Keywords: Thrombophilia Risk Evaluation, Knowledge Representation and Reasoning, Logic Programming, Artificial Neural Networks.

Abstract: Thrombotic disorders have severe consequences for the patients and for the society in general, being one of the main causes of death. These facts reveal that it is extremely important to be preventive; being aware of how probable is to have that kind of syndrome. Indeed, this work will focus on the development of a decision support system that will cater for an individual risk evaluation with respect to the surge of thrombotic complaints. The Knowledge Representation and Reasoning procedures used will be based on an extension to the Logic Programming language, allowing the handling of incomplete and/or default data. The computational framework in place will be centered on Artificial Neural Networks.

1 INTRODUCTION

Thrombophilia or Venous ThromboEmbolism (VTE) may be defined as an increased tendency towards hypercoagulability and venous thrombosis, i.e., it refers to a predisposition to thromboembolism (Favaloro et al. 2009). Thrombophilia is a common clinical condition with high morbidity and mortality, comprising Deep-Vein Thrombosis (DVT) and Pulmonary Embolism (PE) (Cohen et al. 2007). The incidence of VTE is estimated at 56-160 per 100,000 people/year (East and Wakefield, 2010). VTE is a multifactorial disease and these risks are generally distinguished as either heritable or acquired, although sometimes this distinction is unclear (Rosendaal, 1999; Favaloro et al., 2009).

Venous thrombosis could be correlated with some genetic defects, namely mutations that result in deficiency of natural coagulation inhibitors, as well as mutations with increased level/function of coagulation factors (Reitsma and Rosendaal, 2007). Inherited risk factors include deficiencies/defects in natural anticoagulants, such as antithrombin, protein C and protein S (Mondal et al. 2010; Cafolla et al., 2011), and genetic polymorphisms such as prothrombin G20210A and factor V Leiden (Reitsma and Rosendaal, 2007), that lead to a condition designated as activated protein C resistance (Agrawal et al. 2009). Inherited AntiThrombin (AT) deficiency is an uncommon autosomal dominant disorder. Most cases remain heterozygous. Homozygosity for AT deficiency is rare and is almost always fatal in utero. Protein C (PC) deficiency is an autosomal dominant inherited disorder associated with spontaneous and recurrent thrombotic events. Patients with protein C and S deficiency are at increased risk for venous thromboembolic disease, occasional arterial thrombosis (Mondal et al., 2010). Factor V Leiden (FVL) increases the risk of thrombosis in PC-deficient type I families (Cafolla et al. 2011). Other mutations or polymorphisms associated with increased risk of thrombosis are methylenetetra-hydrofolate reductase 677C (Rosendaal, 1999).

Acquired thrombophilia risk factors include antiphospholipid antibodies, detected as lupus anticoagulants and/or anticardiolipin antibodies and/or anti-beta-2-glycoprotein-I antibodies. Environmental
or acquired thrombophilia risk factors include also previous history or concomitant disease, age, immobility, surgery, obesity, smoke, cancer hormone use, and pregnancy or postpartum states (Rosendaal, 1999; Heit et al. 2002; Goldhaber, 2010).

This work reports the founding of a computational framework that uses knowledge representation and reasoning techniques to set the structure of the information and the associate inference mechanisms. We will centre on a Logic Programming (LP) based approach to knowledge representation and reasoning (Neves. 1984; Neves et al. 2007), complemented with a computational framework based on Artificial Neural Networks (ANNs) (Cortez et al., 2004).

ANNs are computational tools which attempt to simulate the architecture and internal operational features of the human brain and nervous system. ANNs can be defined as a connected structure of basic computation units, called artificial neurons or nodes, with learning capabilities. Multilayered feed-forward neural network architecture is one of the most popular ANNs structure often used for prediction as well as for classification. This architecture is molded on three or more layers of artificial neurons, including an input layer, an output layer and a number of hidden layers with a certain number of active neurons connected by modifiable weights. In addition, there is also a bias, which is only connected to neurons in the hidden and output layers. The number of nodes in the input layer sets the number of independent variables, and the number of nodes in output layer denotes the number of dependent variables (Haykin, 2008).

