International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-4, November 2019

System to assist in the Diagnosis of Diabetes using Ontology and Machine Learning

Traore Issa, Oumtanaga Souleymane, Claude Lishou

Abstract: ***Diabetes mellitus has become a public health problem in both developed and developing countries. If it is not treated early, diabetes-related complications in many vital organs of the body can become fatal. Its early detection is very important for early treatment that can prevent the disease from progressing to such complications. This article focuses on designing a system to assist in the diagnosis of diabetes disease based on medical ontology and automatic learning. The proposed method uses automatic learning algorithms as a classifier for the diagnosis of diabetes based on a medical data set. The ontology suggests a pre-processing of a coherent, consistent, interoperable and shareable knowledge basis of data and the machine learning method focuses on classification based on symptoms and medical tests. Based on the experimental results, DDAS not only offers better performance in predicting and diagnosing diabetes in individuals, but also has better accuracy in recommending useful treatment to patients.***

Keywords**: Ontology Machine learning, Decision tree, diabete, Classification, Clinical decision support system**

I. INTRODUCTION

Diabetes is a disease in which blood sugar (glucose) is not metabolized in the body. This increases blood glucose levels to alarming levels. This is known as hyperglycemia. In this state, the body is unable to produce enough insulin. The other possibility is that the body may not be able to respond to the insulin produced. Diabetes is incurable; it must be controlled. A person with diabetes can develop serious complications such as nerve damages, heart attacks, kidney failures and stroke. According to 2017 statistics, it is estimated that 8.8% of the world's population has diabetes 6. This percentage is expected to increase to 9.9% by 2045 [2]. A lot of research has been conducted on the Clinical Decision Support System (CDSS) based on massive and distributed electronic health record data [3], [4]. These systems propose to solve problems of ambiguity and semantic inconsistency by performing a detailed analysis of patients' symptoms for parameter classification by automatic learning methods. It is important to adapt strategies for the prevention, detection, treatment or monitoring of diabetes in individual patients according to the complete medical profile [5], [6] and [7]. The aid systems' approach allows for better results in the detection and follow-up of previous profiles and aspects of type 1 and 2 diabetes, including complications, symptoms, laboratory tests, interactions, Treatment Plan frameworks (TP) and glucose related diseases.

Our approach provides an effective model based on the logic of formal ontology description built with Protected 5.2, and Machine Learning to help physicians make an effective decision for the diagnosis and follow-up of diabetic patients. The proposed method is based on automatic learning, Decision Tree.

The remaining part of the document is organized as follows: The second part presents the related work and motivation, the third part deals with the methods and techniques settled for the construction of the Diabetes Diagnostic Assistance System (DDAS), the fourth part presents the results and discussions of the experiment done, and finally a conclusion to this article.

II. RELATED TO WORK

A. Diagnostic aid systems

The potential of e-learning for the health sector is immense. Although diabetes has serious consequences for the human body, a large proportion of its cases and complications could be prevented by good blood glucose control and early diagnosis. This requires the use of a diagnostic aid system to facilitate decision-making and minimize uncertainty about the patient's current or future condition. The diagnosis of diabetes seems to be difficult under computer constraints, as there are several types (type 1 and type 2 diabetes) of the same disease. In addition, some of the main signs of diabetes may appear secondarily or late. Several categories of computer systems can solve this type of problem as follows:

- Expert systems [8], [9] is a decision support tool capable of reproducing the cognitive mechanisms of an expert or group of experts. In other words, an expert system is able to answer questions by reasoning from known facts and rules;
- Classification systems [10], [11] are based on identifying the classes to which objects belong basing on certain parameters. It is suitable for automated decision-making problems related to a large number of human activities [12]. This type of system has been successful in the scientific community thanks to the techniques of automatic learning.

Early diagnosis is the first step towards managing this disease.

