Logic Programming and Artificial Neural Networks in Pharmacological Screening of Schinus Essential Oils

José Neves, M. Rosário Martins, Fátima Candeias, Diana Ferreira, Silvia Arantes, Júlio Cruz-Morais, Guida Gomes, Joaquim Macedo, António Abelha, Henrique Vicente

Abstract—Some plants of genus Schinus have been used in the folk medicine as topical antiseptic, digestive, purgative, diuretic, analgesic or antidepressant, and also for respiratory and urinary infections. Chemical composition of essential oils of S. molle and S. terebinthifolius had been evaluated and presented high variability according to the parts of the plant studied and with the geographical and climatic regions. The pharmaceutical properties, namely antimicrobial, anti-tumoural and anti-inflammatory activities are conditioned by chemical composition of essential oils. Taking into account the difficulty to infer the pharmacological properties of Schinus essential oils without hard experimental approach, this work will focus on the development of a decision support system, in terms of its knowledge representation and reasoning procedures, under a formal framework based on Logic Programming, complemented with an approach to computing centered on Artificial Neural Networks and the respective Degree-of-Confidence that one has on such an occurrence.

Keywords—Artificial neuronal networks, essential oils, knowledge representation and reasoning, logic programming, Schinus molle, Schinus terebinthifolius raddi.

I. INTRODUCTION

Essential oils are secondary metabolites produced by flowers, leaves, stems, seeds, fruits or bark of aromatic plants. Usually, these compounds are liquid, volatile, limpid, usually with lower density than water and soluble in organic solvents. Essential Oils (EOs) play an important role in the protection of plants against herbivores and some insects and have also important antibiotic, antiviral, antifungal and insecticide properties [1], [2]. A large number of EOs and their individual components have been used as natural food flavorings, as food preservatives and as pharmaceutical agents, because of their functional properties [1], [3], [4].

Schinus L. species are trees from the Anacardiaceae family characterized by pungent-smell essential oils concentrated especially in fruits. The genus Schinus is native to South America particularly to the coast of Brazil and includes approximately 29 species [5]. Schinus molle L., also known as pink pepper, is naturalized in Southern Europe, including Portugal, as an ornamental plant [6].

Essential oil of S. molle and S. terebinthifolius, extracted from leaves and berries, had been characterized mainly by the high presence of monoterpens hydrocarbons, namely, myrcene, α-phellandrene, β-phellandrene and limonene [6]–[9]. The composition of these EOs can be different according with geographic and climatic factors and with the part of the plant (fruit or leaves) [6]–[9]. All parts of these plants have been used in traditional medicine for the treatment of several pathologies. Schinus plants were used in the folk medicine as topical antiseptic, digestive, purgative, diuretic, as analgesic and antidepressant and also for respiratory and urinary infections [10], [11].

Some studies about S. molle and S. terebinthifolius EOs highlight the biological properties, namely antimicrobial [6], [9], [12], antioxidant [5], [6], [8], anti-tumoural [5], [8] and anti-inflammatory activities [13], [14], and correlated them with the chemical composition.

Taking into account the geographical and seasonal variability of Schinus EOs chemical composition and the difficulty to infer their pharmacological properties without experimental assays for each EO, the present study was conducted with the objective to characterize 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 (KRR) [15], [16], complemented with a computational framework based on Artificial Neural Networks (ANNs) [17].

II. KNOWLEDGE REPRESENTATIONS AND REASONING

Many approaches to KRR have been proposed using LP, namely in the area of Model Theory [18]–[20], and Proof Theory [15], [16]. In this work it is followed the proof theoretical approach in terms of an extension to the LP language to KRR. An Extended Logic Program is a finite set of clauses in the form:
\[ p = p_1, \ldots, p_n, \neg q_1, \ldots, \neg q_m \] (1)

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

where \( ? \) is a domain atom denoting falsity, the \( p_i, q_i \), and \( p \) are classical ground literals, i.e., either positive atoms or atoms preceded by the classical negation sign \( \neg \). Under this emblematic formalism, every program is associated with a set of abducibles [18], [20] given here in the form of exceptions to the extensions of the predicates that make the program. Once again, \( LP \) emerged as an attractive formalism for knowledge representations and reasoning tasks, introducing an efficient search mechanism for problem solving.

Due to the growing need to offer user support in decision making processes some studies have been presented [21], [22] related to the qualitative models and qualitative reasoning in Database Theory and in Artificial Intelligence research. With respect to the problem of \( KRR \) using \( LP \), a measure of the Quality-of-Information (QoI) of such programs has been object of some work with promising results [23], [24]. The QoI with respect to the extension of a predicate \( i \) will be given by a truth-value in the interval \([0, 1]\).

