

 <p>ISSN NO. 2320-5407</p>	<p>Journal Homepage: - www.journalijar.com</p> <h2>INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)</h2> <p>Article DOI: 10.21474/IJAR01/ 10145 DOI URL: http://dx.doi.org/10.21474/IJAR01/ 10145</p>	 <p>INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR) ISSN 2320-5407 Journal Homepage: http://www.journalijar.com Journal DOI: 10.21474/IJAR01</p>
-------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

RESEARCH ARTICLE

APPLICATION OF DECISION TREES IN THE PREDICTION OF DISEASE CASES IN TRADITIONAL MEDICINE.

Traore Issa¹, N'Guessan Behou Gerard², Achiepo Odilon³ and Kouame Euloge⁴.

1. Research professor at The University of Felix Houphouet-Boigny, Cote d'Ivoire.
2. Research professor at The Virtual University of Cote d'Ivoire.
3. Research professor The University of Peleforo Gon Coulibaly, Cote d'Ivoire.
4. Research professor The Virtual University, Cote d'Ivoire.

Manuscript Info

Manuscript History

Received: 01 October 2019
Final Accepted: 03 November 2019
Published: December 2019

Key words:-

Ontology, Decision Tree, Traditional Medicine, Owl, Imprecision.

Abstract

Traditional medicine is a source of health care accessible and affordable in Africa. It includes the traditional knowledge that meet primary health care needs. This knowledge based on experience is not structured and is filled with rigid and inadequate data that often lead to uncertainties and fatal errors. In this article, an ontology based on the knowledge of traditional medicine is developed. Thus, we propose a methodology to build a decision tree that corrects inaccuracies of traditional medicine. This consideration is based on the idea of probabilistic allocation of objects in different nodes of the tree based on a cut-off criterion. This is classification of diseases based on symp-toms. This work is a big step forward in the use of Semantic Web technologies, illustrated by a concrete case modeled and tested with a diagnostic problem in the case of traditional medicine.

Copy Right, IJAR, 2019,. All rights reserved.

Introduction: -

The field of African Traditional Medicine (ATM) is very difficult because the disease is considered as a "physical imbalance, mental or social" [1] and [2]. Unlike modern medicine, which bases its therapy on tests and formal examinations, diagnostics of traditional medicine are based on observation, oral questions and experience. Their therapeutic strategies in this area are mostly intuitive. If some of these practices are benign, others are dangerous. The information collected is the data used to establish a diagnosis. This data is not structured and organized.

One of the main challenges of machine learning is to design effective classification systems [3] based on a data set representative of a population of data. Machine learning is a branch of artificial intelligence [4], which deals with the ability to learn, to perform a task from the observation of a number of data in the field. The goal is how to identify the different classes of traditional medicine in diagnostics to facilitate decision making.

To address this concern, we propose the construction of ontology in the field of traditional medicine from the oral responses of a patient. This method has a particular advantage in including tracks of inaccurate data in a structured process with Ontology Web Language Description Logics (OWL-DL) [5]. Similarly, simplicity implies a low level of computation, the complexity can be adapted to the exploration of large databases. The strategy of classification with decision trees [6] is an effective way to determine if a patient has a disease from a group of symptoms.

Corresponding Author:- Traore Issa.

Address:- 1.Research professor at The University of Felix Houphouet-Boigny, Cote d'Ivoire.

This document is structured as follows: Section 2 presents the state of the art and the motivation of our work. Section 3 describes our ontological model developed with Protégé 5.2, while Section 4 shows an example of cases followed by mathematical modeling to solve the diagnostic problem. The results and discussion presented in section complete this work.

Motivation:-

In many African countries, traditional medicine is still the main source of health. The evolution of country progress indicators defined by the World Health Organization (WHO) strategy for traditional medicine is increasing dramatically [7]. This ancient medicine is based on the therapeutic methods handed down by tradition or acquired by the experience of a community. The goal of traditional medicine researchers is to provide prevention, effective diagnosis and treatment of disease. To this end, Launois and al [8] studied the so-called second generation tools, support tools for decision and have codified them to facilitate their implementation in areas where there are real public health issues. Traore et al. in [9], used semantic distances to search for medical information. As for Vili and al [10], they have the basic features of decision trees and a successful alternative approach to traditional induction focusing on existing applications and future potential in medicine.

