

HaemoKBS: A knowledge-based system for real-time, continuous categorisation of adverse reactions in blood recipients



Augusto Ramoa^{a,b}, Jorge Condeço^a, Florentino Fdez-Riverola^{b,c,d}, Anália Lourenço^{b,c,d,e,*}

^aInstituto Português do Sangue e da Transplantação IP, Rua de Bolama, 133, 4200-139 Porto, Portugal

^bEscuela Superior de Ingeniería Informática, Edificio Politécnico, Campus Universitario As Lagoas s/n, University of Vigo, 32004 Ourense, Spain

^cCINBIO – Centro de Investigaciones Biomédicas, University of Vigo, Campus Universitario Lagoas-Marcosende, 36310 Vigo, Spain

^dSING Research Group, Galicia Sur Health Research Institute (IIS Galicia Sur), SERGAS-UVIGO, Hospital Álvaro Cunqueiro 36312, Vigo, Spain

^eCEB – Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal

ARTICLE INFO

Article history:

Received 12 November 2019

Revised 20 March 2020

Accepted 3 April 2020

Available online 28 May 2020

Keywords:

Haemovigilance

Blood recipients

Adverse reactions

Expert knowledge

Machine learning

Knowledge validity

Knowledge and reasoning adaptation

ABSTRACT

This work introduces HaemoKBS, a novel Haemovigilance decision support system for adverse reactions in blood recipients. Machine learning inference and rule-based reasoning were applied to build the underlying decision support models, namely to automatically extract evidence from different types of data included in hospital notifications and incorporate *a priori* expert knowledge. The ultimate aim is to dynamically learn and improve the reasoning abilities of the system and thus, be able to provide educated recommendations to hospital notifiers along with understandable explanations on the acquired knowledge. Experiments over the records of the Portuguese National Haemovigilance System from the last 10 years demonstrate the practical usefulness of HaemoKBS, which will contribute to a better depiction of the adverse reactions and to flag any incomplete notification enforcing data quality.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction and motivation

The use of blood as a therapeutic agent has always been an essential medical therapy, namely for several pathologies, accidents, and surgical procedures. However, such therapy often entails risks [1]. Although the major causes of transfusion-related morbidities, such as blood group incompatibility and transmission of infectious agents, are now significantly reduced, real-time and continuous monitoring is still required for this and other possible adverse reactions [2].

In 1994, the French Blood Agency started setting a network of healthcare professionals with specific roles to signal, notify and review transfusion-related adverse reactions [3]. This monitoring system became the first national Haemovigilance system worldwide. Since then, most of the developed countries have implemented and legislated Haemovigilance programs [4]. The aims are mainly two-fold: (i) to monitor the quality and safety of blood

components throughout the supply-value transfusion chain of national healthcare facilities (i.e. from donor to recipient); and, (ii) to trigger appropriate and timely actions when there is a potential threat to the safety of transfusion.

Existing Haemovigilance systems are usually implemented at the national level. Thus, the number of institutions handling blood transfusions, the number of hospital notifiers (i.e. trained healthcare staff) responsible for notifying about the occurrence of adverse reactions, and the number of annual notifications may vary considerably from country to country [5]. Likewise, the supporting software is quite diverse in nature and degree of automation. Nevertheless, national Haemovigilance organisations are bound to similar laws. For example, current European Union legislation explicitly specifies the level of detail that the Haemovigilance systems of member states should ensure [6].

Aside from legal policies, well-known international medical institutions are invested in the standardisation of this information and its supporting tools [7]. The International Society of Blood Transfusion has a working party on Haemovigilance [8] and the Centers for Disease Control and Prevention released a set of standards for the classification of transfusion reactions [9].

Since the general protocol of notification of adverse reactions in patients transfused with blood products can be considered fairly

* Corresponding author at: ESEI: Escuela Superior de Ingeniería Informática, Edificio Politécnico, Campus Universitario As Lagoas s/n, 32004 Ourense, Spain.

E-mail addresses: augusto.ramoa@ipst.min-saude.pt (A. Ramoa), jorge.condeco@ipst.min-saude.pt (J. Condeço), riverola@uvigo.es (F. Fdez-Riverola), analialourenco@uvigo.es (A. Lourenço).

established for most adverse reactions, computational processing is welcomed to some extent, notably the introduction of intelligent methods of validation that enforce data quality. Therefore, the present work proposes a novel knowledge-based system, named HaemoKBS, which supports the work of notifiers and Haemovigilance experts throughout the process of notifying and curating adverse reactions in blood recipients. The system integrates both expert knowledge and machine learning inference in order to validate and act upon notifications adequately. The main contribution lays on augmenting the discriminatory power of categorisation by continuously learning from previous experience (i.e. validated notifications) as well as prior knowledge (i.e. expert know-how). The rationale is that, by combining the automatic categorisation of adverse reactions with well-established knowledge, the manual effort required from national experts may be reduced while enforcing data quality in a sustainable, educated and continuous way.

The Portuguese Haemovigilance System, which is fully compliant with EU laws and international Haemovigilance guidelines, is used as a case study. Since 2007, this system receives adverse reaction notifications via a restricted access web application. From 2008 to 2017, the system registered, on average, 427 adverse reactions per year [10,11]. Currently, all notifications are manually reviewed by a panel of national Haemovigilance experts that validate the coherence of the designated adverse reaction based on the reported medical observations.

The remainder of the paper is organised into the following sections. Section 2 presents an overview of decision-making systems applied to domains with similar characteristics. Section 3 introduces the general architecture of the proposed system, namely its internal knowledge base and curation workflow. Section 4 demonstrates the validity and potential of the developed system using the notifications collected in the Portuguese National Haemovigilance system. Section 5 states the conclusions drawn from the implementation of HaemoKBS as well as depicts future developments.

2. Related work

Over the past three decades, decision support systems have been successfully established in numerous domains. There is a variety of system types, all of which have the ultimate goal of assisting humans to make appropriate decisions for practical (and often challenging) problems. Among these, expert systems and recommender systems follow the knowledge-driven approach, i.e. take into account various aspects of the domain and integrate different types of knowledge within reasoning to tackle the problem with appropriate actions [12].

Medical decision support systems were introduced as an innovative and valuable way for providing clinicians and patients with varied clinical knowledge and enhancement of the overall quality of care [13,14]. In this regard, several studies have argued the case for implementing systems that are able to provide motivated suggestions [15] as well as encode clinical practice guidelines (i.e. emulate the process followed by experts to take decisions) [16]. Diagnosis prediction and treatment learning and recommendation are long-term lines of research, including challenges in the combination of static and longitudinal data features and the added value of the application of machine learning methods [17]. Nonetheless, considerable attention is also paid to the design of surveillance and reporting systems that rapidly detect and take appropriate action to adverse and undesirable events. Spontaneous reporting has been the main source to detect such events for years, but data quality issues, such as the over-reporting of highly common events, missing and incomplete data, duplicated reporting, voluntary submission, and the high cost of human expert's assessment, raise the

need for automated and intelligent mechanisms of notification and validation.

Pharmacovigilance, i.e. the detection and continuous monitoring of adverse drug reactions or other drug-related problems, is actively engaged in implementing surveillance knowledge-based systems [18]. Databases, such as the Adverse Event Reporting System of the Food and Drug Administration, contain reports on adverse events that have been spontaneously communicated by clinical staff and consumers. Nevertheless, the success of surveillance systems lies in the quality of the submitted reports and manual expert validation [19,20]. Usually, evaluators try to find additional, similar cases in the system, in other databases or in the medical literature and search for any common trend, causal relationship, or pattern of events to identify potential risk factors or any other specificities. One of the main challenges in determining the causality of such events is that there is no gold standard (i.e. ground truth) to compare the relative performance of computational algorithms and expert's assessment. Therefore, increasing attention is paid to the application of retrospective data (including administrative claims, electronic health records, laboratory test results, radiology test results, and other alike) as well as to the encoding of standardised assessments and any regulations (often expressed as a set of specific 'yes/no' questions regarding event features) typically used by human experts (and many times, generated by previous experience). A recent review pinpointed the specificities and challenges of decision support in Pharmacovigilance, including the methods and frameworks used for processing such data, and the machine learning algorithms developed for the detection of adverse drug reactions [21].

To the best of the authors' knowledge, there is no previous attempt to introduce automated decision making in Haemovigilance systems. Current systems rely solely on manual expert curation to guarantee the timely and accurate characterisation of adverse reactions in blood recipients and blood donors [5]. So, in this domain, data quality is still heavily dependent on the prompt and comprehensive notification of the potential reaction as well as the experience of the curators in assessing cases that do not fully comply with the common, standardised description of the reactions.

Decision support can be introduced in Haemovigilance systems in various, combinable ways with the ultimate objectives of enhancing data quality, reducing costs with expert curation, and enforcing efficient monitoring. Most notably, lessons can be learned from developments in similar (purpose and concern-related) domains, such as Pharmacovigilance, as well as from other health expert systems that aim to emulate the behaviour of human experts, facilitate human collaboration, and capture implicit domain knowledge for educated decision-making.