Manuscript published on November 30, 2019.
* Correspondence Author
Traoré Issa*, his Institut of Mathematics research (IMAR), Felix Houphouet-Boigny University, Abidjan, Côte d'Ivoire, Email: issa.traore@ufhb.edu.ci
Oumtanaga Souleymane, Laboratory for Informatics and Telecommunications Research (LARIT), INPHB, 08 BP 475 Abidjan 08 (225), Cote d'Ivoire. Email: oumtana@gmail.com
Claude Lishou, Génie Électrique, Cheikh Anta Diop University, Dakar | UCAD · ESP, Dakar, Senegal Email: claude.lishou@gmail.com

However, a diagnosis involves several variables, which makes it difficult to establish an accurate and timely diagnosis and to construct accurate personalized treatment plans. To ensure an accurate diagnosis, a medical diagnostic aid system is settled to minimize possible errors that may occur during the diagnosis of a disease.

B. Proposed approach

But with the presence of redundant attributes or loudly noisy attributes in databases, system performance can be compromised. This requires the use of variable selection techniques that aim to extract an optimal subset of the most relevant characteristics or parameters to improve learning and ensure good system performance. The main contributions of this article can be summarized as follows:

1) The construction of a Diabetes Diagnostic Ontology (DDO). That is to say, an ontological knowledge base of the medical field, in particular that of diabetes, with the Protected 5.2 tool;
2) An architecture of the Diabetes Diagnostic Assistance System (DDAS) is also presented.
3) The use of automatic learning for the classification of knowledge extracted from the DDO in order to diagnose a patient;

To do this, the DDO is built on the basis of a medical data set from the UCI's automatic learning laboratory [13]. Symptom descriptions are used to diagnose the presence of diabetes or not.

III. METHODOLOGY

An electronic health record system requires an integrated decision support capability and ontologies are rapidly becoming necessary for the design of efficient, reliable, scalable, reusable and intelligently semantic knowledge bases. Ontology is a global or abstract representation of a field.

A. Implementation of the DDO

The construction of the DDO is based on the methodology of Gomez Pérez and al. [14]. It is based on three steps:

- Extraction of terms and their properties with the text mining method that takes as input the corpus of texts and resources from patients' questionnaires and laboratory tests. Then the syntactic analysis and extraction of pairs (object, property) and triplets (object, relationship, property) present in the same syntactic phrases come ;
- Construction of the ontology core: it consists in using pairs to build a hierarchy of concepts with the Analysis of Concepts Formal (ACF);
- Extraction of transversal relationships: it consists in taking as entry the triplets extracted from the text, then the Concept Relational Analysis (ARC) to extract the transversal relationships.

Thus, the DDO is built from medical diabetic resources such as: number of fat, plasma glucose concentration, insulin dose, BMI etc. This phase determines the aspects that DDO must cover see Fig. 1 The specification of requirements can be defined informally by a set of competency questions defined by a diabetes diagnostic expert. To properly diagnose diabetes, doctors need to know the answers to the following questions:

- What are the results of the patient's laboratory tests?
- What are the patient's demographics?
- What are the patient's current complications?
- What are the patient's symptoms?
- What are the results of the patient's physical examination?
- What drugs are currently being taken by patients who may affect glucose levels or the functioning of the pancreas?
- What are the chemical ingredients in these drugs?
- What are the possible diagnoses for diabetic patients?

The CDSS must have the ability to use the answers to these questions to diagnose a patient. These questions will guide the next steps, including the evaluation of ontology.