Yuh-Jyh Hu and al [11], studied the control of patients' analgesia. They showed that the accuracy of predicting the consumption and analgesic requirements of the PCR (PCR dose only) by a set of decision trees was 80.9% and 73.1%, respectively. Chinami Matsumoto and al [12], expose a diagnostic algorithm using SELDI-TOF MS. Indeed, 53.8% of people of Oketsu syndrome were reviewed and confirmed by a decision tree. These were distinguished with 97.2% and 83.3% accuracy in the training set and the test set, respectively.

Methodology: -

The population which is based on our methodology consists of a set of diseases. This population is divided into three classes {yellow_fever, malaria, ictere}. The descriptions are made with the following three attributes: fever, headache and vomiting which are decimal values to attributes and sometimes logic as shown in Table 1 below. The table has 15 observations. Indeed it is to explain the behavior of individuals in relation to a set of diseases {malaria, yellow_fever, ictere} from symptomatic forecasts. It's about finding a partitioning of individuals which is represented in the form of a decision tree. The aim is to produce the most homogeneous possible groups of individuals from the perspective of the variable to be predicted. The database is established with Protege 5.2 (figure 4). The first objective has been achieved. The patient's data are structured in the ontology in (figure 1).

The second objective is to classify the implemented data. The purpose is to find the corresponding decision tree as shown in (figure 2). The nodes of the tree correspond to the tested attributes.

```
<?xml version="1.0"?>
<!DOCTYPE rdf:RDF [
  <ENTITY owl "http://www.w3.org/2002/07/owl#" >
  <ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
  <ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
  <ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
]>
<rdf:RDF xmlns="http://www.semanticweb.org/gerard/ontologies/2019/9/untitled-ontology-7#"
xml:base="http://www.semanticweb.org/gerard/ontologies/2019/9/untitled-ontology-7"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
<owl:Ontology rdf:about="http://www.semanticweb.org/gerard/ontologies/2019/9/untitled-ontology-7"/>
<!--
////////////////////////////////////
// Classes
```

Figure 1:- OWL knowledge structuring

Suppose we have the following attributes:

1. F for fever with $F = \{f_1, f_2, f_3, \dots, f_n\}$
2. C for headaches with $C = \{c_1, c_2, c_3, \dots, c_n\}$ and
3. V for vomiting with $V = \{v_1, v_2, v_3, \dots, v_n\}$.

For the construction of this tree, we must consolidate the attributes as a class. The first vertex, located at the first level, is called the "root" of the tree. This node contains all the attributes of the knowledge base. The branches are assigned to different possible values of the attribute like that (1):

$$\{x, x', x'', z, z', z'', t, t'\} \tag{1}$$

The sheets correspond to the different classes. The measurement of the whole of information involved in disease research can now be carried out.

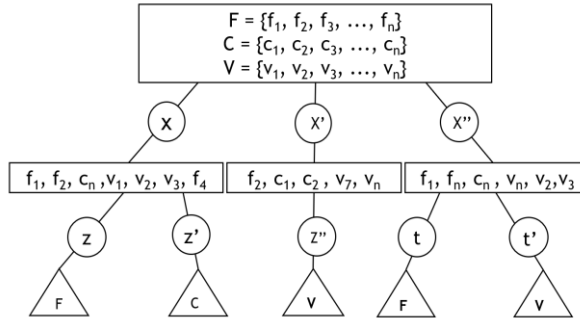


Figure 2:- Example of decision tree structure

The process stops when the elements of a node have the same value for the target variable.