The so-called hybrid decision support systems that combine expert-defined static rules with supervised learning are well-established in medical domains. In particular, the number of medical applications of case-based reasoning systems has increased considerably over the last decade [22–24]. Generally speaking, a case can be defined as a set of meaningful domain features coupled with information about the applied "solution", i.e. a situation-action domain-specific pattern [25]. The success of the methodology lies in three main capabilities, namely: the ability to produce a rich and evolvable representation of both implicit (i.e. previous cases) and explicit knowledge (i.e. human expertise); the ability to implement efficient and flexible ways to retrieve existing data; and, the ability to apply analogical reasoning to solve new situations [25]. This is well-described in the review of Begum and colleagues, which summarises 34 case-based reasoning medical systems, in terms of purpose-oriented properties and construction characteristics [26]. Also interesting, another review explored intrinsic weaknesses affecting the application of this methodology

to the medical domain, notably difficulties in feature mining, presence of competence gaps affecting case base contents, challenges in the definition of a suitable adaptation strategy, and other complementary issues related to the increasing complexity of the domain [27].

On top of all these developments, it is also worth mentioning the development of rule-based systems that model clinical knowledge using sets of rules, in the form of IF-THEN statements [28]. That is, based on the available cases, the system selects one rule from the knowledge base to be fired in order to generate new knowledge or execute the corresponding actions. Forward chaining is the most extended reasoning mechanism for clinical prognosis, monitoring and control rule-based applications, i.e. checking the left-hand side of the rules against the available cases and launching the right-hand side of the rule meeting the conditions. Likewise, and regardless of the architecture of the implemented systems, the potential of machine learning methodologies has been widely tested for improving disease detection, clinical decision support, and workflow efficiencies [29]. Finally, various methodologies have been proposed to extract and acquire medical knowledge, including the use of theoretical models to interchange knowledge between rules [30,31].

3. HaemoKBS: application architecture and operational process

The main aim of HaemoKBS is to help in the automatic classification and validation of the notified adverse reactions based on the provided observations and prior expert knowledge. As illustrated in Fig. 1, the hospital notifier reports a collection of medical observations that are used by the hybrid classification model to produce a prioritised list of possible reactions. The higher is the number of similar symptoms and signs, combined with different ponderations for specific reactions, the higher is the probability of the reaction type (as an example, in a hypotensive reaction, low blood pressure should be overvalued if it appears as a single symptom and undervalued if accompanied by other symptoms). After the notifier makes a decision (which may not necessarily lie on the reaction

with top probability), the reported case is matched against the decision tree corresponding to the selected reaction type. If it complies with expert knowledge, the notification is considered valid and saved in the knowledge base. Otherwise, the system will alert the hospital notifier of the discrepancies so that the notification can be amended or submitted as-is. Whenever a notification is submitted without conformity with prior knowledge, it is sent to the Haemovigilance expert for manual validation. The Haemovigilance expert may accept or reject the proposed classification, with the added possibility of postponing the decision until additional evidence becomes available (e.g. after asking the hospital notifier for further details); when the notification is finally validated, it is stored in the knowledge base for future reuse, namely to retrain and apply the predictive model. The national expert decides whether retraining is necessary, and may then provide additional, useful information for the update process.

3.1. Knowledge storage and notification definition

The knowledge base of the HaemoKBS system keeps all the validated notifications. Each notification stored in this database has a unique identifier and is defined by the information features described in Table 1. Apart from the timestamps of the notification and the occurrence of the adverse reaction, the notifier specifies the age and gender of the patient and characterises the reaction in terms of imputability and severity of the reaction, i.e. the strength of the association between the transfusion and the occurrence of adverse reaction and the severity of the episode, as well as signs, symptoms and exam results.

The Adverse Transfusion Reaction (ATR) (e.g. anaphylaxis, bacterial infection, acute haemolytic reaction, or hypotensive transfusion reaction, coded numerically for simplicity) is described by the feature tuple $\{I, S, G, A, SG, SM, M, O\}$ as follows: I indicates how strong is considered the association between the transfusion and the adverse reaction (possible values are “evaluation not possible”, “excluded”, “possible”, “probable” and “definite”); S specifies the degree of severity of the adverse reaction (possible values include

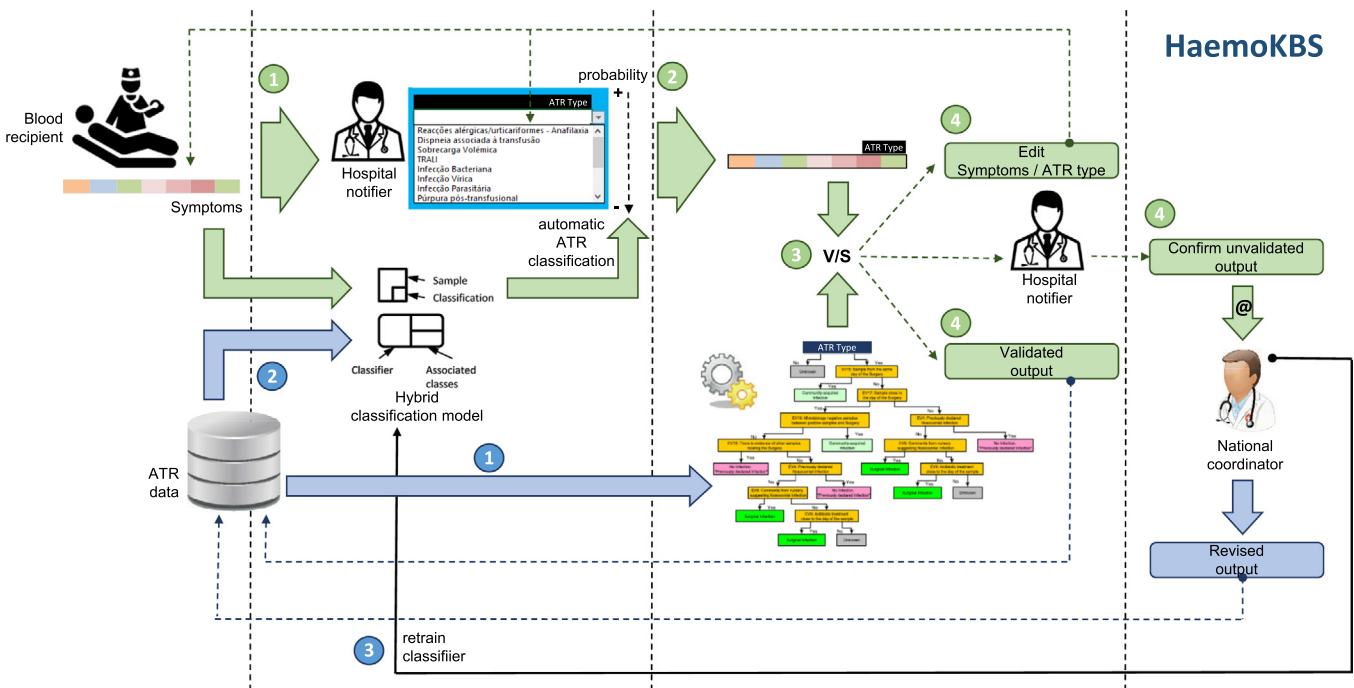


Fig. 1. General architecture of the knowledge-based decision support system proposed for the notification of adverse transfusion reactions (ATR) in recipients.

Table 1
Features included in an adverse reaction notification and used by HaemokBS.

Attribute	Type	Description
Reaction code (RC)	Nominal	Identification of adverse reaction type
Imputability (I)	Nominal	Strength of the association between the transfusion and the adverse reaction. Possible values: evaluation not possible, excluded, possible, probable and definite
Severity (S)	Ordinal	Degree of the severity of the adverse reaction. Possible values: non-severe, severe, life-threatening, death
Reaction type (ATR)	Nominal	An adverse reaction can occur from during the transfusion to months later. Possible values: acute and delayed
Gender (G)	Nominal	Gender of the patient
Age (A)	Nominal	Age (in years) of the patient in the moment of transfusion was discretised into 3 data bins (i.e. $0 < \text{age} < 18$, $18 \leq \text{age} < 65$, and ≥ 65).
Signs (SG)	Boolean	Medical signs observed in the patient by the healthcare staff. Possible values: present and absent
Anguish/Anxiety; Arrhythmia; Bradycardia; Bronchospasm/Sibilant; Cough; Cyanosis; Dark urine; Disorientation; Dyspnoea; Epistaxis; Erythema/Generalized Redness; Erythema/Localized Redness; Fever/Temperature rise; Generalized Oedema; Haemorrhage; Hypertension; Hypotension; Hypothermia; Jaundice; Localized Angioedema; Localized Oedema; Maculopapular rash/Urticaria; Oligoanuria; Paleness; Psychomotor agitation; Quincke's Oedema; Skin Rash; Sweating/Diaphoresis; Syncope/Loss of consciousness; Tachycardia; Tachypnoea/Polypnea/Hyperpnoea; Tonic-clonic seizure; Urticaria; Vomiting		
Symptoms (SM)	Boolean	Symptoms mentioned by the patient to the healthcare staff. Possible values: present and absent
Abdominal pain; Arthralgia; Diarrhoea; Headaches; Low back pain; Nausea; Orthopnoea; Pain/Burning in the venous access; Shivers/Tremors; Thoracic Pain; Vertigo/Dizziness		
Medications (M)	Boolean	Medications administered to the patient. Possible values: present and absent
Adrenaline; Anti-histamines/steroids; Antipyretics; Bronchodilators; Diuretics; Fluid Therapy; Hypotensors; Metoclopramide; Oxygen therapy		
Others (O)	Boolean	Diagnostic test results and other characteristics that help in the diagnosis and identification of the adverse reactions. Possible values: present and absent
Acute lung oedema; Cardiac arrest; Cardiac insufficiency; Decreased O2 sat; Evidence of Positive Fluid Balance; Other unlisted findings; Shock; Viral Infection, Absent before transfusion		

/ indicates alternative feature naming conventions.