Several studies have shown how ANNs could be successfully used to model data and capture complex relationships between inputs and outputs (Caldeira et al., 2011; Vicente et al., 2012; Salvador et al., 2013).

With this paper we make a start on the development of a diagnosis assistance system for thrombophilia risk detection using LP complemented with ANNs.

2 RELATED WORK

Many studies presenting the concept of uncertainty and/or "imperfect data" like Hunter (1999) and Zhang and Goodchild (2002) shows that there is an emergent interest in the problem of uncertainty as compared to accuracy or error in data. The notion of uncertainty is broader than error or accuracy and includes these more restricted concepts. While accuracy is the closeness of measurements or computations to their “true” value or some value agreed to be the “truth”, uncertainty can be considered any aspect of the data that results in less than perfect knowledge about the phenomena being studied (Hong et al., 2014). On the one hand, it is consensual that when the data are uncertain, it is need a different representation and uncertainty can be reduced by “acquiring additional information or improving the quality of the information available” (Hunter, 1999), i.e., in almost all decisions that one may take, the information is not always exact, but indeed imperfect, in the sense that we handle estimated values, probabilistic measures, or degrees of uncertainty. On the other hand, knowledge and belief are generally incomplete, contradictory, or even error sensitive, being desirable to use formal tools to deal with the problems that arise from the use of partial, contradictory, ambiguous, imperfect, nebulous, or missing information (Neves, 1984; Neves et al., 2007; Hong et al., 2014). Some general models have been presented where uncertainty is associated to the application of Probability Theory (Li et al., 2007), Fuzzy Set Theory (Schneider, 1999), Similarities (Freire et al., 2002; Liao, 2005). Other approaches for knowledge representation and reasoning have been proposed using the Logic Programming paradigm, namely in the area of Model Theory (Gelfond and Lifschitz, 1988; Kakas et al., 1998; Pereira and Anh, 2009) and Proof Theory (Neves, 1984; Neves et al., 2007). The evaluation of knowledge that stems out from logic programs becomes a point of research. In this sense, the evaluation of knowledge that stems out from logic programs becomes a point of research. Lucas (2003) and Hommerson (2008) work is a good example of quality evaluation using logic. The author used abduction and temporal logic for quality checking of medical guidelines, proposing a method to diagnose potential problems in a guideline, regarding the fulfillment of general medical quality criteria at a meta-level characterization. They explored an approach, which uses a relational translation to map the temporal logic formulas to first-order logic and a resolution-based theorem prover (Schneider, 1999). In another research line, the Quality-of-Information concept (QoI) (Lucas, 2003; Machado et al., 2010) demonstrated their applicability in dynamic environments and for decision-making purposes. The objective is to built a quantification process of the QoI and an assessment of the argument values of a given predicate with relation to their domains (here understood as Degree-of-Confidence (DoC)), that stems from a logic program or theory during the evolution process when searching for solutions in order to solve a problem in environments with default data. Our main contribution relies on the fact that at
the end, the extensions of the predicates that make the universe of discourse are given in terms of DoCs predicates that stand for one’s confidence that the predicates arguments values fit into their respective domains. This approach potentiates the use of diverse computational paradigms, like Case Based Reasoning (Carneiro et al., 2013), Artificial Neural Networks (Vicente et al., 2012; Salvador et al., 2013), Particle Swarm (Mendes et al., 2004), just to name a few. It also incapsulates, in itself, a new vision of Multi-value Logics, once a proof of a theorem in a conventional way, is evaluated to the interval \([0,1]\). Indeed, some interesting results have been obtained, namely in the fields of Coronary Risk Evaluation (Rodrigues et al., 2014), Hyperactivity Disorder (Pereira et al., 2014) and Length of Hospital Stay (Abelha et al., 2014) among others.