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:

\[ \text{predicate}_i \leftarrow U_{1 \leq j \leq m} \text{clause}_j (x_1, \ldots, x_n) : \text{QoI}_i : \text{DoC}_i \] (3)

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:

\[ f_i (\ldots, f_2 (\ldots, \ldots, f_n (\ldots)) \text{ where } (n \geq 0) \] (4)

Assuming that a clause denotes a happening, a clause has as argument all the attributes that make the event. The argument values may be of the type unknown or members of a set, or may be in the scope of a given interval, or may qualify a particular observation. Taking into account the following clause where the first argument stands for itself, with a domain that ranges in the interval \([0, 12]\), the value of the second may fit into the interval \([5.5, 7]\) with a domain that ranges between 2.5 and 10, and the value of the third argument is unknown, being represented by the symbol \( \ominus \), with a domain that ranges in the interval \([0, 2]\). Let us consider that the case data is given by the extension of predicate \( f_i \), given in the form:

\[ f_i : x_1, x_2, x_3 \rightarrow [0, 1] \] (5)

where “\( f \)” and “\( \ominus \)” is one’s notation for sets, “\( 0 \)” and “\( 1 \)” denote, respectively, the truth values false and true. Therefore, one may have:

\[
\begin{align*}
\neg f_1(x_1, x_2, x_3) & \iff \neg f_1(x_1, x_2, x_3) \\
\neg f_1(\text{[6.6], [5.5, 7], [0.2]}) & : 1 : \text{DoC} \\
\text{attribute’s values} & \text{for } x_1, x_2, x_3 \\
\text{[0,12] [2,5,10] [0,2]} & \text{attribute’s domains for } x_1, x_2, x_3
\end{align*}
\]

Once the clauses or terms of the extension of the predicate are established, the next step is to set 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:

\[
\begin{align*}
\neg f_1(x_1, x_2, x_3) & \iff \neg f_1(x_1, x_2, x_3) \\
\neg f_1(\text{[0.5, 0.5], [0.4, 0.6], [0.1]}) & : 1 : \text{DoC} \\
\text{attribute’s values ranges} & \text{for } x_1, x_2, x_3 \\
\text{once normalized} & \\
\text{[0,1] [0,1] [0,1]} & \text{attribute’s domains for } x_1, x_2, x_3 \\
\text{once normalized}
\end{align*}
\]

The Degree of Confidence (DoC) is evaluated using the equation \( \text{DoC} = \sqrt{1 - \Delta I^2} \), as it is illustrated in Fig. 1. Here \( \Delta I \) stands for the length of the arguments intervals, once normalized. Therefore, one may have:
Neural Network given in Fig. 3.

The information regarding case 1 (one). Moving on, the next solving methodology, we will look at a relational model, since it provides a basic framework that fits into our expectancies [25], and is understood as the genesis of the LP approach to KRR [15].

As a case study, consider a database given in terms of the extensions of the relations (or tables) depicted in Fig. 2, which stands for a situation where one has to manage information in order to predict pharmacological activity of essential oils of Schinus plants, namely antimicrobial, analgesic, anti-tumoural and anti-inflammatory properties.

Under this scenario some incomplete and/or unknown data is also available. For instance, in case 1, the percentage of monoterpenes hydrocarbons ranges in the interval [58, 72], while the data regarding to antioxidant tests are unknown.

The values presented in the Hippocratic Screening and Antioxidant Activity columns of Pharmacological Activity Predict table are the sum of the correspondent values present in Hippocratic Screening and Antioxidant Activity Tests tables, ranging between [0, 6] and [0, 12], respectively. Now, we may consider the relations given in Fig. 2, in terms of the pharm_act predicate, depicted in the form:

\[
\text{pharm}_\text{act}: M_\text{monoterpenes}_\text{Hydrocarbons}, M_\text{monoterpenes}_\text{Oxigenated}, S_\text{sesquiterpenes}_\text{Hydrocarbons}, S_\text{sesquiterpenes}_\text{Oxigenated}, CL_\text{50}, DL_\text{50}, H_{\text{hippocratic Screening}}, A_{\text{antioxidant Activity}} \rightarrow \{0,1\}
\]

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

\[
\neg \text{pharm}_\text{act}(MH, MO, SH, SO, CL_{50}, DL_{50}, HS, AA) \iff \neg \text{pharm}_\text{act}(MH, MO, SH, SO, CL_{50}, DL_{50}, HS, AA)
\]

\[
\text{pharm}_\text{act}([58, 72], [0.5, 2], [2.4], [13.15], 48, 2500, 1, 1) : 1 :: \text{DoC}
\]

\[
[0.100, 0.100, 0.100, 0.100, (25, 3000), (100, 5000), (0.6), (0.12)]
\]

\[
\text{pharm}_\text{act}(0.990, 0.999, 0.999, 0.999, 0.999, 1, 1, 1, 0) : 1 :: 0.873
\]

