A case-based reasoning view of thrombophilia risk

João Vilhena [a], Henrique Vicente [a,e], M. Rosário Martins [b], José M. Grañeda [c], Filomena Caldeira [c], Rodrigo Gusmão [c], João Neves [d], José Neves [e,*]

[a] Departamento de Química, Escola de Ciências e Tecnologia, Universidade de Évora, Évora, Portugal
[b] Departamento de Química, Escola de Ciências e Tecnologia, Laboratório HERCULES, Universidade de Évora, Évora, Portugal
[c] Serviço de Patologia Clínica do Hospital do Espírito Santo de Évora EPE, Évora, Portugal
[d] Drs. Nicolas & Asp, Dubai, United Arab Emirates
[e] Centro Algoritmi, Universidade do Minho, Braga, Portugal

ARTICLE INFO

ABSTRACT

Thrombophilia stands for a genetic or an acquired tendency to hypercoagulable states that increase the risk of venous and arterial thromboses. Indeed, venous thromboembolism is often a chronic illness, mainly in deep venous thrombosis and pulmonary embolism, requiring lifelong prevention strategies. Therefore, it is crucial to identify the cause of the disease, the most appropriate treatment, the length of treatment or prevent a thrombotic recurrence. Thus, this work will focus on the development of a diagnosis decision support system in terms of a formal agenda built on a logic programming approach to knowledge representation and reasoning, complemented with a case-based approach to computing. The proposed model has been quite accurate in the assessment of thrombophilia predisposition risk, since the overall accuracy is higher than 90% and sensitivity ranging in the interval [86.5%, 88.1%]. The main strength of the proposed solution is the ability to deal explicitly with incomplete, unknown, or even self-contradictory information.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Thrombophilia described genetic or acquired tendency to hypercoagulable states that increase the risk of venous thrombosis and, in some cases arterial thrombosis [1]. Venous ThromboEmbolism (VTE) is the most common vascular disease after acute myocardial infarction and stroke. It is represented by two main clinical events: Deep Venous Thrombosis (DVT) and Pulmonary Embolism (PE), which often constitute a unique clinical picture in which PE follows DVT [2]. Primary hypercoagulable states are generally inherited abnormalities of coagulation such as antithrombin III deficiency, protein C and protein S deficiency. High risk thrombophilia includes natural coagulation inhibitors (antithrombin deficiency, protein C deficiency, protein S deficiency and homozygosity for factor V Leiden or the prothrombin G20210A gene mutation, compound heterozygosity for factor V Leiden and the prothrombin gene mutation) [3,4]. Secondary hypercoagulable states are frequently acquired disorders in patients with underlying systemic diseases or clinical conditions, known to be associated with an increased risk of thrombosis (e.g., malignancy, pregnancy, use of oral contraceptives, myeloproliferative disorders, hyperlipidemia, diabetes mellitus and abnormalities of blood vessels). The most common acquired thrombophilia is antiphospholipid antibody syndrome [5]. Furthermore, the long distance travel is often associated with an up to 4-fold increased risk of VTE compared to non-travelers [6]. Air pollution was also associated to the hypercoagulable states in patients with DVT [7].

The incidence of VTE is estimated at 56–160/100.000 people/year [8], and strongly age-dependent, rising nearly 1% per year in old age [9]. Studies show that about 70% of patients presenting a first episode of VTE are over 60 years old, and the rate of recurrence is higher when the first episode of VTE occurs before 60 years old [10]. Patients with VTE have laboratorial abnormalities and clinical conditions that are associated with an increased risk of thrombosis (prethrombotic states) or have recurrent thrombosis without recognizable predisposing factors (thrombosis-prone). The stated above highlights how it is important to assess the patient's clinical probability for VTE. In this context, the Wells score has been commonly used, namely in patients with suspected DVT or PE [11]. Based on Wells score the physician can order the D-dimer test as a marker of endogenous fibrinolysis [11]. Follow-up of patients for prolonged periods after an initial DVT or PE has revealed a startling fact, i.e., VTE often is a chronic illness requiring lifelong prevention strategies. Thus, it

is crucial to identify the cause, the most appropriate treatment, how long the treatment should be or how to prevent a thrombotic recurrence [12]. In the case of patients with a high risk of clinical complications preventive actions should be taken. The screening for thrombophilia is mandatory in three high-risk groups, namely women who are prescribed oral oestrogen preparations, pregnancy and the puerperium and patients undergoing major orthopaedic surgery [13]. In addition, a significant number of arterial and venous thrombotic episodes, especially among young individuals, occur without a plausible explanation [2]. So, more studies are necessary to reach a correct identification of factors associated with these diseases in order to assess the individual risk of thrombosis, and promote more targeted prophylactic and therapeutic alternatives.

In this study the complex pathophysiologic features of these hypercoagulable states (i.e., genetic, environmental and acquired risk factors) are discussed, and the problem was tackled with Artificial Intelligence (AI) based methodologies and techniques for problem solving. Thus, this work will focus on the development of an AI grounded Decision Support System aiming the early diagnosis of thrombophilia and signalize patients with hypercoagulable states. The computational framework was built on top of a Logic Programming Case Base approach to knowledge representation and reasoning, which caters for the handling of incomplete, unknown, or even self-contradictory information. Clustering methods centered on an analysis of attribute's similarities were used to distinguish and aggregate historical data according to the context under which it was added to the Case Base, therefore enhancing the prediction process.