“non-severe”, “severe”, “life-threatening”, and “death”); G represents the gender of the transfusion recipient (i.e. “female” and “male”); A indicates the age of the transfusion recipient (i.e. age in years of the recipient in the moment of transfusion, which was later discretized into 3 bins); $SG = \{SG_1, SG_2, \dots, SG_{34}\}$ is the set of observed signs (e.g. bradycardia, cyanosis or epistaxis, represented as Boolean variables); $SM = \{SM_1, SM_2, \dots, SM_{11}\}$ is the set of manifested symptoms (e.g. headaches, nausea or vertigo, all represented as Boolean variables); $M = \{M_1, M_2, \dots, M_9\}$ is the set of medications ministered to the patient (e.g. antipyretics, diuretics or hypotensors, all represented as Boolean variables); and, $O = \{O_1, O_2, \dots, O_7\}$ collects other possible manifestations of the adverse reaction (e.g. acute lung oedema or decreased oxygen saturation, all represented as Boolean variables).

3.2. Expert knowledge representation and inference on prior notifications

The Portuguese Blood and Transplantation Institute has published standardised descriptions, namely signs, symptoms and other features, of the most common adverse reactions in their blood recipients [32]. Such knowledge is represented as rules stored in the knowledge base and a rule-based engine implemented in-house enables the execution of these rule sets inside the system. Each adverse reaction type is represented by a decision tree, from which a WHEN–THEN rule set is easily derived by defining a new rule for each path from the root to a leaf node. Following this strategy, the leaf node contains the decision part of the rule whereas the path from the root to the leaf node determines its conditions (i.e. the corresponding collection of signs, symptoms and other possible features). As illustrative examples, Figs. 2 and 3 present the decision trees established for respiratory reactions and

haemolytic reactions. In particular, Fig. 2 describes respiratory reactions as reactions happening hours after the transfusion and, most specifically, states that transfusion-associated circulatory overload (TACO) and transfusion-related acute lung injury (TRALI) happen within the first 6 h whereas transfusion-associated dyspnoea (TAD) occurs between the 6 to 24 h after. Differentiation between TACO and TRALI reactions depends on the predominance of other medical observations (e.g. presence of pulmonary oedema or oxygen saturation below 90%). Likewise, Fig. 3 divides haemolytic reactions into those occurring within the first 24 h and delayed reactions. Then, clinical tests (e.g. DAT test) and the observation of physical causes of haemolysis are determinant to establish whether the reaction is immune or not.

As depicted in Fig. 1, the hospital notifier is free to accept the reaction code recommendation provided by the predictive model of HaemokBS (typically the one with the highest score) or make another decision. Regardless, the notification will be always validated against the expert knowledge, i.e. notification data will be matched against the decision tree built by human experts for that adverse reaction type.

Consistent notifications will be automatically validated and inserted in the knowledge base. In case of discrepancy, the failed decision rule helps to explain which observations are missing or are inconsistent with that reaction type. The hospital notifier is given then the opportunity to make two possible amendments: (i) to update the set of observations described in the notification (i.e. rectify any of the values), and/or (ii) to choose another reaction. If no amendment is made, the inconsistent notification will be sent to the Haemovigilance expert for manual curation.

Validated notifications are inserted in the knowledge base and are applied to re-train the predictive models whenever the national expert considers it appropriate.

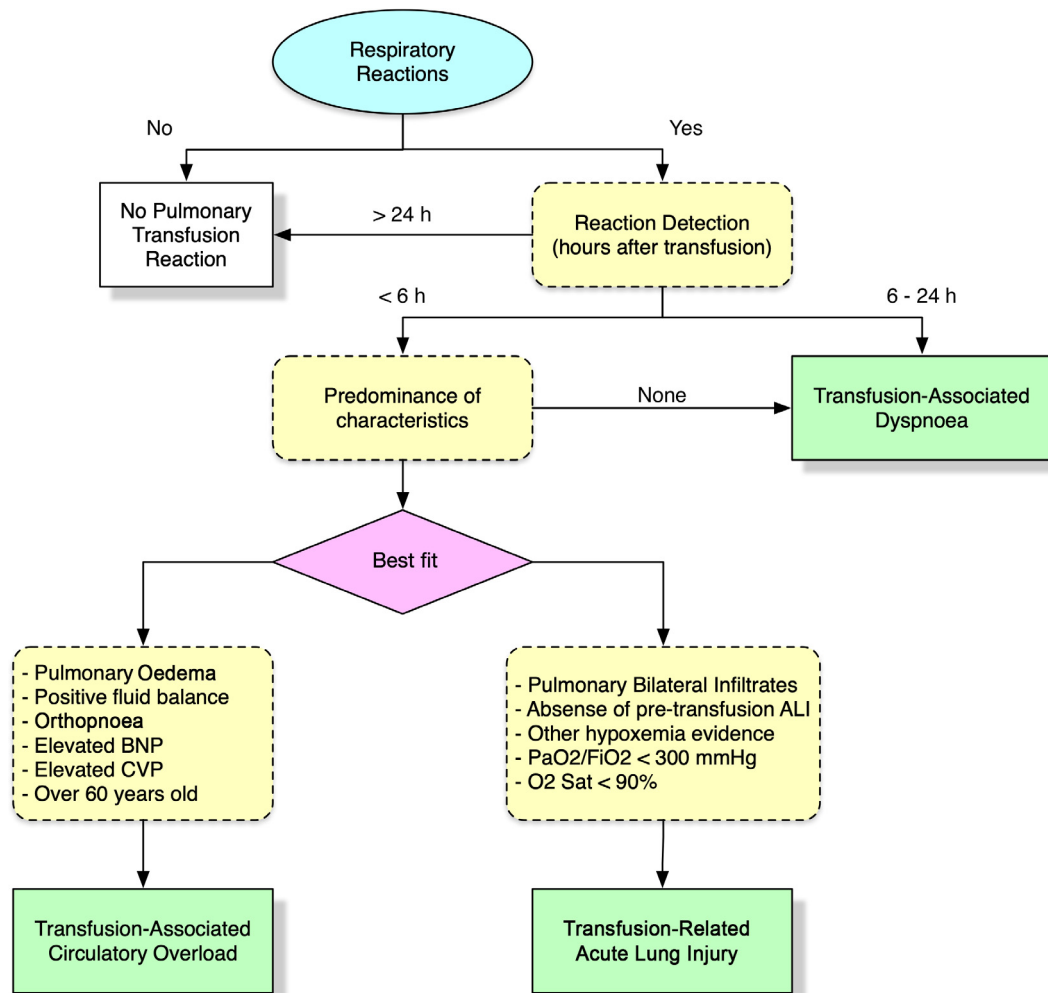


Fig. 2. Manually defined decision tree used by experts to classify respiratory reactions, i.e. transfusion-associated dyspnoea (TAD), transfusion-associated circulatory overload (TACO) and transfusion-related acute lung injury (TRALI).

4. Results and discussion

In this section, the predictive performance of the models proposed for the classification of adverse reactions in blood recipients is evaluated. Both the experimental protocol and the results are presented and discussed.

4.1. Experimental setup

The records of the Portuguese National Haemovigilance System supported the training and testing of the predictive models. These records are not published individually, as they contain sensitive information, nevertheless they are compiled by the Portuguese Haemovigilance System and a descriptive report is published annually [10].

The primary goal was to predict the most probable types of adverse reactions based on the observations supplied by the hospital notifiers.

Experiments considered notifications from the last 10 years of activity, which have been manually curated by experts. Fig. 4 shows the distribution of these notifications per type of adverse reaction throughout the years. This dataset contains a total of 3,829 notifications distributed as follows: the febrile non haemolytic transfusion reactions (i.e. ATR code 15) account for 49% of the

notifications, followed by the acute allergic reactions (i.e. ATR code 16), which represent 26% of the notifications; each of the other of adverse reactions represents less than 10% of the dataset. Most notably, notifications of transfusion-related acute lung injuries (i.e. TRALI, with ATR code 9), viral infections (i.e. ATR code 10), bacterial infections (i.e. ATR code 7) and delayed haemolytic transfusion reaction (i.e. ATR code 14) are almost inexistent (below 1%).

This dataset was divided into a training set (2,873 examples) and a test set (956 examples). Initially, experiments focused on validating the performance of predictive models trained to recognise all possible adverse reaction types (i.e. the implementation of the discriminative model), however, these predictive models achieved poor performance. As such, the training set was further divided into three subsets, namely a subset containing only the notifications of the most frequent reactions (i.e. haemolytic transfusion reactions and acute allergic reactions), another one keeping the seldom occurring reactions (i.e. TRALI, viral and bacterial infections and delayed haemolytic transfusion reactions), and a third one containing the rest of the notifications (i.e. those concerning ordinarily occurring reactions).