3 KNOWLEDGE REPRESENTATION AND REASONING

We follow the proof theoretical approach and an extension to the Logic Programming (LP) language, to knowledge representations and reasoning. An Extended Logic Program (ELP) is a finite set of clauses in the form:

\[ p \leftarrow p_1, \ldots, p_n, \neg q_1, \ldots, \neg q_m \] (1)

\[ \neg (p_1, \ldots, p_m, \neg q_1, \ldots, \neg q_m) \quad (n, m \geq 0) \] (2)

where \( p \) is a domain atom denoting falsity, the \( p_i, q_j \), and \( p \) are classical ground literals, i.e., either positive atoms or atoms preceded by the classical negation sign \( \neg \) (Neves, 1984). Under this representation formalism, every program is associated with a set of abducibles (Kakas et al. 1998; Pereira and Anh, 2009) given here in the form of exceptions to the predicates arguments values fit into their respective domains. This approach potentiate the use of diverse computational paradigms, like Case Based Reasoning (Carneiro et al., 2013), Artificial Neural Networks (Vicente et al., 2012; Salvador et al., 2013), Particle Swarm (Mendes et al., 2004), just to name a few. It also incapsulates, in itself, a new vision of Multi-value Logics, once a proof of a theorem in a conventional way, is evaluated to the interval \([0,1]\). Indeed, some interesting results have been obtained, namely in the fields of Coronary Risk Evaluation (Rodrigues et al., 2014), Hyperactivity Disorder (Pereira et al., 2014) and Length of Hospital Stay (Abelha et al., 2014) among others.

\[ \text{QoI}_i = \frac{1}{\text{Card}} \] (4)

where Card denotes the cardinality of the abducibles set for \( i \), if the abducibles set is disjoint. If the abducibles set is not disjoint, the QoI is given by:

\[ \text{QoI}_i = \frac{1}{\text{Card}_i} = \frac{1}{c_i^1 + \cdots + c_i^{\text{Card}_i}} \] (5)

where \( c_i^j \) is a card-combination subset, with Card elements. The next element of the model to be considered is the relative importance that a predicate assigns to each of its attributes under observation, i.e., \( w_i^k \), which stands for the relevance of attribute \( k \) in the extension of \( \text{predicate}_i \). It is also assumed that the weights of all the attribute predicates are normalized, i.e.:

\[ \sum_{1 \leq k \leq n} w_i^k = 1, \forall_i \] (6)

allowing one to set:

\[ \text{predicate}_i(x_1, \ldots, x_n) \iff V_i(x) \] (8)

It is now possible to engender the universe of discourse, according to the information given in the logic programs that endorse the information about the problem under consideration, according to productions of the type:

\[ \text{predicate}_i - \bigcup_{1 \leq i \leq m} \text{clause}_{ij}(x_1, \ldots, x_n) \iff \text{QoI}_i \iff \text{DoC}_i \] (9)

where \( U \) and \( m \) stand, respectively, for “set union” and the cardinality of the extension of \( \text{predicate}_i \). On the other hand, \( \text{DoC}_i \) denotes one’s confidence on the attribute’s values of a particular term of the extension of \( \text{predicate}_i \), whose evaluation will be illustrated below. In order to advance with a broad-spectrum, let us suppose that the Universe of Discourse is described by the extension of the predicates:

\[ f_1(\cdots), f_2(\cdots), \ldots, f_k(\cdots) \quad \text{where} \quad (n \geq 0) \] (10)

Assuming we have a clause that is mapped into a case, that clause has as argument all the attributes that make the case. The argument values may be of
the type unknown or members of a set, may be in the scope of a given interval or may qualify a particular observation. Let us consider the following clause where the second argument value may fit into the interval $[3,5]$ with a domain of $[0,8]$, the value of the third argument is unknown, which is represented by the symbol $\Delta$, with a domain that ranges in the interval $[5,15]$, and the first argument stands for itself, with a domain that ranges in the interval $[0,3]$. Let us consider that the case data is given by the extension of predicate $f_1$, given in the form:

$$f_1: x_1, x_2, x_3 \rightarrow \{0,1\}$$

(1)

where “$\times$” and “$\triangleright$” is one’s notation for sets, where “0” and “1” denote, respectively, the truth values “false” and “true”. One may have:

$$\neg f_1(x_1, x_2, x_3) \leftarrow not f_1(x_1, x_2, x_3)$$

where the $Y_i$ stand for themselves.