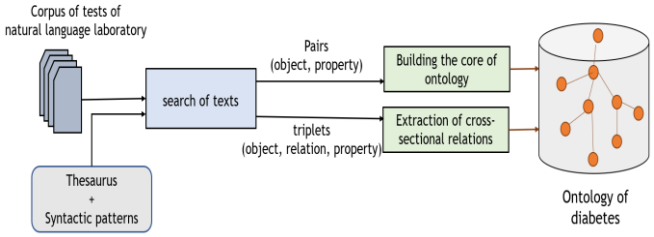

Fig. 1. Construction Ontology of Diabetes Diagnosis (DDO)

Thus, the constructed DDO does not contain instances or individuals in most cases see Fig.2. In our design, ontology contains only concepts and properties that must be inspired by existing standard ontologies and terminologies in [15], [16]. This ontology is serialized in OWL2 format with the Protected Tool5.2. In order to standardize the terminology used to refer to document management concepts and integrate it with other terminology sources, all DDO concepts are annotated with standard concept identifiers, synonyms and definitions collected from SCT, UMLS and RxNorm, where available.

As a result, DDO concepts are annotated with many types of additional information, such as SCT concept IDs, UCI UMLS, CUI RxNorm, text definitions and alternative terms (synonyms). Synonym annotation is used to specify other concept names in DDO. The design principles of the W3C [17]. The metric data collected using Protected is used to predict diabetes diagnoses using automatic learning.

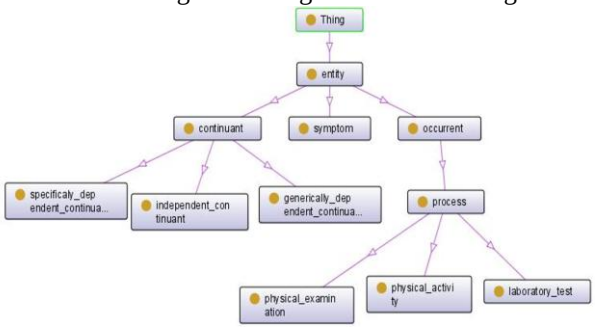

Fig. 2. Architecture of DDO

Retrieval Number: D4421118419/2019©BEIESP
DOI:10.35940/ijrte.D4421.118419
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

When we use ontology to build a CDSS connected to an Electronic Health Record (EHR), the instantiation of ontology is performed according to each set of personalized patient conditions and characteristics. This customization facilitates personalized diagnosis and treatment.
Such an ontology is an important step towards developing a smarter research agenda for the management of chronic diseases, particularly diabetes. EHRs ensure that individuals can access their medical records and laboratory test results to track, monitor and manage their long-term illnesses. In the section below, an overview of the CDSS architecture provides an understanding of the important steps in detecting a patient with diabetes.

B. Structure of the CDSS model

In this section, we present the architecture of the CDSS model. It is a system that must be able to diagnose a patient with diabetes by entering a number of clinical descriptions. According to Fig. 3 below, the DDO data are classified by the automatic learning method. Several automatic learning methods exist, the one with the lowest error rate is chosen to classify diabetic knowledge.

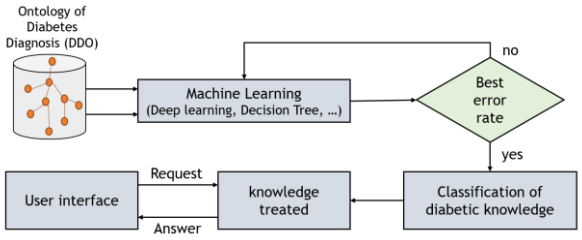

Fig. 3. Architecture du CDSS

The user enters the clinical descriptions obtained after the clinical laboratory tests. The CDSS system processes the user's request and gives the answer. This architecture integrates a cognitive layer of semi-structured documents in the OWL language and thus allows artificial intelligence to be naturally introduced into the management of patients with diabetes. Thus, the aim of this study is to develop a theoretically sound and semantically intelligent knowledge base to solve problems related to the diagnosis of diabetes.

C. Classification of DDOs with Machine Learning

For the implementation of an ontology, ontological development tools and languages are used. The OWL2 language is chosen to describe our ontology model. The ontology was implemented using the Protected Ontology Editor 5.2 with protégé [18] and its consistency was verified using a set of reasoners, including FACT ++, Racer and Pellet. OWL2 was chosen to formulate the ontology because it provides the maximum expressiveness capacity that can be offered, while guaranteeing the total computing capacity. The main components of DDO are presented in Fig. 4, depending on the ontology, the patient's diagnosis involves the verification of many conditions. The data identified by experts, the literature and guidelines as relevant to the diagnosis of diabetes are classified into eight groups, described in the ontology in eight general classes: disease, laboratory test, physical examination, demography, symptom, disorder, drug, and chemical substance.