While the unbalanced distribution of adverse reaction types challenged the machine learning processes, sampling techniques (or similar) were not considered because, in countries with High Development Index [33] (such as the present case), Haemovigi-

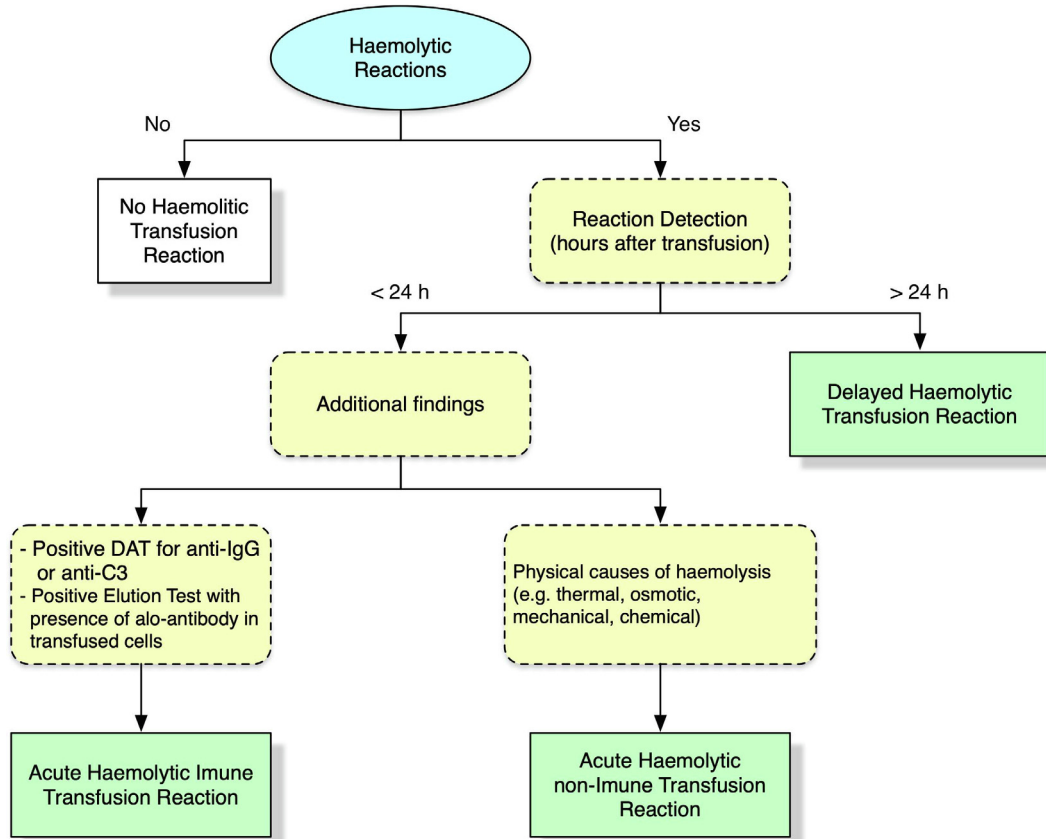


Fig. 3. Manually defined decision tree used by experts to classify haemolytic reactions, i.e. delayed haemolytic transfusion reaction, acute haemolytic immune transfusion reaction and acute haemolytic non-immune transfusion reaction.

lance is well-established and the number of notifications of adverse reactions is expected to be low and, in particular, adverse reactions such as TRALI have become highly unlikely to occur.

Therefore, the proposed classification model was devised as a hybrid decision-maker that handled the whole spectrum of possible adverse reactions while providing specialised outputs for the three subsets of reactions. Most notably, and as represented in Fig. 5, the constructed hybrid classification model includes a discriminative model that predicts the most probable ATR category for the notified data, plus three models specialised in the prediction of the specific ATR code: prevalent reactions (ATR codes 15 and 16), ordinary reactions (ATR codes 1, 17, 8, 18, 23 and 3) and rare reactions (ATR codes 9, 10, 7 and 14).

All experiments conducted 10 fold cross-validation over the training set and used the test set to perform a final validation over previously unseen examples [34]. Exception made for the rare reaction types (covering less than 2% of the dataset), for which experiments carried out leave-on-out validation. The classification models were developed with classic predictive algorithms, including decision trees (DT) [35,36], Naïve Bayes (NB) [37], k-nearest neighbours (k-NN) [37], support vector machines (SVM) [38,39], random forest (RF) [40], adaptive boosting algorithms (AdaBoost) [41–43], artificial neural networks (ANN) [44] and 5 Nearest Neighbours (5NN) [45]. A grid search was applied to perform hyperparameter optimisation, i.e. parameter combinations were explored and the parameters that resulted in the best models were chosen [46,47].

The micro-averaged values of Cohen’s kappa coefficient (Eq. (1)), accuracy (Eq. (2)), error rate (Eq. (3)), precision (Eq. (4)), recall (Eq. (5)), specificity (Eq. (6)), and F-score (Eq. (7)) were used as evaluation metrics [48].

$$k \equiv \frac{p_0 - p_e}{1 - p_e} = 1 - \frac{1 - p_0}{1 - p_e}, \tag{1}$$

$$Accuracy = \frac{\text{number of true positives and true negatives}}{\text{total of predictions}} \tag{2}$$

$$Error\ rate = \frac{\text{number of false positives and false negatives}}{\text{total of predictions}} \tag{3}$$

$$Precision = \frac{\text{number of true positives}}{\text{number of predicted positives}} \tag{4}$$

$$Recall = \frac{\text{number of true positives}}{\text{number of actual positives}} \tag{5}$$

$$Specificity = \frac{\text{number of true negatives}}{\text{number of actual negatives}} \tag{6}$$

$$F\text{-score} = 2 \times \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{7}$$

Experiments were conducted using the RapidMiner implementations of the aforementioned machine learning algorithms [49].

4.2. Results of model building

Tables 2–5 summarise the evaluation results for the machine learning models, i.e. a total of 28 classification models were considered for the prediction of adverse reactions in blood recipient based on 7 machine learning methods and 4 training sets.

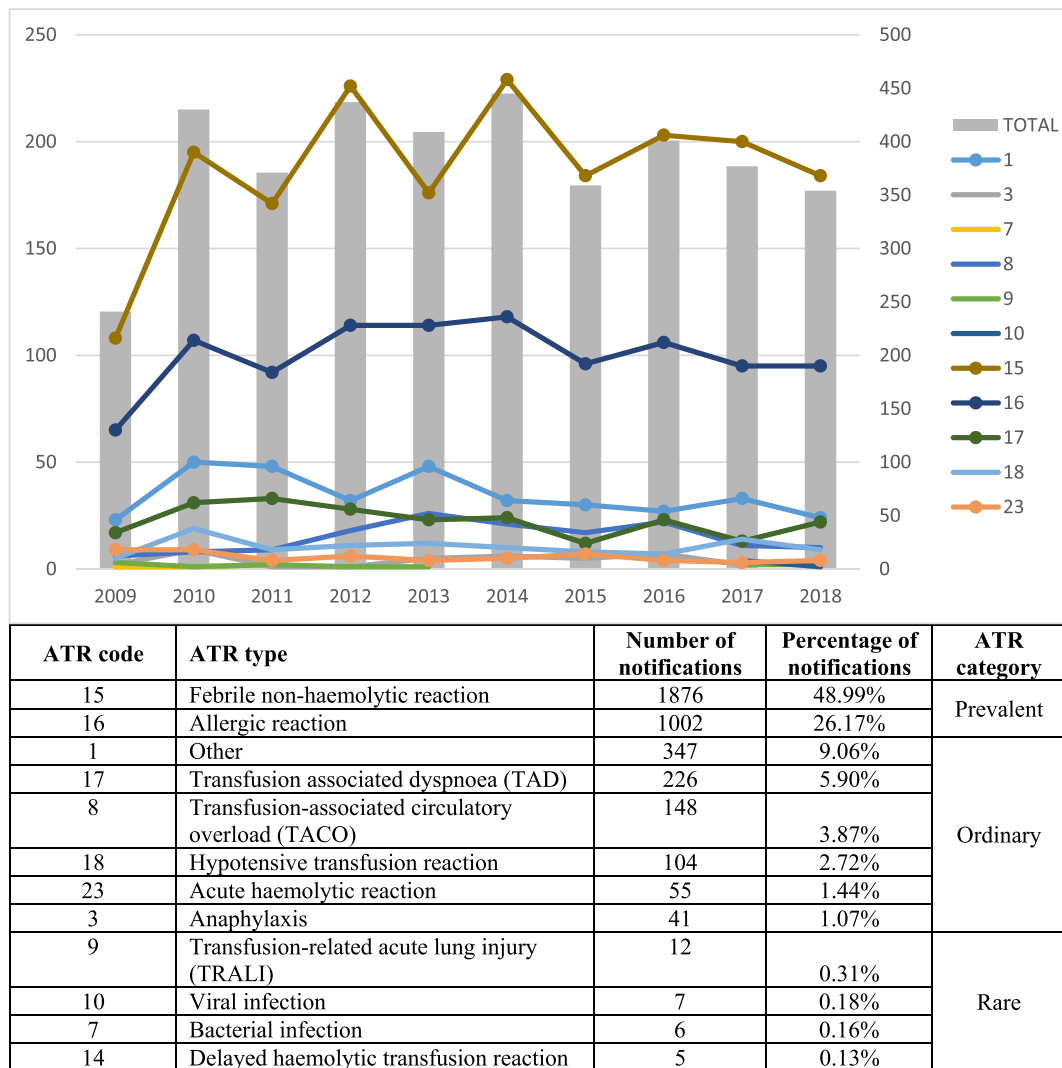


Fig. 4. Distribution of the adverse reaction notifications recorded by the Portuguese National Haemovigilance System in the last decade.