\[
\neg f_1(x_1, x_2, x_3) \leftarrow not f_1(x_1, x_2, x_3)
\]

$$f_1 = \begin{bmatrix} 2 & 0 & 2 & 0 \\ 3 & 0 & 3 & 0 \\ \end{bmatrix}, x_2 = \begin{bmatrix} 3 & 0 & 5 & 0 \\ 8 & 0 & 8 & 0 \\ \end{bmatrix}, x_3 = \begin{bmatrix} 5 & 5 & 15 & 5 \\ 15 & 5 & 15 & 5 \\ \end{bmatrix}
$$

Once the clauses or terms of the extension of the predicate are established, the next step is to transform all the arguments, of each clause, into continuous intervals. In this phase, it is essential to consider the domain of the arguments. As the third argument is unknown, its interval will cover all the possibilities of the domain. The first argument speaks for itself. Therefore, one may have:

\[
\neg f_1(x_1, x_2, x_3) \leftarrow not f_1(x_1, x_2, x_3)
\]

$$f_1 = \begin{bmatrix} 2 & 0 & 2 & 0 \\ 3 & 0 & 3 & 0 \\ \end{bmatrix}, x_2 = \begin{bmatrix} 3 & 0 & 5 & 0 \\ 8 & 0 & 8 & 0 \\ \end{bmatrix}, x_3 = \begin{bmatrix} 5 & 5 & 15 & 5 \\ 15 & 5 & 15 & 5 \\ \end{bmatrix}
$$

...}

The Degree of Confidence (DoC) is evaluated using the equation $DoC = \sqrt{1 - \Delta^2}$, as it is illustrated in Figure 1, where $\Delta$ stands for the length of the argument’s intervals, once normalized.

\[
\neg f_{1_{DoC}}(x_1, x_2, x_3) \leftarrow not f_{1_{DoC}}(x_1, x_2, x_3)
\]

$$f_{1_{DoC}} = \begin{bmatrix} 0.67, 0.67 \end{bmatrix}, [0.375, 0.625], [0.1] \leftarrow not f_{1_{DoC}}(x_1, x_2, x_3)
$$

...}

4 A CASE STUDY

In order to exemplify the applicability of our model,
we will look at the relational database model, since it provides a basic framework that fits into our expectations (Liu and Sun, 2007), and is understood as the genesis of the LP approach to Knowledge Representation and Reasoning (Neves, 1984).

As a case study, consider the scenario where a relational database is given in terms of the extensions of the relations depicted in Figure 2, which stands for a situation where one has to manage information about thrombophilia risk detection. Under this scenario some incomplete and/or default data is also available.

For instance, in the Venous Thromboembolism Predisposing database, the Body Mass Index (BMI) is unknown, while the Blood Group Predisposition ranges in the interval [0.08,0.14].

In this study, to ensure the scalability of the method, the extension of the relational database includes the features, obtained by both objective and subjective methods, which were pointed relevant by the research done so far. Thus, physicians will fill the tables that link to the Venous Thromboembolism Predisposing one while executing the health check. The clinics may populate some issues, others may be perceived by additional exams (e.g. this happens earlier) and some incomplete and/or default data is also available.

The Body Mass Index (BMI) is evaluated using the equation $BMI = \frac{Body\ Mass}{Height^2}$ (WHO, 2014). In the Venous Thromboembolism Predisposing database, the domain of Body Mass Index column is in the range [0,3], wherein 0 (zero) denotes BMI < 25; 1 (one) stands for a BMI ranging in interval [25,30]; and 2 (two) denotes a BMI ≥ 30. Age/Heredity Predisposition column is based on Table 1, adapted from Sacher (1999). These predisposition values are clustered by age group and by heredity. Thus, the value of this parameter for the [0,40] age group is in the range [0,0.5] for general population, and in the range [0.05,0.5] for population with genetic antecedents. The blood group predisposition parameter, which is also evidenced in the Thrombosis Predisposing database, is based on Table 2 adapted from Spiezia (Spiezia et al. 2013). Its values are in the range [0.08,0.14] for O blood group, and in the range [0.18,0.30] for non-O blood groups.