The classification models of Depp Learning and Decision Tree were used and evaluated. The following clinical descriptions of diabetes were used:

- Npreg: number of large ones;
- Glu: plasma glucose concentration;
- BP: diastolic blood pressure (mm Hg) ;
- SKIN: triceps skin fold thickness (mm) ;
- Insulin: insulin dose (mu U/ml);
- BMI: body mass index (weight in kg/(height)2 ;
- PED: pedigree function (heredity);
- AGE: age (year).

The automatic learning experiments are conducted on Python3.7. The given knowledge sets are stored in OWL and imported into OWLReader2 [19]. Owlready2 is an ontologically oriented programming module under Python 3, including an optimized RDF quadstore. Import OWL 2.0 ontologies in NTriples, RDF / XML or OWL / XML format.

IV. RESULT AND DISCUSSION

A. Results

In this section, we present the results of the analysis of the experiment. The Decision Tree classification models are used to identify patients with and without diabetes see Fig. 4. The Decision Tree thus provides data on the percentage of the existence of diabetes or not. The solution of a classification problem and more generally of a modelling problem is carried out by comparing models in order to choose the most suitable to solve the problem.

Model Evaluation is therefore an unavoidable prerequisite for selection.

Modern artificial intelligence methods such as in-depth learning can complement the expertise of pathologists to ensure consistent diagnostic accuracy. We have developed a computer-based approach based on in-depth learning that automatically assigns a score that identifies whether a patient has diabetes or not. The choice of evaluation meth od is essential in the diagnosis of the disease in a patient [20],[21]. The performance of a test or model can be assessed by many indicators. Computer-assisted diagnosis holds great promise for clinical decision-making in personalized oncology. In the logical continuation of the analysis of the ROC curves, the calculation of the areas under the curves makes it possible to evaluate the models numerically and more precisely.

B. The evaluation criteria

The data classification performance was evaluated by calculating the true positive (VP), true negative (VN), false positive (FP) and false negative, the percentage sensitivity (SE), the specialty (SP) and the classification rate (TC) their respective definitions are as follows:

- VP: diabetic classified as diabetes
- VN: not diabetic classified as non-diabetic
- FP: not diabetic classified as diabetic
- NF: diabetic classified as non-diabetic

Sensitivity is the ability to give a positive result when diabetes is present.