As mentioned previously, the predictive models covering all possible adverse reaction types presented poor performance (Table 2). Notably, models showed poor ability to predict bacterial infections (ATR code 7), TRALI (ATR code 9), and delayed haemolytic transfusion reactions (14). Febrile non-haemolytic transfusion reactions (ATR code 15) and acute allergic reactions (ATR code 16), i.e. the most commonly notified reactions, obtained the best results, i.e. precision around or above 0.9 and recall, most of the times, well above 0.9. Models showed modest values of performance for TACO (ATR code 8), TAD (ATR code 17) and other acute reactions (ATR code 1), i.e. most precision and recall values below 0.6. Overall, the top-performing models were RF (a grid search supported hyperparameter optimization considering combinations of the number of estimators $\in [50, 100, 150, 200, 250, 300, 500]$, maximum depth $\in [20, 25, 30, 35, 40]$, and minimum samples in a leaf $\in [1, 10, 20, 50, 100]$, and the best parameters were number of estimators = 200, maximum depth = 35, and minimum samples in a leaf = 1) and SVM (a grid search supported hyperparameter optimization combining values of $C \in [0.001, 0.01, 0.1, 1, 10, 100]$ and values of $\gamma \in [0.001, 0.01, 0.1, 1]$ and the best parameters were $C = 10$ and $\gamma = 0.01$) with a kappa metric of 0.75 and a classification rate of 16% and 17%, respectively. The most noticeable difference between these two models was the predictive ability to classify acute allergic reactions (ATR code 16), which are the

second most common reactions. Considering the poor performance of the RF model for this case (i.e. recall of 38%), the SVM model was chosen as the discriminative component of the hybridised classification model. Notably, this model is used to determine which category-specific classification model should be applied to produce the final prediction, i.e. the prevalent type classifier, the ordinary type classifier, or the rare type classifier.

These results also offered relevant insights into the differentiation of similar reactions based on notification data. For example, false negatives on respiratory reactions were deemed two-fold, i.e. TAD with ATR code 17 and TACO with ATR code 8, are often due to misclassifications among themselves, whereas TRALI (ATR code 9) was confused with acute haemolytic transfusion reaction (ATR code 23) or other acute reaction (ATR code 1). The characterisation of respiratory reactions is still looking for a final consensus among experts. Working groups are addressing this characterisation, specially TRALI and TACO reactions [50–53], and therefore, to an extent, it was somewhat expected and justifiable misclassifications between them. The confusion with acute haemolytic transfusion reactions will be addressed in future developments with the harmonization of the diagnostic test results and other characteristics, namely by implementing a rule that strictly forbids a given classification when a specific relevant test result (i.e. a positive result in the direct antiglobulin test) is not available.

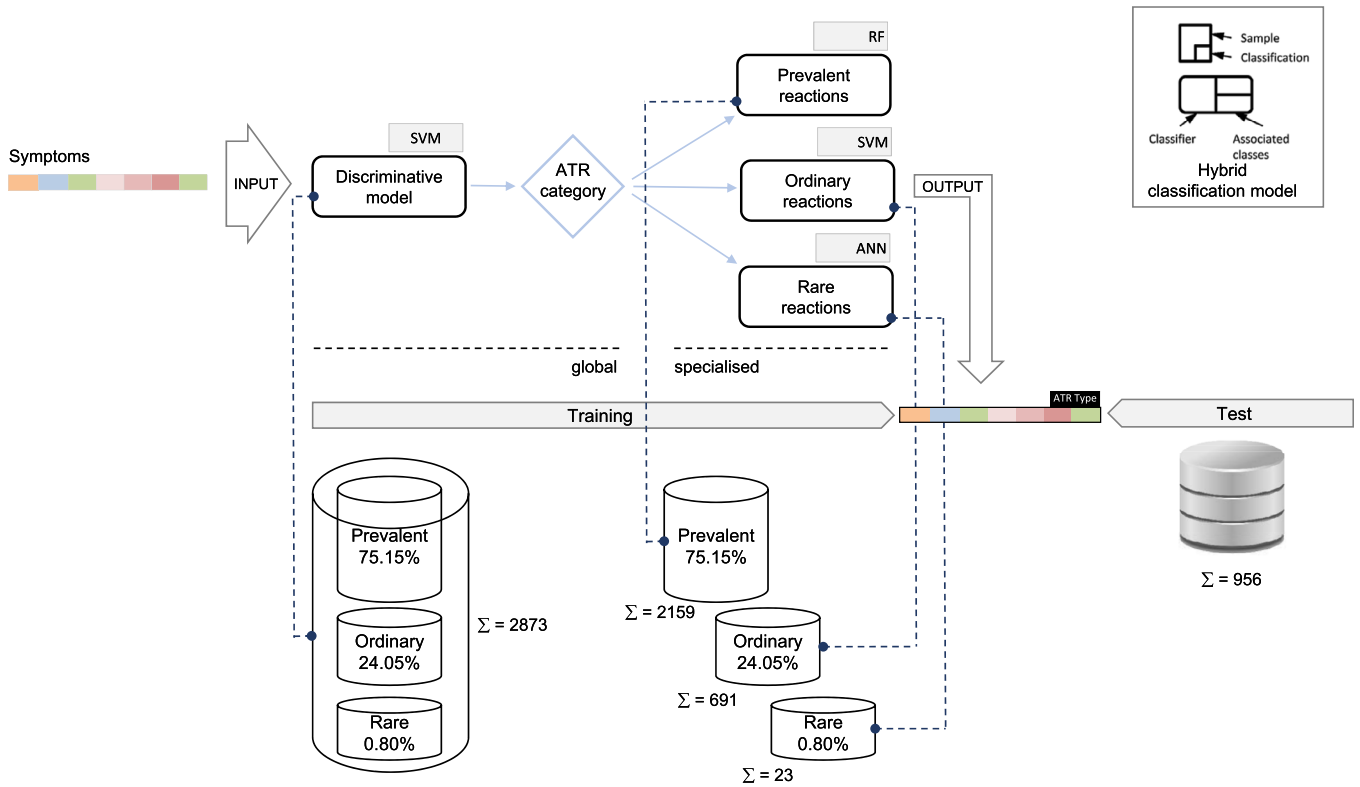


Fig. 5. Hybrid classification model architecture (in the upper right corner, the hybrid classification model showed in Fig. 1, acting as a visual aid to zoom in on the general architecture of HaemoKBS).

Table 2 Performance of the global predictive models covering all ATR types by 10-fold cross-validation.

	Kappa	% accuracy	ATR code														Precision/Recall
			15	16	1	17	8	18	23	3	9	10	7	14			
DT	0.71	80.61	0.9025	0.9168	0.493	0.5041	0.4894	0.5882	0.64	0.2069	0	1	0	0			
			0.9339	0.8644	0.5423	0.7176	0.2072	0.5128	0.3902	0.1935	0	0.8	0	0			
NB	0.75	83.08	0.9255	0.8895	0.5689	0.5297	0.5481	0.7662	0.7297	0.3438	0.1429	0.6667	0	0			
			0.936	0.8989	0.4923	0.6294	0.5135	0.7564	0.6585	0.3548	0.1111	0.8	0	0			
SVM	0.75	83.15	0.9062	0.8971	0.6818	0.552	0.7119	0.7195	0.7368	0.8	0	1	0	0			
			0.941	0.9043	0.4038	0.7176	0.3784	0.7564	0.3415	0.129	0	0.8	0	0			
RF	0.75	83.78	0.8771	0.7119	0.6818	0.552	0.7195	0.8971	0.8	0.7368	0	0	0	1			
			0.9794	0.3784	0.4038	0.7176	0.7564	0.9043	0.129	0.3415	0	0	0	0.8			
AB	0.73	82.6	0.8824	0.8768	0.5907	0.569	0.567	0.7258	0.619	0.75	0	1	0	0			
			0.9652	0.9082	0.4385	0.5824	0.4955	0.5769	0.3171	0.0968	0	0.6	0	0			
ANN	0.74	82.84	0.9188	0.9092	0.5375	0.5102	0.5755	0.6974	0.7059	0.2632	0	0.5	0.2	0			
			0.9488	0.8923	0.4962	0.5882	0.5495	0.6795	0.5854	0.1613	0	0.2	0.2	0			
5NN	0.72	81.73	0.8537	0.892	0.5515	0.5439	0.6	0.7302	0.8333	1	0	1	0	0			
			0.9787	0.9003	0.35	0.5471	0.4865	0.5897	0.122	0.0323	0	0.8	0	0			

Predictive models specialised in the most prevalent reactions, i.e. febrile non-haemolytic transfusion reactions (ATR code 15) and acute allergic reactions (ATR code 16), obtained very good results, i.e. a kappa over 0.92 and most values of precision and recall above 0.95 (Table 3). The RF was the best performing model with a rate of error in the classification of 2.4%.