\[
[0.58, 0.72], (0.005, 0.02), (0.02, 0.04), (0.13, 0.15), (0.008, 0.008), (0.49, 0.49), (0.17, 0.17), (0.1)
\]

\[
[0.1], [0.1], [0.1], [0.1], [0.1], [0.1], [0.1], [0.1], [0.1], [0.1], [0.1], [0.1]
\]

\[
\neg \text{pharm}_\text{act}(MH, MO, SH, SO, CL_{50}, DL_{50}, HS, AA) \iff \neg \text{pharm}_\text{act}(MH, MO, SH, SO, CL_{50}, DL_{50}, HS, AA)
\]

\[
\text{pharm}_\text{act}(0.990, 0.999, 0.999, 0.999, 0.999, 1, 1, 1, 0) : 1 :: 0.873
\]

\[
[0.58, 0.72], (0.005, 0.02), (0.02, 0.04), (0.13, 0.15), (0.008, 0.008), (0.49, 0.49), (0.17, 0.17), (0.1)
\]

\[
[0.1], [0.1], [0.1], [0.1], [0.1], [0.1], [0.1], [0.1], [0.1], [0.1], [0.1], [0.1]
\]

where its terms make the training and test sets of the Artificial Neural Network given in Fig. 3.
build a database of study cases that may be used to train and pharmacological activity is needed. In Fig. 3 it is shown how the numeric variables.

Table II, in turns, presents a brief statistical characterization of each variable and the data type, i.e., numeric or nominal. The dataset with 32 records. Table I shows a brief description of variables were selected allowing one to have a multivariable critical in the prediction of pharmacological activity. Fifteen classification models is the coincidence matrix, a matrix of identity one. In the other layers we used the sigmoid function. 

executions were performed for each one of them. To ensure statistical significance of the attained results, 30 (thirty) experiments were applied in all tests. 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.

A common tool to evaluate the results presented by the classification models is the coincidence matrix, a matrix of size $L \times L$, where $L$ denotes the number of possible classes.

### General Information

<table>
<thead>
<tr>
<th>#</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Month of Plant Crop</th>
<th>Part of Plant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>38°31'40'' N</td>
<td>8°01'15'' W</td>
<td>10</td>
<td>leaves</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$n$</td>
<td>38°31'40'' N</td>
<td>8°01'15'' W</td>
<td>10</td>
<td>fruit</td>
</tr>
</tbody>
</table>

### Chemical Composition

<table>
<thead>
<tr>
<th>#</th>
<th>Monoterpenes Hydrocarbons (%)</th>
<th>Monoterpenes Oxigenated (%)</th>
<th>Sesquiterpenes Hydrocarbons (%)</th>
<th>Sesquiterpenes Oxigenated (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[58,72]</td>
<td>[0.5,2]</td>
<td>[2,4]</td>
<td>[13,15]</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$n$</td>
<td>[91,98]</td>
<td>[0,1]</td>
<td>[0,1]</td>
<td>[1,3]</td>
</tr>
</tbody>
</table>

### Pharmacological Activity Prediction

<table>
<thead>
<tr>
<th>#</th>
<th>Monoterpenes Hydrocarbons</th>
<th>Monoterpenes Oxigenated</th>
<th>Sesquiterpenes Hydrocarbons</th>
<th>Sesquiterpenes Oxigenated</th>
<th>$CL_{50}$</th>
<th>$DL_{50}$</th>
<th>Hippocratic Screening</th>
<th>Antioxidant Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[58,72]</td>
<td>[0.5,2]</td>
<td>[2,4]</td>
<td>[13,15]</td>
<td>48</td>
<td>2500</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$n$</td>
<td>[91,98]</td>
<td>[0,1]</td>
<td>[0,1]</td>
<td>[1,3]</td>
<td>67</td>
<td>2000</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>

### Toxicity Tests

<table>
<thead>
<tr>
<th>#</th>
<th>$CL_{50}$</th>
<th>$DL_{50}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>48</td>
<td>2500</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$n$</td>
<td>67</td>
<td>2000</td>
</tr>
</tbody>
</table>

### Hippocratic Screening

<table>
<thead>
<tr>
<th>#</th>
<th>Catalepsy</th>
<th>Postural Reflexes</th>
<th>Tail Sensitivity</th>
<th>Pinene Reflex</th>
<th>Motor Activity</th>
<th>Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$n$</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

### Antioxidant Activity Tests

<table>
<thead>
<tr>
<th>#</th>
<th>Radical Scavenging</th>
<th>Lipid Protection</th>
<th>Haemoglobin Protection</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$n$</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Fig. 2 An extension of the relational model. In Hippocratic Screening table 0 (zero) and 1 (one) denote, respectively, normal and abnormal responses. The values of Antioxidant Activity Tests ranges in the interval [0, 4], where 0 (zero) stands for without antioxidant activity and 4 (four) denotes a very strong antioxidant activity. The $CL_{50}$ and $DL_{50}$ are expressed in mgdm$^{-1}$ and mgkg$^{-1}$ (p. o.), respectively.

