Expert Systems with Applications 41 (2014) 831–840

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Evolutionary computation in the identification of risk factors. Case of TRALI

María Dolores Torres [a,*], Aurora Torres [a], Felipe Cuellar [b], María de la Luz Torres [b], Eunice Ponce de León [a], Francisco Pinales [a]

[a] Universidad Autónoma de Aguascalientes, Av. Universidad # 940, Ciudad Universitaria, C.P. 20131 Aguascalientes, Ags, Mexico
[b] Centenario Hospital Miguel Hidalgo, Galeana Sur # 465, Col. Obraje, C.P. 20230 Aguascalientes, Ags, Mexico

ARTICLE INFO

Keywords:
Genetic algorithm
Evolutionary Hybrid Algorithm
TRALI's risk factors
Testor
Typical testor
Informational weight

ABSTRACT

This paper presents the use of an evolutionary algorithm hybridized with the concepts of testor and typical testor in determining factors associated with transfusion related acute lung injury (TRALI). Although nowadays many cases of this syndrome remain ignored or misdiagnosed, this is the leading cause of morbidity and mortality related to transfusion in the United States.

This research was conducted with data from 174 cases collected in the Centenary Hospital Miguel Hidalgo in the city of Aguascalientes, Mexico, in the period 2007 to 2010.

The proposed algorithm works with information from the model known as "two hits", in which the first hit is the original disease and the second corresponds to the blood transfusion. This algorithm was strengthened with mechanisms that let it do an efficient search in the whole solution space. In addition to the calculation of the informational weight, the algorithm also establishes the cutoff point that determines the variables that impact the most.

From the results given by the algorithm and the cutoff proposed by the medical staff, a strategy for the treatment of patients that should be transfused was proposed.

This study confirmed some of the risk factors previously reported in the literature, and also made an interesting discovery.

1. Introduction

The influence that a group of variables has on a response variable is a common topic in several areas. When the response variable is numeric, it is feasible to apply tools such as multiple regressions; however, when the response variable is dichotomous or polytomous, the problem must be addressed by other techniques, such as logistic regression. Regardless of the technique used to determine the effect of a group of variables over another, this process is known as risk factors determination.

Identifying risk factors for any health disorder is an area in which researchers have invested many hours. The identification of risk factors is important because it allows physicians to avoid or quickly identify the presence of any adverse medical condition. By having the opportunity to know in advance that a serious condition can occur in a patient, health experts have precious time to establish treatment or actions that could reduce its impact.

This paper describes the application of a hybrid genetic algorithm in the determination of informational weight of variables related to "transfusion related acute lung injury" (TRALI). The used TRALI's etiology in this work is a model known as "two hits"; consisted of two events: the first of them is related to the clinical picture of the patient (first hit), and the second is the transfusion of blood products (second hit) (Rodríguez-Moyado, 2011a). The determination of informational weight implies the identification of TRALI's risk factors and the establishment of an assessment to each variable.

The use of conventional statistical techniques (binomial or multinomial logistic regression), may produce unsatisfactory results when analyzed independent variables are highly correlated (McGee, Reed & Yano, 1984). Another aspect that should be considered, in order that logistic regression makes sense, is that monotonous relationship should exist between the explanatory variables and the response variable (Domínguez & Aldana, 2001). Also according to Hsieh (Hsieh, 1989), the number of subjects needed to use logistic regression smoothly, must be greater than $10(k+1)$ where k is the number of explanatory variables. This apparently innocuous restriction makes impossible the analysis of rare syndromes by these techniques. Due to this, it is desirable

* Corresponding author. Tel.: +52 449 9107417.
E-mail addresses: mdtorres@correo.uaa.mx (M.D. Torres), atorres@correo.uaa.mx (A. Torres), cuellarluis@hotmail.com (F. Cuellar), tishats@hotmail.com (María de la Luz Torres), eponce@correo.uaa.mx (Eunice Ponce de León), fjpinales@correo.uaa.mx (F. Pinales).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.08.013

to develop other techniques to quantify the importance of the relationship between each predictor variable and the dependent variable, without having to meet the mentioned requirements.

Artificial intelligence has provided several tools to make efficient use or interpretation of data, to different areas; without the need to have expert knowledge of the phenomenon under study. With the use of metaheuristics, countless complex problems have been solved efficiently. This paper describes the process and tools used for informational weight determination of TRALI related variables using a metaheuristic hybridized with the concepts of testor and typical testor. A testor is a set of features capable of distinguishing between two classes (may be more); and a typical testor is a minimal testor. These concepts are presented in Section 3.

Because the determination of all typical testors immersed in a basic matrix is a problem of exponential complexity according to Sanchez and Lazo (2008), the use of approximate solution techniques such as genetic algorithm is perfectly justified. Specifically, the approached problem was initially described by 31 variables, so the amount of possible subsets of variables is 2,147,483,647. The single generation of this many subgroups, even without assessing their ability to describe the syndrome, would take a long time and would be very computational resources consuming. As Torres showed (Torres, 2010), this problem is so hard that some metaheuristics without additional information can be executing for several days and find no one typical testor. In this work, Torres used a basic matrix of 32,768 rows and 29 features.