Discriminating between ordinary reactions, i.e. transfusion-associated dyspnoea (ATR code 17), hypotensive transfusion reaction (ATR code 18), anaphylaxis (ATR code 3), acute haemolytic transfusion reaction (ATR code 23), TACO (ATR code 8) and other acute reactions (ATR code 1), was somewhat more challenging (Table 4). At the best-case scenario, classification error was around 25%, mainly due to the poor predictive performance exhibited for anaphylaxis (ATR code 3) and TACO (ATR code 8), which had values of recall often below 0.5. Results for acute haemolytic transfusion reaction (ATR code 23) and transfusion-associated dyspnoea (ATR

code 17) also presented limited precision and/or recall abilities. Overall, the SVM was the best-performing model.

As shown in Table 5, the classification of the rare reactions (i.e. those with seldom representation in the knowledge base) was challenging even for specialised models, i.e. half of the models achieved a kappa value between 0.5 and 0.6. The ANN and the DT were the only models able to produce decent predictions for viral infections (ATR code 10), TRALI (ATR code 9) and bacterial infections (ATR code 7) in a consistent manner. Most notably, those models were the only ones able to classify delayed haemolytic transfusion reactions (ATR code 14) satisfactorily.

4.3. Results from the test phase

Based on the training phase results, the final hybridised classification model included a SVM model for discriminating the most

Table 3
Performance of the specialised predictive models for the prevalent ATR category: febrile non-haemolytic transfusion reactions (ATR code 15) and acute allergic reactions (ATR code 16) by 10-fold cross-validation.

	Kappa	% accuracy	ATR code		Precision/Recall
			15	16	
DT	0.92	96.2	0.9675	0.9515	
			0.9744	0.9388	
NB	0.93	96.62	0.9744	0.9509	
			0.9737	0.9521	
SVM	0.94	97.36	0.9707	0.9793	
			0.9893	0.9441	
RF	0.95	97.6	0.9728	0.9821	
			0.9908	0.9481	
AB	0.94	97.22	0.972	0.9727	
			0.9858	0.9468	
ANN	0.93	96.9	0.9725	0.9622	
			0.9801	0.9481	
5NN	0.93	97	0.966	0.9777	
			0.9886	0.9348	

Table 4
Performance of the specialised predictive models for the ordinary ATR category: transfusion-associated dyspnoea (ATR code 17), hypotensive transfusion reaction (ATR code 18), anaphylaxis (ATR code 3), acute haemolytic transfusion reaction (ATR code 23), TACO (ATR code 8) and other acute reactions (ATR code 1) by 10-fold cross-validation.

	Kappa	% accuracy	ATR code						Precision/Recall
			1	17	18	23	3	8	
DT	0.49	62.37	0.7224	0.6058	0.7246	0.3684	0.2258	0.4844	
			0.7808	0.7412	0.641	0.3415	0.2258	0.2793	
NB	0.61	71.06	0.8148	0.6412	0.8267	0.7632	0.3714	0.5631	
			0.8462	0.6412	0.7949	0.7073	0.4194	0.5225	
SVM	0.66	74.96	0.8235	0.6384	0.7753	0.9583	0.8333	0.7105	
			0.8615	0.8412	0.8846	0.561	0.1613	0.4865	
RF	0.62	72.06	0.7942	0.6183	0.7416	0.7826	0.5385	0.7917	
			0.8462	0.8765	0.8462	0.439	0.2258	0.3423	
AB	0.55	67.44	0.7196	0.6078	0.7727	0.8571	0.4444	0.5895	
			0.8192	0.7294	0.6538	0.439	0.129	0.5045	
ANN	0.61	70.91	0.8231	0.6188	0.7831	0.7576	0.4706	0.5641	
			0.8231	0.6588	0.8333	0.6098	0.2581	0.5946	
5NN	0.52	65.56	0.6901	0.5922	0.7436	1	0.6667	0.5714	
			0.8308	0.7176	0.7436	0.1707	0.0645	0.4324	

Table 5
Performance of the specialised predictive models for the rare ATR category: viral infections (ATR code 10), delayed haemolytic transfusion reactions (ATR code 14), TRALI (ATR code 9) and bacterial infections (ATR code 7) by leave-one-out validation.

	Kappa	% accuracy	ATR code				Precision/Recall
			10	14	7	9	
DT	0.71	78.26	1	0.8	0.5714	0.8571	
			0.8	1	0.8	0.6667	
NB	0.58	69.57	0.5	1	0.6667	0.875	
			0.8	0.25	0.8	0.7778	
SVM	0.58	69.57	0.5	0	0.8	0.8889	
			0.8	0	0.8	0.8889	
RF	0.57	69.57	0.4444	0	1	0.9	
			0.8	0	0.6	1	
AB	0.57	69.57	0.6667	0.5	0.6667	0.8	
			0.8	0.5	0.4	0.8889	
ANN	0.76	82.61	1	0.8	0.6667	0.875	
			0.8	1	0.8	0.7778	
5NN	0.52	65.22	0.5	0.3333	0.75	0.875	
			0.8	0.25	0.6	0.7778	

probable type of reaction, a RF model for classifying prevalent reactions (i.e. febrile non-haemolytic reactions and allergic reactions), an ANN model specialised in rare reactions (i.e. viral infections, TRALI, bacterial infections, and delayed haemolytic transfusion reactions), and a SVM for classifying ordinary reactions (i.e. all other ATRs).

The predictive capability of this hybridised classification model was further demonstrated in the test phase, as the data in the test set were completely independent of the training set and not used for model building. From the previously unseen 959 examples,

the model was able to accurately predict the reaction type of 823 notifications. The 136 false predictions are described in Table 6 and Fig. 6. Most errors related to the false prediction of febrile non-haemolytic transfusion reactions (ATR code 15) and transfusion-associated dyspnoea (ATR code 17).

The analysis of these misclassifications confirmed our suspicions regarding the need to implement new notification features and reorganise the signs, symptoms, diagnostic test results, medications and other features currently in use. In more detail:

Table 6
Performance of the hybridised classification model over the test set.

	Final prediction												
	1	3	7	8	9	10	14	15	16	17	18	23	
True class	1	0	0	2	0	0	0	24	7	6	3	1	
	3	1	0	1	0	0	0	1	3	3	0	0	
	7	0	0	0	0	0	0	1	0	0	0	0	
	8	3	0	0	0	0	0	3	0	14	0	0	
	9	0	0	0	0	0	0	0	0	3	0	0	
	10	1	0	0	0	0	0	0	0	0	0	0	
	14	1	0	0	0	0	0	0	0	0	0	0	
	15	5	0	0	0	0	0	0	5	3	2	0	
	16	1	0	0	0	0	0	3	6	2	0	0	
	17	1	0	0	5	0	0	6	5	1	0	0	
	18	0	0	0	0	0	0	0	2	1	0	0	
	23	1	0	0	0	0	0	7	0	1	1	0	

- Anaphylactic reactions (ATR code 3) were misclassified 9 times:
 - o In 3 of these notifications, the severity of the reaction was classified as “Non severe”, which is not correct.
 - o Other 3 of these notifications were classified as acute allergic reactions (ATR code 16), whose feature characterisation is currently very similar.
- TACO (ATR code 8) was misclassified 20 times and, most notably, 14 of these cases were misclassified as TAD (ATR code 17).
- TRALI (ATR code 9) was misclassified 3 times, all of them as TAD (ATR code 17), which is in line with the previously explained difficulty to distinguish between TACO (ATR code 8), TRALI (ATR code 9) and TAD (ATR code 17) reactions [50–53].
- Febrile non-haemolytic transfusion reactions (ATR code 15) were misclassified 15 times:
 - o In 9 of these cases, the sign fever was not selected, which is not correct.
 - o In other 5 cases, other symptoms, such as rash, were also selected which led the predictive model to consider them more similar to acute allergic reactions (ATR code 16).
- Acute allergic reactions (ATR code 16) was misclassified 12 times:
 - o 6 of these cases were classified as transfusion-associated dyspnoea (ATR code 17), because the symptom dyspnoea was present, which is the most important symptom of this reaction type.
 - o In the remaining 6 cases, classified as other acute reactions (ATR code 1), febrile non haemolytic transfusion reactions (ATR code 15), and hypotensive transfusion reactions (ATR code 18), no symptom specific to allergic reactions (ATR code 16) was present.

- Hypotensive transfusion reactions (ATR code 18) were misclassified 3 times because the cases were over-characterized, i.e. more symptoms and signs were selected.

The study of these cases is essential to define strategies to decrease misclassifications. In this context, the implementation of reaction-targeted rules that hamper notifications to be submitted when key symptoms are not selected by the notifier, as well as a ponderation system that prevents overvaluation of some symptoms in specific reactions (i.e. a symptom when identified together with others, should be less valued than when it is identified alone) seem to be relevant issues to be addressed.

5. Conclusions and further work

This paper introduced a hybrid decision support system, based on machine learning inference and rule-based reasoning, to assist in the categorisation of adverse reactions in blood transfusion recipients. The system indicates the most probable reaction(s) according to the observed signs and symptoms and the knowledge acquired from prior, validated notifications. The hospital notifier may decide to accept the predicted type or to propose another one, i.e. based on his experience or additional observations. Notification categorisation is automatically validated against the expert decision rules defined in accordance with European laws. The hospital notifiers and, ultimately, the Haemovigilance experts manually resolve the non-compliances.