Table 1: Age/Heredity predisposition (%), adapted from Sacher (1999).

<table>
<thead>
<tr>
<th>Age Group</th>
<th>General Population</th>
<th>Genetic Predisposition</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 40</td>
<td>[0.05]</td>
<td>[0.05,0.5]</td>
</tr>
<tr>
<td>40-75</td>
<td>[0.05,0.5]</td>
<td>[0.5,5.5]</td>
</tr>
<tr>
<td>&gt; 75</td>
<td>0.5</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 2: Blood group predisposition (%), adapted from Spiezia et al. 2013.

<table>
<thead>
<tr>
<th>Blood Group</th>
<th>Predisposition</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>[0.08,0.14]</td>
</tr>
<tr>
<td>non-O</td>
<td>[0.18,0.30]</td>
</tr>
</tbody>
</table>

The values presented in the remaining columns are the sum of the respective databases, ranging between [0.3], [0.10], [0.4] and [0.8], respectively for Thrombophilia Genic Factor, Thrombotic Risk Factors, Mutations/Polymorphisms and Earlier Secondary Factors columns. Then, one may have:

$$\text{thromb} : B_{\text{od}yM_{\text{ass}}} \land I_{\text{index}}, A_{\text{ge}} \land H_{\text{eredity}} \land \text{Predisposition}, B_{\text{lood\ group}} \land P_{\text{redisposition}} \land T_{\text{hrombophilia}} \land G_{\text{enic\ Factors}}, T_{\text{hrombotic\ Risk\ Factors}} \land M_{\text{utations}} \land P_{\text{olymorphisms}}, E_{\text{alier\ Secondary\ Factors}} \rightarrow [0,1]$$

where thromb stands for the predicate venous thromboembolism predisposing, 0 (zero) and 1 (one) denote, respectively, the truth values false and true. It is now possible to give the extension of the predicate thromb, in the form:

$$\text{thromb} \left( \text{BMI}, \text{AH}, \text{BP}, \text{TG}, \text{TR}, \text{MP}, \text{ES} \right) \leftarrow \neg \text{thromb} \left( \text{BMI}, \text{AH}, \text{BP}, \text{TG}, \text{TR}, \text{MP}, \text{ES} \right) \left[ \begin{array}{c} \text{attribute's values} \\ \text{attribute's domains} \end{array} \right]$$

$$\text{DoC} \left( \begin{array}{c} [0.3][0.5] [0.08,0.30][0.3][0.10][0.4][0.8] \ldots \\ \text{attribute's values ranges} \end{array} \right)$$

In this program, the first clause denotes the closure of predicate thromb. The next clause corresponds to patient 1, taken from the extension of the venous thromboembolism predisposing relation presented in Figure 2. Moving on, the next step is to transform all the argument values into continuous intervals and then normalize the predicate’s arguments in order to obtain the Degree of Confidence of the thromb predicate. One may have:

$$\text{thromb} \left( \text{BMI}, \text{AH}, \text{BP}, \text{TG}, \text{TR}, \text{MP}, \text{ES} \right) \leftarrow \neg \text{thromb} \left( \text{BMI}, \text{AH}, \text{BP}, \text{TG}, \text{TR}, \text{MP}, \text{ES} \right) \left[ \begin{array}{c} \text{attribute's values} \\ \text{attribute's domains} \end{array} \right]$$

$$\text{DoC} \left( \begin{array}{c} [0.3][0.5] [0.08,0.30][0.3][0.10][0.4][0.8] \ldots \\ \text{attribute's values ranges} \end{array} \right)$$