To better understand the application presented in this paper, Section 2 provides a brief description of the transfusion related acute lung injury and justification of the importance of its early identification. In Section 3, the concepts of typical testor, testor and informational weight are described. Section 4 presents main concepts related with hybrid metaheuristics, whereas Section 5 discusses in detail the used methodology for determining the informational weight of TRALI related variables. Later, in Section 6 it is described the implemented algorithm and operators involved on it. Later, in Section 7 the results and discussion are presented. Finally, in Section 8 the conclusions drawn from this research are discussed.

2. Transfusion related acute lung injury (TRALI)

This paper addresses the identification of risk factors for TRALI with preventive and early identification purposes. This syndrome is characterized by the development of acute respiratory failure and pulmonary edema within 6 h after transfusion, with a wide clinical spectrum of presentation. The term "transfusion-related acute lung injury" was coined in 1985 by Popovsky & Moore (1985).

According to Añon et al., although this is an under-diagnosed and under-reported medical condition, this syndrome is considered the leading cause of transfusion-related death in the United States, and the second one in the United Kingdom (Añón, García, Quintana, González & Bruscas, 2010).

Nowadays many cases of TRALI remain ignored or are misdiagnosed as fluid overload or acute lung injury (ALI) of other etiology according Vlaar et al. (2011), so a strong motivation for the study of this condition is that physicians consider its relationship with transfusion. As stipulated in Rodriguez-Moyado (2011a), doctors of transfusion medicine services must establish a monitoring of patients in surgical and intensive care units, to identify which component is related to the respiratory distress syndrome, identify the donors involved and clarify whether the blood component contains leukocyte antibodies.

The low level of diagnosis of this medical condition, combined with its seriousness, makes it a syndrome that should be avoided, as Cuellar stated (Cuellar, 2012). In the presence of prone patients to this clinical condition, the following actions are recommended: employing blood transfusion components leukocytereduced, with less than 72 h of storage and irradiated; for red blood cells, washed units are recommended (Rodriguez-Moyado, 2011b).

Since its initial description in Barnard (1951) non-cardiogenic lung edema related to transfusion has been reported widely using different names, including non-cardiogenic pulmonary edema, pulmonary hypersensitivity and severe allergic pulmonary edema.

Today, this syndrome has gone from being an almost unknown side effect of transfusion, to the leading cause of transfusion related morbidity and mortality.

The formal definition of this medical condition was developed by the National Heart Lung and Blood Institute Working Group on TRALI US in Toy et al. (2005), and is described as new acute lung injury (ALI)/acute respiratory distress syndrome (ARDS) occurring during or within six hours after blood product administration.

Although the absence of specific disease markers and diagnostic tests has resulted in a large variation in estimations of incidence, TRALI is generally considered to be a rare event.

Some of the most important elements of the TRALI's definition according to medical experts are (Kleinmann et al., 2004):

1. Sudden onset of acute lung injury (ALI).
2. Hypoxemia (PaO2/FiO2 (ratio of partial pressure of arterial O2 to the fraction of inspired O2) $\leqslant$ 300 and must be adjusted downward with increasing altitude or SpO2 $\leqslant$ 90% on room air or other clinical evidence).
3. Bilateral lung infiltrates on frontal chest radiograph.
4. No evidence of left atrial hypertension (i.e., transfusion-associated circulatory overload).
5. Occurrence during or within 6 h after completion of transfusion.
6. No temporal relationship to an alternative risk factor for ALI.
7. New ALI and no other ALI risk factors present including aspiration, multiple trauma, pneumonia, cardiopulmonary bypass, burn injury, toxic inhalation, lung contusion, acute pancreatitis, drug overdose, near drowning, shock and sepsis.
8. If one or more ALI risk factors are present, possible TRALI should be diagnosed (in patients with an alternative ALI risk factor, TRALI is still possible).

There are different types of blood components, however according to Rodríguez-Moyado (2011a), fresh frozen plasma (FFP) figures prominently as a cause of immediate adverse transfusion effects. Moreover, according to Pita and Ramírez (1999) it has been demonstrated abuse of PFC transfusion in Mexico at a rate of 96.23% of 229 cases in a general hospital. According to the above-mentioned, it is necessary to establish awareness programs and access to information that enables health specialists, to predict the presence of these alterations and even avoid them.

There are two proposed etiologies for this syndrome: the first is an immune mediated episode produced by transfused antibodies directed toward Human Leukocyte Antigens (HLA) or Human Neutrophil Antigens (HNA). The second, is a model consisted of two events: the first of them is related to the clinical picture of the patient (first hit), and the second one is the transfusion of blood products (second hit). This model commonly known as "two hits" was taken as the basis for the classification of the studied variables.

For this reason, we have grouped the variables as first hit variables, second hit variables and other variables.

Because each of the two hits occurs at different time, we propose that health specialists assess first hit variables highly related to TRALI (according with obtained results) in a first moment, and if these variables are presented, then they have to try to avoid transfusion. If transfusion is unavoidable, then it has to be