The experimental testbed carried out for 10-year records of the Portuguese National Haemovigilance System demonstrated that the implementation of a decision support system such as the one

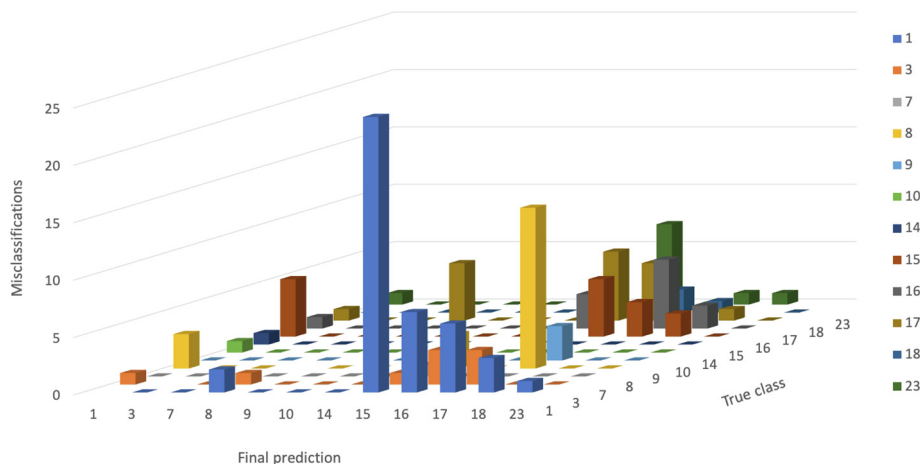


Fig. 6. Graphical representation of the confusion matrix shown in Table 6.

proposed here is of practical use to the operation of national Haemovigilance systems. Most notably, data introduction and validation processes can benefit from some degree of intelligent, expert-guided automation that may reduce the manual workload of experts significantly. Moreover, these experiments brought to light some deficiencies in current reaction characterisation, which will certainly promote additional monitoring during the notification and validation processes.

The implementation of this system is expected to improve the quality of future adverse reaction notifications and, most notably, minimise the manual intervention of Haemovigilance experts in the process. In particular, the notifier will be directed to either characterise the adverse reaction in such a way that all “mandatory” features (i.e. those that experts deemed as necessary to ensure the occurrence of such type of reaction) are checked, or consider other possible adverse reactions if such features do not apply. Therefore, this system can be also seen as a training tool for new (and current) hospital notifiers.

With an expected notification rate of 400 adverse reactions per year in Portugal, it will soon be possible to extract a new dataset and use these (semi-) automatically revised notifications to further improve the predictive models. Another future line of work will be the study of current data on the adverse reaction in donors, which surpassed 10,000 records in the last 10 years. While these adverse reactions have different feature profiles, an approach similar to the proposed here holds the potential to improve the quality of that service as well.

Acknowledgements

SING group thanks CITI (Centro de Investigación, Transferencia e Innovación) from the University of Vigo for hosting its IT infrastructure. This work was partially supported by the Consellería de Educación, Universidades e Formación Profesional (Xunta de Galicia) under the scope of the strategic funding of ED431C2018/55-GRC Competitive Reference Group, the Portuguese Foundation for Science and Technology (FCT) under the scope of the strategic funding of UID/BIO/04469/2013 unit and COMPETE 2020 (POCI-01-0145-FEDER-006684).

References

- [1] M.F. Murphy, D.J. Roberts, M.H. Yazer, eds., *Practical Transfusion Medicine*, 5th ed., John Wiley & Sons, Ltd, Chichester, UK, 2017. Doi: 10.1002/9781119129431.
- [2] O. Garraud, C. Sut, A. Haddad, S. Tariket, C. Aloui, S. Laradi, H. Hamzeh-Cognasse, T. Bourlet, F. Zeni, C. Aubron, Y. Ozier, S. Laperche, T. Peyrard, P. Buffet, D. Guyotat, E. Tavernier, F. Cognasse, B. Pozzetto, G. Andreu, Transfusion-associated hazards: A revisit of their presentation, *Transfus. Clin. Biol.* (2018), <https://doi.org/10.1016/j.tracbi.2018.03.002>.
- [3] M. Carlier, M.-P. Vo Mai, L. Fauveau, N. Ounnoughene, I. Sandid, P. Renaudier, Dix-sept ans d'hémovigilance en France: bilan, perspectives, *Transfus. Clin. Biol.* 18 (2011) 140–150, <https://doi.org/10.1016/j.tracbi.2011.02.024>.
- [4] E.M. Wood, A.L. Ang, A. Bisht, P.H. Bolton-Maggs, A.G. Bokhorst, O. Flesland, K. Land, J.C. Wiersum-Osselton, M.R. Schipperus, P. Tiberghien, B.I. Whitaker, International haemovigilance: what have we learned and what do we need to do next?, *Transf. Med.* (2019) 12582, <https://doi.org/10.1111/tme.v29.410.1111/tme.12582>.
- [5] A. Ramoa, J. Condeço, M.A. Escoval, J.-C. Faber, F. Fdez-Riverola, A. Lourenço, The Evolving Role of Information Technology in Haemovigilance Systems, *J. Healthc. Eng.* 2018 (2018) 1–8, <https://doi.org/10.1155/2018/6183468>.
- [6] European Commission, Commission Directive 2005/61/EC of 30 September 2005 implementing Directive 2002/98/EC of the European Parliament and of the Council as regards traceability requirements and notification of serious adverse reactions and events, 2005. <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32005L0061&from=EN>.
- [7] N. Roubinian, S. Kleinman, Building consensus: steps toward standardised haemovigilance reporting, *Lancet. Haematol.* (2019), [https://doi.org/10.1016/S2352-3026\(19\)30081-X](https://doi.org/10.1016/S2352-3026(19)30081-X).
- [8] APPENDIX B: Proposed Standard Definitions for Surveillance of Non Infectious Adverse Transfusion Reactions, in: *Hemovigilance*, John Wiley & Sons, Ltd, 2012: pp. 351–359. Doi: 10.1002/9781118338179.app2.
- [9] CDC, Ncezd, DHQP, National Healthcare Safety Network Biovigilance Component Hemovigilance Module Surveillance Protocol, 2018. www.cdc.gov/nhsn (accessed July 17, 2018).
- [10] G.C. do SPHV, M. Escoval, J. Condeço, A. Ramoa, A. Sousa, C. Caeiro, I. Lobo, I. Miranda, M. Chin, M. Santos, Relatório de Atividade Transfusional e Sistema Português de Hemovigilância 2018, (2019). http://hemovigilancia.net/files/RA_2018_VF1.2.pdf.
- [11] Grupo Coordenador do SPHV, M. Escoval, J. Condeço, A. Ramoa, A. Sousa, I. Lobo, I. Miranda, M. Chin, M. Santos, G. Sousa, Relatório de Atividade Transfusional e Sistema Português de Hemovigilância 2017, 2018.
- [12] J. Baumeister, A. Striffler, Knowledge-driven systems for episodic decision support, *Knowledge-Based Syst.* 88 (2015) 45–56, <https://doi.org/10.1016/j.knsys.2015.08.008>.
- [13] M. Plebani, A. Aita, A. Padoan, L. Sciacovelli, Decision Support and Patient Safety, *Clin. Lab. Med.* 39 (2019) 231–244, <https://doi.org/10.1016/j.cll.2019.01.003>.
- [14] J.M. Baron, D.E. Kurant, A.S. Dighe, Machine Learning and Other Emerging Decision Support Tools, *Clin. Lab. Med.* 39 (2019) 319–331, <https://doi.org/10.1016/j.cll.2019.01.010>.
- [15] E.H. Shortliffe, J.J. Cimino, eds., *Biomedical Informatics*, Springer London, London, 2014. Doi: 10.1007/978-1-4471-4474-8.
- [16] K.H. Hoang, T.B. Ho, Learning and recommending treatments using electronic medical records, *Knowledge-Based Syst.* (2019).
- [17] A. Borah, B. Nath, Identifying risk factors for adverse diseases using dynamic rare association rule mining, *Expert Syst. Appl.* 113 (2018) 233–263, <https://doi.org/10.1016/j.eswa.2018.07.010>.
- [18] P. Beninger, Pharmacovigilance: An Overview, *Clin. Ther.* 40 (2018) 1991–2004, <https://doi.org/10.1016/j.clinthera.2018.07.012>.
- [19] P. Inácio, A. Cavaco, M. Airaksinen, The value of patient reporting to the pharmacovigilance system: a systematic review, *Br. J. Clin. Pharmacol.* 83 (2017) 227–246, <https://doi.org/10.1111/bcp.13098>.
- [20] D. Kopciuch, T. Zaprutko, A. Paczkowska, P. Ratajczak, Ł. Zielińska-Tomczak, K. Kus, E. Nowakowska, Safety of medicines-Pharmacists' knowledge, practice, and attitudes toward pharmacovigilance and adverse drug reactions reporting process, *Pharmacoepidemiol. Drug Saf.* (2019), <https://doi.org/10.1002/pds.4792>.
- [21] C.Y. Lee, Y.-P.P. Chen, Machine learning on adverse drug reactions for pharmacovigilance, *Drug Discov. Today.* (2019), <https://doi.org/10.1016/j.drudis.2019.03.003>.
- [22] P. Funk, Why Hybrid Case-Based Reasoning Will Change the Future of Health Science and Healthcare, in: *ICCB, 2015*.
- [23] C. Marling, S. Montani, I. Bichindaritz, P. Funk, Synergistic case-based reasoning in medical domains, *Expert Syst. Appl.* 41 (2014) 249–259, <https://doi.org/10.1016/j.eswa.2013.05.063>.
- [24] I. Bichindaritz, C. Marling, Case-based reasoning in the health sciences: What's next?, *Artif. Intell. Med.* 36 (2006) 127–135, <https://doi.org/10.1016/j.artmed.2005.10.008>.
- [25] A. Aamodt, E. Plaza, Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches, *AI Commun.* 7 (2001) 39–59.
- [26] S. Begum, M.U. Ahmed, P. Funk, N. Xiong, M. Folke, Case-Based Reasoning Systems in the Health Sciences: A Survey of Recent Trends and Developments, *Trans. Sys. Man Cyber Part C.* 41 (2011) 421–434, <https://doi.org/10.1109/TSMCC.2010.2071862>.
- [27] S. Montani, Exploring new roles for case-based reasoning in heterogeneous AI systems for medical decision support, *Appl. Intell.* 28 (2008) 275–285, <https://doi.org/10.1007/s10489-007-0046-2>.
- [28] B. Cánovas-Segura, A. Morales, J.M. Juárez, M. Campos, F. Palacios, A lightweight acquisition of expert rules for interoperable clinical decision support systems, *Knowledge-Based Syst.* 167 (2019) 98–113, <https://doi.org/10.1016/j.knsys.2019.01.007>.
- [29] B. Rush, L.A. Celi, D.J. Stone, Applying machine learning to continuously monitored physiological data, *J. Clin. Monit. Comput.* (2018), <https://doi.org/10.1007/s10877-018-0219-z>.
- [30] S. Zolhavarieh, D. Parry, KQA: A Knowledge Quality Assessment Model for Clinical Decision Support Systems, *Stud. Health Technol. Inform.* 245 (2017) 983–986. <http://www.ncbi.nlm.nih.gov/pubmed/29295247>.
- [31] Y.-F. Zhang, Y. Tian, T.-S. Zhou, K. Araki, J.-S. Li, Integrating HL7 RIM and ontology for unified knowledge and data representation in clinical decision support systems, *Comput. Methods Programs Biomed.* 123 (2016) 94–108, <https://doi.org/10.1016/j.cmpb.2015.09.020>.
- [32] Grupo Coordenador do SPHV, Definições padronizadas para a vigilância de reacções transfusionais não infecciosas, 2011. http://hemovigilancia.net/files/Portuguese_ISBT_definitions_jul_2011-verso_01.pdf (accessed July 17, 2018).
- [33] A.D. Sagar, A. Najam, The human development index: a critical review, *Ecol. Econ.* 25 (1998) 249–264, [https://doi.org/10.1016/S0921-8009\(97\)00168-7](https://doi.org/10.1016/S0921-8009(97)00168-7).
- [34] Y. Bengio, Y. Grandvalet, No unbiased estimator of the variance of k-fold cross-validation, *J. Mach. Learn. Res.* 5 (2004) 1089–1105.
- [35] C.E. McCulloch, Generalized linear models, *J. Am. Stat. Assoc.* 95 (2000) 1320–1324.
- [36] G. Shi, Chapter 5 – Decision Trees, in: G. Shi (Ed.), *Data Min. Knowl. Discov.* Geosci, Elsevier, Oxford, 2014, pp. 111–138, <https://doi.org/10.1016/B978-0-12-410437-2.00005-9>.
- [37] R.G. Cowell, P. Dawid, S.L. Lauritzen, D.J. Spiegelhalter, *Probabilistic Networks and Expert Systems*, Springer-Verlag, New York, 1999, doi: 10.1007/b97670.