### IV. ARTIFICIAL NEURAL NETWORKS

Several studies have shown how Artificial Neural Networks (ANNs) could be successfully used to structure data and capture complex relationships between inputs and outputs [26]–[28]. ANNs simulate the structure of the human brain being populated by multiple layers of neurons. As an example, let us consider the first case presented in Fig. 2, where one may have a situation in which the prediction of pharmacological activity is needed. In Fig. 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 pharmacological activity and 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 ANN.

The dataset holds information about the factors considered critical in the prediction of pharmacological activity. Fifteen variables were selected allowing one to have a multivariable dataset with 32 records. Table I shows a brief description of each variable and the data type, i.e., numeric or nominal. Table II, in turns, presents a brief statistical characterization of the numeric variables.

The dataset used in the training phase was divided in exclusive subsets through the 4-folds cross validation. In the implementation of the respective dividing procedures, ten executions were performed for each one of them. To ensure statistical significance of the attained results, 30 (thirty) experiments were applied in all tests. 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 identity one. In the other layers we used the sigmoid function.

A common tool to evaluate the results presented by the classification models is the coincidence matrix, a matrix of size $L \times L$, where $L$ denotes the number of possible classes.

This matrix is created by matching the predicted and target values. $L$ was set to 2 (two) in the present case. Table III present the coincidence matrix (the values denote the average of the 30 experiments). Table III shows that the model accuracy was 84.4% (27 instances correctly classified in 32). Thus, the predictions made by the ANN model are satisfactory and therefore, the generated model is able to predict pharmacological activity of *Schinus* essential oils.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monoterpenes hydrocarbons</td>
<td>Percentage of monoterpenes hydrocarbons</td>
<td>Numeric</td>
</tr>
<tr>
<td>Monoterpenes oxigenated</td>
<td>Percentage of monoterpenes oxigenated</td>
<td>Numeric</td>
</tr>
<tr>
<td>Sesquiterpenes hydrocarbons</td>
<td>Percentage of sesquiterpenes hydrocarbons</td>
<td>Numeric</td>
</tr>
<tr>
<td>Sesquiterpenes oxigenated</td>
<td>Percentage of sesquiterpenes oxigenated</td>
<td>Numeric</td>
</tr>
<tr>
<td>$CL_{50}$</td>
<td>Lethal concentration 50%</td>
<td>Numeric</td>
</tr>
<tr>
<td>$DL_{50}$</td>
<td>Lethal dose 50%</td>
<td>Numeric</td>
</tr>
<tr>
<td>Catalepsy</td>
<td>Has muscular rigidity</td>
<td>Nominal</td>
</tr>
<tr>
<td>Postural reflexes</td>
<td>Has automatic movements that control the equilibration</td>
<td>Nominal</td>
</tr>
<tr>
<td>Tail sensitivity</td>
<td>Has tail reflex</td>
<td>Nominal</td>
</tr>
<tr>
<td>Pineal reflexes</td>
<td>Has pineal sensitivity</td>
<td>Nominal</td>
</tr>
<tr>
<td>Motor activity</td>
<td>Has motor response</td>
<td>Nominal</td>
</tr>
<tr>
<td>Behavior</td>
<td>Response of the organism to various stimuli</td>
<td>Nominal</td>
</tr>
<tr>
<td>Radical scavenging</td>
<td>Has ability to scavenging free radicals</td>
<td>Nominal</td>
</tr>
<tr>
<td>Lipid protection</td>
<td>Has ability to inhibit lipid oxidation</td>
<td>Nominal</td>
</tr>
<tr>
<td>Haemoglobin protection</td>
<td>Has ability to inhibit the Fe$^{3+}$ oxidation</td>
<td>Nominal</td>
</tr>
</tbody>
</table>
and techniques for problem solving. This work presents the founding of a computational framework that uses powerful KRR techniques to set the structure of the information and the associate inference mechanisms. Indeed, this method brings a new approach that can revolutionize prediction tools in all its variants, making it more complete than the existing methods and tools available.

The KRR presented above are very versatile and capable of covering every possible instance by considering incomplete, contradictory, and even unknown data. Indeed, the new paradigm for KRR enables the use of the normalized values of the interval boundaries and their DoC values, as inputs to the ANN. The output translates the prediction of pharmacological activities and the confidence that one has on such a happening.

Future work may recommend that the same problem must be approached using others computational frameworks like Genetic Programming [16], Case Based Reasoning [29] or Particle Swarm [30], just to name a few.

REFERENCES