In this program, the first clause denotes the closure of predicate thromb. The next clause corresponds to patient 1, taken from the extension of the venous thromboembolism predisposing relation presented in Figure 2. Moving on, the next step is to transform all the argument values into continuous intervals and then normalize the predicate’s arguments in order to obtain the Degree of Confidence of the thromb predicate. One may have:
### Personal Information

<table>
<thead>
<tr>
<th>#</th>
<th>Age</th>
<th>Gender</th>
<th>Body Mass (Kg)</th>
<th>Height (m)</th>
<th>Blood Group</th>
<th>Strong Family Story</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>77</td>
<td>M</td>
<td>88</td>
<td>1</td>
<td>O</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>n</td>
<td>29</td>
<td>F</td>
<td>65</td>
<td>1,68</td>
<td>A</td>
<td>1</td>
</tr>
</tbody>
</table>

### Thrombophilia Genic Factors (Major)

<table>
<thead>
<tr>
<th>#</th>
<th>Antithrombin III</th>
<th>Protein C</th>
<th>Protein S</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>n</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

### Venous Thromboembolism Predisposing

<table>
<thead>
<tr>
<th>#</th>
<th>Body Mass Index</th>
<th>Age/Hereditity Predisposition</th>
<th>Blood Group Predisposition</th>
<th>Thrombophilia Genic Factors</th>
<th>Thrombotic Risk Factors</th>
<th>Mutations/Polymorphisms</th>
<th>Earlier Secondary Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0.5</td>
<td>[0.08,0.14]</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>n</td>
<td>0</td>
<td>[0.05,0.5]</td>
<td>[0.18,0.30]</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

### Molecular Analysis of Mutations/Polymorphisms

<table>
<thead>
<tr>
<th>#</th>
<th>Factor V Leiden mutation</th>
<th>Prothrombin 20210a mutation G/A</th>
<th>Methylene tetrahydrofolate reductase 677C/T</th>
<th>PAI-1 5G/4G Gene Polymorphism 675G/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>n</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

### Earlier Secondary Factors Predisposing to Thrombosis

<table>
<thead>
<tr>
<th>#</th>
<th>Smoke</th>
<th>Immobilization/Hospitalization</th>
<th>Air travel</th>
<th>Surgery</th>
<th>Liver disease</th>
<th>Infection</th>
<th>Oncologic pathology</th>
<th>Pregnancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>n</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

### Thrombotic Risk Factors

<table>
<thead>
<tr>
<th>#</th>
<th>Prothrombin</th>
<th>Factor VII</th>
<th>Factor VIII</th>
<th>Factor V Leiden</th>
<th>Fibrinogen</th>
<th>Plasminogen</th>
<th>Heparin cofactor II</th>
<th>Homocysteine</th>
<th>Phopholipid Antibodies</th>
<th>Previous Venous Thrombosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>n</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 2: Extension of the Relational Database model. In Molecular Analysis of Mutations/Polymorphisms and Earlier Secondary Factors Predisposing to Thrombosis databases, 0 (zero) denotes absence and 1 (one) denotes presence. In Thrombophilia Genic Factors (Major) database, 0 (zero) and 1 (one) denotes, respectively, functional and non-functional values. In the first eight columns of the Thrombotic Risk Factors database, 0 (zero) and 1 (one) denotes, respectively, normal and increased values, while in remaining columns denotes, respectively, absence and presence.

The logic program referred to above, is now presented in the form:

```prolog
\[ \neg \text{thromb}_{doc}(BMI, AH, BP, TG, TR, MP, ES) \leftarrow \text{not thromb}_{doc}(BMI, AH, BP, TG, TR, MP, ES) \]

\[
\text{thromb}_{doc}(0, 0.96, 1, 1, 1, 1, 1) \quad \text{attribute's confidence values} \\
:: 1 : 0.85 \\
[0.1] [0.1, 0.1] [0.0, 0.27] [0.0] [0.0] [0.25, 0.25] \quad \text{attribute's values ranges once normalized} \\
[0.1] [0.1] [0.1] [0.1] [0.1] [0.1] \quad \text{attribute's domains once normalized}
\]
```