2. Background

2.1. Case based reasoning

A Case Based Reasoning (CBR) methodology for problem solving stands for an act of finding and justifying a solution to a given problem based on the consideration of similar past ones, by reprocessing and/or adapting their data or knowledge [21,22]. In CBR – the cases – are stored in a Case Base, and those cases that are similar (or close) to a new one are used in the problem solving process. There are several examples on the literature concerning the use of CBR as a problem-solving methodology to be used in problems in Medicine. Different researchers have reviewed more than thirty CBR systems/projects [23,24] revealing that CBRs have been widely employed in the medical domain, including disease diagnosis, classification, treatment and management. Indeed, a CBR approach used in mental health problems, in order to predict the effect of treatments of patients with anxiety disorders showed 65% of correct predictions in the absence of similarity restrictions, while for scenario with similarity restrictions (i.e., under the condition that the prediction was based only on cases with a similarity of at least 0.62), the accuracy increased to 80% [25]. Another study presents a fuzzy ontology-based semantic CBR system for a decision support system to answer complex medical queries related to semantic understanding of medical concepts and handling of vague terms in diabetes diagnosis. The proposed system exhibits an overall accuracy of 97.7% higher than the accuracies obtained with other artificial intelligence based tools like the k-nearest neighbour, with $k = 3$ (68.3%), decision trees (90.0%), support vector machines (76.7%), Bayesian classifier (76.7%) and artificial neural networks (71.7%) [26]. A study of [27] combines CBR and multi-agent systems. The multi-agent architecture aims to take into account the whole cycle of clinical decision-making, i.e., adaptable to different medical aspects like the diagnosis, prognosis, treatment and therapeutic monitoring of gastric cancer. In the multi-agent architecture, the ontological agent type uses the knowledge domain in order to ensure proficiency in the extraction of similar clinical cases and provide treatment suggestions to patients and physicians. CBR, in turn, is used to memorize and to restore experience data aiming to solve similar problems [27].
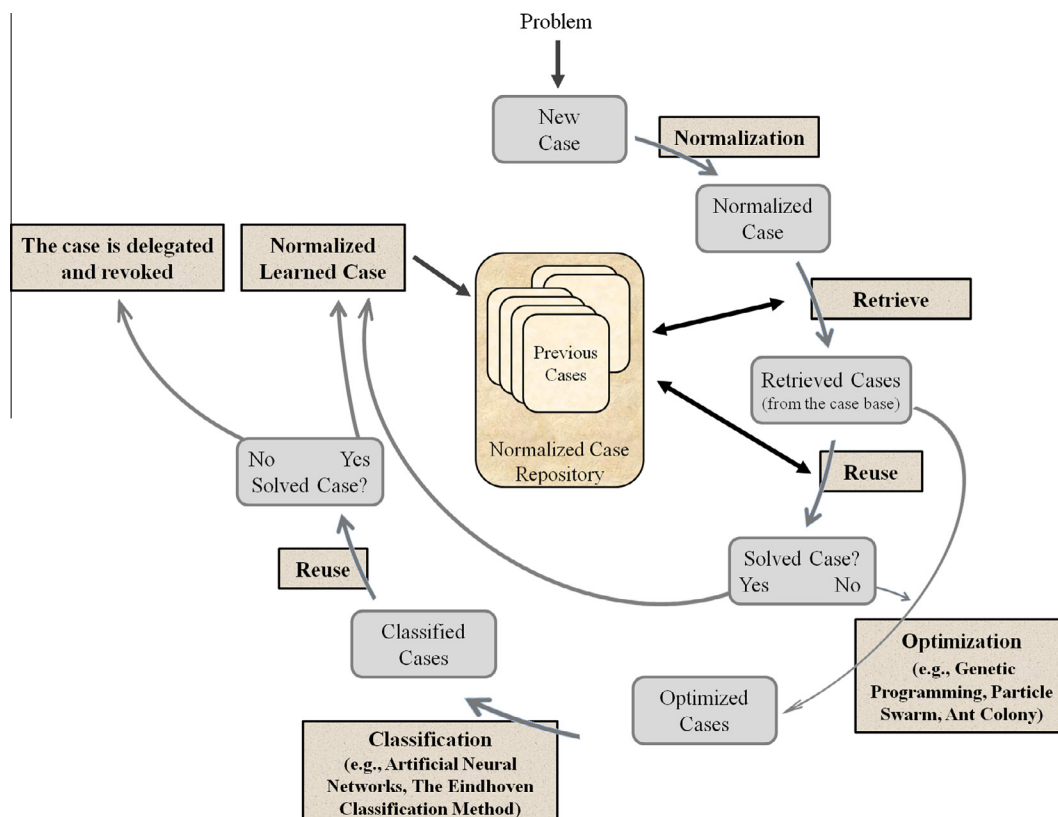

Fig. 1. An extended view of the CBR cycle.