Modelization: -

In this document, we use the entropy measure or information gain. This measure is used at each node of the decision tree. The highest information gain is chosen as the split attribute. Apply equation (2) to calculate the information necessary to classify a tuple in the set of diseases D, equation (3) is applied to find the necessary information (after using the X attribute of the symptoms to divide set D into partitions v).

$$H(D) = - \sum_{i=1}^m P_i * \log_2(P_i) \tag{2}$$

$$Gain(X) = H(D) - \sum_{j=1}^v \left| \frac{D_j}{D} \right| H(D_j) \tag{3}$$

The information obtained by connecting to attribute X can be obtained by subtracting the result from equation (2), which is shown in equation (3). Equation (2) is applied to the data set only once at the beginning. Then, equations (3) are applied to each attribute. This approach is presented by the algorithm (figure 3) below.

```

line number      Algorithm: Probability of choosing the node
-----
1  i: index of the table;
2  t: an element of the node;
3  N: the size of the array;
4  C: the set of classes;
5  Ci: class i of a node;
6  T[N]: the initial base;
7  T[S]: The new database;
8  DEBUT
9  Do
10  Read (T [N]);
11  for (l = 0; i <= N; i++) do
12  P (k | t); //calculated the probability of belonging
13  If (T in Ci) then
14  The <- t;
15  else
16  T [S] <- T [i];
17  End if
18  end for
19  While T [S]! = { }
20  END
    
```

Figure 3:-Algorithm method for selecting the node

The attribute that gives us the highest gain is used as a splitting attribute. The process is repeated at each level of the tree until we reach nodes of leaves that represent the classification of diseases.

Simulation:-

To ensure the classification of traditional medical knowledge, a structured database of diseases and clinical signs {fever, headache, vomiting and diarrhea} have been developed. Each clinical sign is characterized by a set of random values. This table contains more than 170 variables. In addition, we considered three types of diseases: malaria, ictere and yellow_fever. (figure 4) below was carried out based on the previous database. We used the statistical software R Ross Ihaka Robert Gentleman in 1988 it set up.

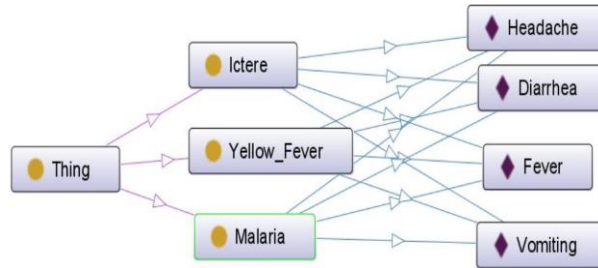


Figure 4:- Ontology of the field of the diseases studied

The interpretation made from this decision is that if the rates of all clinical signs is greater than the expected threshold, that is to say, 0.01 automatic decision will Ictere. By cons, if that rate is less than 0.01, it leaves us in a vague situation that requires several questions. When the decision is not clear, we recalculate this threshold in the possible cases:

1. If it is less than 0.01, now the decision will be ictere
2. Otherwise the decision will be yellow_fever.

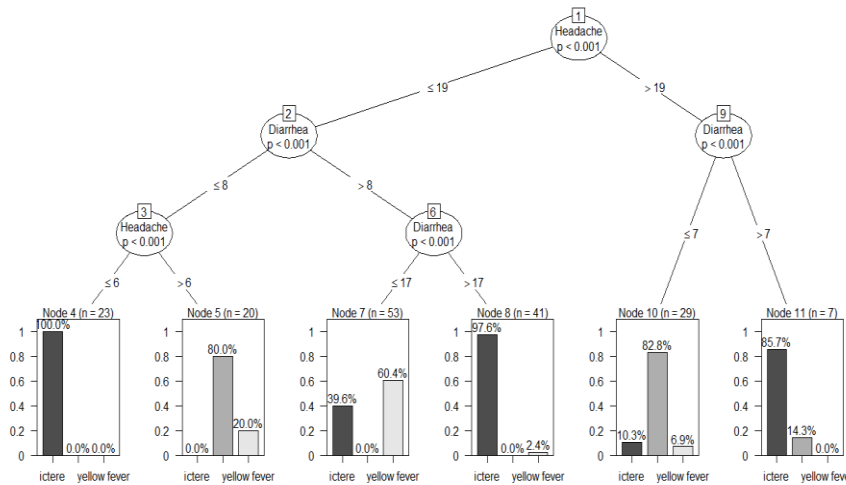


Figure 5:- Classification of diseases according to the frequency of appearance

The data classification (figure 5) performance was evaluated by calculating the true positive (VP), true negative (VN), false positive (FP) and false negative (FN), the percentage sensitivity (S_c), the specialty (S_p) and the classification rate (T_a) their respective definitions are as follows:

1. VP: “true positive” for correctly predicted event values.
2. FP: “false positive” for incorrectly predicted event values.
3. VN: “true negative” for correctly predicted no event values.
4. FN: “false negative” for incorrectly predicted no event values.