...
\text{thromb}(BM, AH, BP, TG, TR, MP, ES) \leftarrow \text{not thromb}(BM, AH, BP, TG, TR, MP, ES)
\text{thromb}\left(\begin{array}{l}
0, [0.05,0.5], [0.18,0.30], 2, 2, 1, 1
\end{array}\right) \cdot 1 :: DoC
\begin{array}{llllllllllll}
[0,3] & [0,5] & [0.08,0.30] & [0,3] & [0,10] & [0,4] & [0,8]
\end{array}
\text{attribute's values}
\text{domains}

\text{1st interaction: transition to continuous intervals}
\text{thromb}\left(\begin{array}{l}
0, [0.05,0.5], [0.18,0.30], [2,2], [2.2], [0.4], [1.1]
\end{array}\right) \cdot 1 :: DoC
\begin{array}{llllllllllll}
[0,3] & [0,5] & [0.08,0.30] & [0,3] & [0,10] & [0,4] & [0,8]
\end{array}
\text{attribute's values ranges}
\text{domains}

\text{2nd interaction: normalization}
\text{thromb}\left(\begin{array}{l}
0, [0.01,0.1], [0.45,1], [0.67,0.67], [0.2,0.2], [0.1], [0,125,0.125]
\end{array}\right) \cdot 1 :: DoC
\begin{array}{llllllllllll}
[0,1] & [0,1] & [0,1] & [0,1] & [0,1] & [0,1] & [0,1]
\end{array}
\text{attribute's values one normalized}
\text{domains one normalized}

\text{DoC calculation: DoC} = \sqrt{1 - \Delta F}
\text{thromb}_{DoC}\left(\begin{array}{l}
1, 0.996, 0.838, 1, 1, 0, 1
\end{array}\right) \cdot 1 :: 0.83
\begin{array}{llllllllllll}
[0,0][0.01,0.1][0.45,1][0.67,0.67][0.2,0.2][0.1][0,125,0.125]
\end{array}
\text{attribute's values ranges one normalized}
\text{domains one normalized}

In Figure 3 it is shown how the normalized values of the interval boundaries and their DoC and QoI values work as inputs to the ANN. The output translates the \textit{venous thromboembolism predisposition risk}, and DoC the confidence that one has on such a happening. In addition, it also contributes to build a database of study cases that may be used to train and test the ANNs.

In this study were considered 300 patients from the south of Portugal, with an age average of 52 years, ranging from 27 to 82 years old. The gender distribution was 46% and 54% for male and female, respectively. To ensure statistical significance of the attained results, 20 runs were applied in all tests. In each simulation, the available data were randomly divided into two mutually exclusive partitions, i.e., the training set with two-thirds of the available data and, the test set with the remaining one-third of the cases. The back propagation algorithm was used in the learning process of the ANN. As the output function in the pre-processing layer it was used the sigmoid function.

The model accuracy was 97.6% for the training set (203 correctly classified in 208) and 93.5% for test set (86 correctly classified in 92).
6 CONCLUSIONS AND FUTURE WORK

Diagnosing venous thromboembolism predisposition risk has shown to be a hard task, as the parameters that cause the disorder are not fully represented by objective data. Therefore, it is mandatory to consider many different conditions with intricate relations among them. These characteristics put this problem into the area of problems that may be tackled by Artificial Intelligence based methodologies and techniques to problem solving.

This methodology for problem solving and the computational techniques used have the advantage of allowing one to consider incomplete and/or unknown information, a marker that is not present in existing systems. Future work may recommend that the same problem must be approached using others computational frameworks like Case Based Reasoning (Carneiro et al. 2013), Genetic Programming (Neves et al., 2007) or Particle Swarm (Mendes et al. 2004), just to name a few.

ACKNOWLEDGEMENTS

This work is funded by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within projects PEst-OE/EEI/UI0752/2014 and PEst-OE/QUI/UI0619/2012.

REFERENCES


East, A., Wakefield, T., 2010. What is the optimal duration
Thrombophilia Screening - An Artificial Neural Network Approach