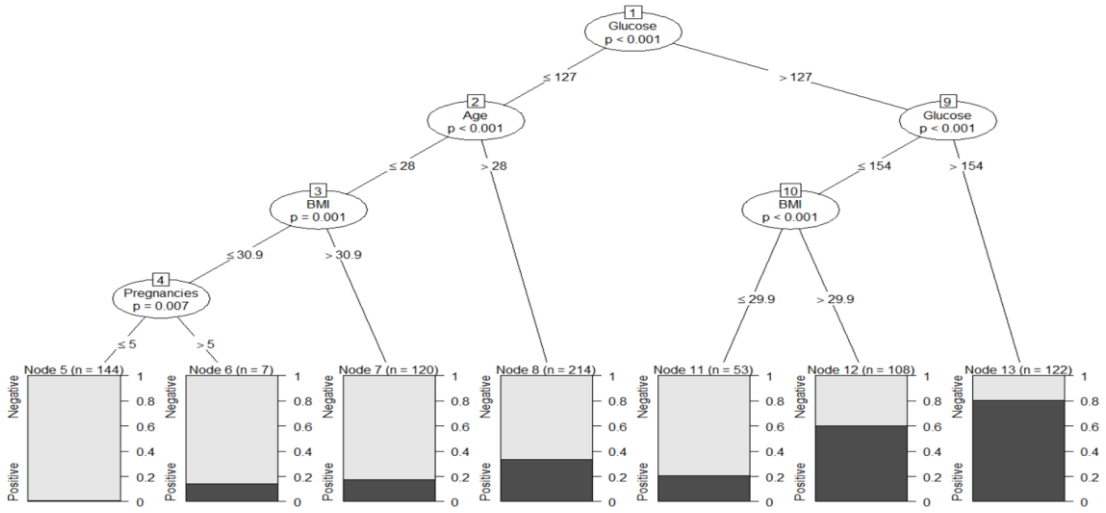

Fig.4. Classification of diseases according to the frequency of appearance

S_e is given by the formula:

$$S_e = \frac{VP}{VP+FN} \quad (1)$$

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

$$S_p = \frac{VN}{VN+FP} \quad (2)$$

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

$$T_a = \frac{VP+VN}{VP+VN+FP+FN} \quad (3)$$

Finally, according to rule (1), (2) and (3) we have the in Table-I below, the classification experiments are conducted on the Diabetes dataset:

Table- I: Measures of Classification Models

	positive_test	***negative_test***	***sum***
positive_test	VP =90	FN =159	249
negative_test	FP =18	VN =501	519
somme	108	660	768

The terms of the tables used for the analysis are calculated and the following results are obtained:

$S_e = 36,14$; $S_p = 96,53$ and $T_e = 76,95$

The model (Decision Tree) accuracy is: 76,95

Fig. 4 also shows that glucose levels, BMI and age take an important role in the diabetes control system.

The training set accuracy of Diabetes data set is 76,95% and the testing accuracy is 82.2% for the Decision Tree. The classifier from the cross-validation accuracy it is noticed that there is significant improvement in the accuracy if the number of training samples increases.

V. CONCLUSION

The aim of this study was to develop a CDSS, a theoretically sound and semantically intelligent knowledge basis for solving problems related to the diagnosis of diabetes. Such a System can lead to the emergence of a new patient-centred CDSS class that can help physicians diagnose diabetics quickly and accurately.

DDO provides a standard ontology that can support interoperability between the CDSS and health care systems. In addition, it can be used in combination with a rule basis to build a rule-based diabetes diagnosis system. 6

Ontology includes diabetes complications, laboratory tests, symptoms, physical examinations, demographic data and diagnoses. The CDSS thus developed uses machine learning methods, in particular Decision Tree, to diagnose diabetes with an accuracy of 98.1%. Future work will focus on using DDO to build a much richer CDSS system with other medical ontology. The objective will be to have a more complete DDO th666rough semantic alignment. We will continue the study until we reach the treatment of diabetes.

REFERENCES

1. World Health Organization (WHO). *Global status report on noncommunicable diseases 2014.* Geneva: WHO, 2015, http://www.who.int/nmh/publications/ncd-status-report-2014/en/, 17 February 2017
2. P. Romero-Aroca, A. Valls, A. Moreno, R. Sagarra-Alamo, J. Basora-Gallisan, E. Saleh, M. Baget-Bernaldiz, , and D. Puig, *A Clinical Decision Support System for Diabetic Retinopathy Screening: Creating a Clinical Support Application* , 11 Jan 2019
3. C. Valverde, M. Garcia, R. Hornero, M. Lopez-Galvez. *Automated detection of diabetic retinopathy in retinal images*, Indian J Ophthalmol,64, p.26–32, 2016
4. A. Malhotra, E. Younesi, M. Gündel, B. Müller, M. Heneka, M. Hofmann-Apitius, *ADO: A disease ontology representing the domain knowledge specific to Alzheimer's disease.* Alzheimers Dement 10 (2), pp.238–246, 2014
5. R. Alizadehsani, M. Roshanzamir, M. Abdar, A. Beykikhoshk, A. Khosravi, *A database for using machine learning and data mining techniques for coronary artery disease diagnosis*, Scientific Data volume 6, (227), 2019
6. Neerincx MA, Kaptein F, M. A.Van Bekkum, *Ontologies for social, cognitive and affective agent-based support of child's diabetes self-management*, pp.35–38
7. R. Imhanlahimi E1 and A. John-Otumu, *Application of expert system for diagnosing medical conditions: a methodological review*, European Journal of Computer Science and Information Technology, Vol.7, No.2, pp.12-25, April 2019
8. M. Elhoseny, K. Shankar, and J. Uthayakumar, *Intelligent Diagnostic Prediction and Classification System for Chronic Kidney Disease*, Scientific report, 03 July 2019