Sensitivity is the ability to give a positive result when diabetes is present. S_e is given by the formula:

$$S_e = \frac{VP}{VP+FN} \quad (4)$$

S_p specificity is the ability to give a negative result when diabetes is absent. S_p is given by the formula:

$$S_p = \frac{VN}{VN+FP} \quad (5)$$

The T_a classification rate is the percentage of correctly classified knowledge. T_a is given by the formula:

$$T_a = \frac{VP+VN}{VP+VN+FP+FN} \quad (6)$$

The prediction of the test sample by the decision tree leads to the confusion matrix below and gives an estimated error of 4.67% (figure 6).

```
Call:
  randomForest(x = dataset[, 1:4], y = dataset$Diseases)
  Type of random forest: classification
  Number of trees: 500
  No. of variables tried at each split: 2

  OOB estimate of error rate: 4.67%
Confusion matrix:
      ictere malaria yellow_fever class.error
ictere      46      0          4      0.08
```

Figure 6:- Confusion matrix for the decision tree model

The results of the confusion matrix suggest that our approach is encouraging because we have an accuracy rate of 95.33%.

Conclusion:-

This work presents a methodology for the analysis and classification of traditional medicine knowledge. We have built an ontology in the field of traditional medicine with 5.2 protection to better structure the data. It contains more than 170 traditional knowledge variables that are generally treated informally by traditional physicians. A classification of this data was performed using decision tree technology. This approach has made knowledge in the field of traditional medicine more understandable and facilitates decision-making.

Our future work will focus on dosage and dosage of remedies related to the therapy of traditional medicine.

References:-

1. Ali Messaoud, Mohamed Ben Messaoud, Abdennaceur Kachouri and Faisal SELLAMI: Classification of cardiac arrhythmias fuzzy from ECG signals SETIT (2005).
2. Behou Gerard N'Guessan, proposalm éthodologies to value traditional medicine based on an ontology. Computer [cs]. Université d'Abobo-Adjamé, (2016).
3. Simon Marcellin: decision trees in a situation of asymmetry, , Thesis, ERIC laboratory and society Fenics,(2008).
4. Laurent Heutte, Simon Bernard, Sebastien Adam, Emilia Oliveira: From the selection of decision trees in Random Forests. Antoine Tabbone and Thierry Paquet. International Conference on Frontiers in Writing and Paper, France. Research Group in Written Communication, pp.163-168, (2008).
5. Oumtanaga Souleymane, Traore Issa, Babri Michel: Specifying a model of semantic web service composition; International Journal on Computer Science and Engineering (IJCSSE);Vol.3, Issue.10, ISSN 0975-3397, pp.3393 à 3402, (2011).
6. Simon Marcellin: Decision Trees in a situation of asymmetry, These, September 2, 2008
7. WHO:Traditional Medicine Strategy 2014-2023, ISBN 978 92 4 150609 0, Licence: CC BY-NC-SA 3.0 IGO, (2013).
8. Launois R., Payet. Skid frança: The evaluation of health technologies, second-generation protocols', 1 REES France, 28, rue d'Assas, 75006 Paris, France.

9. Traoré Issa, Souleymane Oumtanaga, Babri Michel, Claude Lischou: Approach to a conceptual model of indexation in a weak ontology described in RDFS, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 6, No 1, pp225-235, (2011).
10. Vili Podgorelec, Peter Kokol, Bruno Stiglic, Ivan R.: Decision Trees: An Overview and Their Use in Medicine, Journal of Medical Systems, Vol. 26, No. 5, (2002).
11. Yuh-Jyh Hu, Tien H.: Decision tree-based learning to predict patient controlled analgesia consumption and readjustment, BMC Medical Informatics and Decision Making (2012).
12. Chinami Matsumoto, Tetsuko K., Kazuo O.: A Proteomic Approach for the Diagnosis of 'Oketsu' (blood stasis) has Pathophysiologic Concept of Japanese Traditional (Kampo) Medicine, ; 5 (4) 463-474, eCAM (2008).