- [38] M.A. Hearst, S.T. Dumais, E. Osuna, J. Platt, B. Scholkopf, Support vector machines, *IEEE Intell. Syst. Their Appl.* 13 (1998) 18–28.
- [39] C. Cortes, V. Vapnik, Support-vector networks, *Mach. Learn.* 20 (1995) 273–297, <https://doi.org/10.1007/BF00994018>.
- [40] L. Breiman, Random Forests, *Mach. Learn.* 45 (2001) 5–32, <https://doi.org/10.1023/A:1010933404324>.
- [41] I. Mukherjee, C. Rudin, R.E. Schapire, The Rate of Convergence of AdaBoost, *J. Mach. Learn. Res.* 14 (2013) 2315–2347, <http://jmlr.org/papers/v14/mukherjee13b.html>.
- [42] Y. Freund, R.E. Schapire, Experiments with a New Boosting Algorithm, in: Proc. Thirteenth. Int. Conf. Mach. Learn., Morgan Kaufmann, 1996: pp. 148–156.
- [43] D.L. Shrestha, D.P. Solomatine, Experiments with AdaBoost.RT, an Improved Boosting Scheme for Regression, *Neural Comput.* 18 (2006) 1678–1710, <https://doi.org/10.1162/neco.2006.18.7.1678>.
- [44] D.F. Specht, Probabilistic neural networks, *Neural Networks.* 3 (1990) 109–118, [https://doi.org/10.1016/0893-6080\(90\)90049-Q](https://doi.org/10.1016/0893-6080(90)90049-Q).
- [45] M.N. Murty, V.S. Devi, Nearest Neighbour Based Classifiers, in: Pattern Recognit. An Algorithmic Approach, Springer London, London, 2011: pp. 48–85. Doi: 10.1007/978-0-85729-495-1_3.
- [46] H. Larochelle, D. Erhan, A. Courville, J. Bergstra, Y. Bengio, An Empirical Evaluation of Deep Architectures on Problems with Many Factors of Variation, in: Proc. 24th Int. Conf. Mach. Learn., ACM, New York, NY, USA, 2007: pp. 473–480. Doi: 10.1145/1273496.1273556.
- [47] J. Bergstra, R. Bardenet, Y. Bengio, B. Kégl, Algorithms for Hyper-Parameter Optimization, in: J. Shawe-Taylor, R.S. Zemel, P. Bartlett, F. Pereira, K.Q. Weinberger (Eds.), 25th Annu. Conf. Neural Inf. Process. Syst. (NIPS 2011), Neural Information Processing Systems Foundation, Granada, Spain, 2011. <https://hal.inria.fr/hal-00642998>.
- [48] G. Hripcsak, A.S. Rothschild, Agreement, the f-measure, and reliability in information retrieval, *J. Am. Med. Inform. Assoc.* 12 (2005) 296–298, <https://doi.org/10.1197/jamia.M1733>.
- [49] R. Markus, Hofmann Klinkenberg, *RapidMiner: Data Mining Use Cases and Business Analytics Applications*, CRC Press Taylor & Francis Group, 2014.
- [50] R.B. Klanderman, J.J. Bosboom, Y. Migdady, D.P. Veelo, B.F. Geerts, M.F. Murphy, A.P.J. Vlaar, Transfusion-associated circulatory overload—a systematic review of diagnostic biomarkers, *Transfusion* 59 (2019) 795–805, <https://doi.org/10.1111/trf.15068>.
- [51] M.S. Cho, S. Sharma, Transfusion-related Acute Lung Injury (TRALI), in: Treasure Island (FL), 2019.
- [52] P.H.B. Bolton-Maggs, Conference report: International Haemovigilance Seminar and the SHOT Annual Symposium, 10–12 July 2018, *Transfus. Med.* (2018), <https://doi.org/10.1111/tme.12569>.
- [53] A.P.J. Vlaar, P. Toy, M. Fung, M.R. Looney, N.P. Juffermans, J. Bux, P. Bolton-Maggs, A.L. Peters, C.C. Silliman, D.J. Kor, S. Kleinman, A consensus redefinition of transfusion-related acute lung injury, *Transfusion.* (2019), <https://doi.org/10.1111/trf.15311>.



Jorge Manuel Condeço Ribeiro is a PhD graduate of the Department of Computer Science at the University of Vigo and Director of the Porto Centre of Blood And Transplantation - Portuguese Blood and Transplantation Institute. His main research interests include hemovigilance, medical statistics, BPMN and decision support systems.



Florentino Fdez-Riverola received the BS degree in computer science from University of Oviedo. He received the MS and PhD degrees from University of Vigo, where he is Professor in the Department of Computer Science and Coordinator of the Next Generation Computer System group (<http://sing-group.org/>). He is author of 210 scientific contributions, of which more than one hundred were published in journals indexed in the SCI. His publications hold 1581 citations from 1227 different documents.



Anália Maria Garcia Lourenço is a faculty member of the Department of Computer Science and a researcher affiliated to the Centre of Biomedical Research Centre (CINBIO), at the University of Vigo and the Centre of Biological Engineering, at the University of Minho. Her main research interests include data science, machine learning, computational intelligence, bioinformatics and systems biology.



Augusto Manuel Oliveira Ramoa Rodrigues is a PhD student of the Department of Computer Science at the University of Vigo and web applications developer at the Portuguese Blood and Transplantation Institute in Porto, Portugal. His main research interests include bioinformatics, hemovigilance, ETL and decision support systems.