9. O. Souleymane, T. Issa, Babrı Michel, *Specifying a model of semantic web service composition*, International Journal on Computer Science and Engineering (IJCSE);Vol.3, Issue.10, ISSN 0975–3397, pp.3393-3402, 2011
10. Z. Chen , Z. Zhang, R. Zhu, Y. Xiang, *Diagnosis of patients with chronic kidney disease by using two fuzzy classifiers*, Chemometrics and Intelligent Laboratory Systems. 153, pp.140–145, 2016
11. S. Gopika, and M. Vanitha, *Efficiency of Data Mining Techniques For Predicting Kidney Disease*, International Journal of Engineering and Technology (IJET). 9, pp.3586–3591, 2017
12. https://archive.ics.uci.edu/ml/index.php
13. G.-Pérez, *Evaluating ontology evaluation*, IEEE Intelligence Systems, 19(4), pp.74-76, 2004
14. H. S. Ali, D. Kwak, *DMTO: a realistic ontology for standard diabetes mellitus treatment*, Kyung Sup Kwak, J. Biomedical Semantics, 2018
15. A mobile health monitoring-and-treatment system based on integration of the SSN sensor ontology and the HL7 FHIR standard
16. Shaker H. Ali El-Sappagh, Farman Ali, Abdeltawab M. Hendawi, J. Jun-Hyeog, *BMC Medical Informatics and Decision Making*, Kyung-Sup Kwak, 2019
17. http://www.rifca.nl/wp-content/uploads/2016/12/http://www.w3.org/TR/swbp-specified-values
18. http://www.protege.stanford.edu
19. https://pypi.org/project/Owlready2/
20. A. Jakka, V. Rani, *Performance Evaluation of Machine Learning Models for Diabetes Prediction*, International Journal of Innovative Technology and Exploring Engineering (IJITEE), Blue Eyes Intelligence Engineering & Sciences Publication, ISSN: 2278-3075, Volume-8 Issue-11, p.1976–1980, September 2019
21. Z. Tao, W. Xie, L. Xu, X. He, Y. Zhang, M.You, *A machine learning-based framework to identify type 2 diabetes through electronic health records*, International Journal of Medical Informatics, 97, pp.120-127, 2017

AUTHORS PROFILE

Traoré issa, received the doctorate at the university cheikh anta diop (UCAD). He is currently working as a searcher in Institut of Mathematics research (IMAR) in Felix Houphouet-Boigny University. His research interests include telecommunication, Big data, System & Network security and Machine Learning. He teaches data processing and the telecommunication in UFHB.

Oumtanaga Souleymane, is a Professor in Data processing and telecommunication. He is professor at the National Polytechnic Institute Felix Houphouët-Boigny, (INP-HB) one of the most prestigious tertiary institutions in Côte d'Ivoire. and is the director of LARIT Laboratory. He is making a lot of research in telecommunication, network and web mining

Claude Lishou, is a Professor in data processing. He also teaches at the university Cheick Anta Diop (UCAD) of Dakar in Senegal. He supervises the students in Thesis of doctorate. He is my teacher adviser for my thesis, he is making in many fields as follows : mathematics, computing and telecommunication. He is in charge of the technological computing doctorate school of UCAD in Dakar.