Semantic Web Based ST Algorithm Framework for Diagnosing and Monitoring of Breast Cancer on an Intelligent Personal Agent (IPA)

O. N. Oyelade
Department of Computer Science, Faculty of Physical Science Ahmadu Bello University, Zaria, Nigeria
olaide_oyelade@yahoo.com

A. A. Obiniyi
Department of Computer Science, Faculty of Physical Science Ahmadu Bello University, Zaria, Nigeria
aaobiniyi@yahoo.com

S. B. Junaidu
Department of Computer Science, Faculty of Physical Science Ahmadu Bello University, Zaria, Nigeria
abuyusra@gmail.com

Abstract— *Clinical diagnosis and monitoring are gradually shifting into the homes of patients. These are now easily attained because of increased efforts geared towards creation of more approximate medical diagnostic reasoning algorithms (MDRAs). Researchers have crafted different medical algorithms for developing healthcare delivery systems, with most of these algorithms being built on the knowledge clinicians have learnt through study and experience during diagnostic procedures. Some of these models are based on statistical, mathematical, fuzzy and rule based techniques. Despite the differences in their underlying approaches, they are all oriented towards a MDRA with higher precision. In this research in progress, we develop an enhanced MDRA that is particularly addressing the limitations of the reasoning functions of an MDRA called Select and Test (ST) algorithm. The logical inference making process of ST is being limited by its use of simple logical constructs and some applications of mathematical methods. Therefore, the semantic based MDRA framework this paper presents, builds on the ST reasoning structures, aided by using the semantic web concept. We then model the knowledgebase using an ontological approach, design and implement a coordinated rule system for effective reasoning, and uses semantic web based rule/reasoning engines for rule implementation and inference making respectively. This enhanced framework adds a monitoring agent that autonomously improves both its knowledge base and to actualize its monitoring task. We use our enhanced MDRA as a test bed for breast cancer diagnosis, and designed a set of metrics for comparing the result of our improved ST algorithm with the existing ST algorithm.*

Keywords—Semantic Web, Inference making, Ontology, Rule set and Diagnosis

I. Introduction

Though cancer and some other non-communicable diseases have been known since antiquity, but the need for accurate diagnosis has necessitated further researches that will help clinicians to understand the basis of their clinical presentations. Cancer is a disease that occurs when there is an uncontrolled division of cell, and then grows out into other tissues that are healthy. The uncontrolled growth also known as cancer cells [1] is usually treated by using surgery, chemotherapy or radiotherapy [2] [3] [4]. Cancer diagnosis, in particular, has been diagnosed early enough through Pap smear test, colon cancer screening and mammograms, and all these tests have helped in curtailing the death tolls resulting from unmonitored effects on cervical, colon, and breast cancer respectively. However, increasingly exposing cancer patients to sensitive diagnostic tools can lead to over diagnosis [5]. Hence, a continuous use of a monitoring/diagnosis system that is self-operative by patient alongside the professional guide of a clinician can help reduce the exposures of patient to some dangerous therapies. In [6], the author talked about biologically tailoring therapy to individuals, this also known as personalized (cancer) medicine, is the most promising area for modern (cancer) therapy. This proposal hinges on stand that promotes the deployment of an intelligent personal agent (IPA) for monitoring and diagnosis of ailment like cancer.

Much effort has been channeled towards building automated systems for adding convenience and easing the financial cost of accessing some health services. Semantic Web has offered some technological support in driving this course through the design of domain specific ontologies. The health domain has much prospect from the inter- and intra-domain application of Semantic Web technologies because the health sector depends on the interoperability of information from many disciplines. But medical ontology developers can still use popular ontology languages like RDF/RDFS or OWL/OWL2 to model knowledge base. This research models all data collection nodes using some of these ontology languages. Another advantage of the semantic web is the power of inference making and reasoning over ontologically modeled knowledge base.

Therefore, we use semantic web rule languages to encode our rule systems for effective interoperability with the knowledge base.
Meanwhile, more algorithms have been designed for diagnosing ailments, and some of these algorithms are fashioned after mathematical, statistical, fuzzy and rule-based models. The Semantic web is a web of meanings, having more affinity for rule-based models. It enables machines as well as people to understand, work and share data in an autonomous way. As a result, taxonomies, metadata, classifications, context and ontology have been the basic building blocks of the Semantic Web [7]. A combination of a formal ontology for the medical domain with a fine-grained contextual inference making and reasoning algorithm is an exceptional tool in incorporating autonomous (health) systems into applications in the areas of biomedicine, life sciences clinical research, health care, biological sciences and translational medicine.

This research in progress argues that employing the use of some semantic web technologies in ST algorithm will yield a higher precision and an inference making medical diagnostic reasoning system. It therefore develop a framework that can be harnessed perceptively and constructively in realizing an adaptive intelligent personal agent (IPA). IPAs are potential tools for closing the gap of differentiated access to qualitative healthcare delivery among citizenry, especially in developing nations. The enhancement carried out on the ST algorithm enables the use of semantic web technology, and as well to add some mathematical models whose overall aim is approximation of diagnostic tasks. This also helps us deploy a monitoring agent in the framework for a continuous gathering of data that will later aid diagnostic process. The resulting IPA will monitor and provide diagnostic role; specifically, helping patients with cancer diagnose and monitor themselves. Given the demand of image processing, this research will not consider patient's mammogram as an input parameter.

II. Related Work

In the literature, we have found out that there are medical experts systems and their underlying algorithms. We have divided this section into three subsections: (i) medical expert systems, (ii) medical diagnostic reasoning algorithms, and (iii) ST Model, acclaimed medical algorithms with higher approximation.

A. Medical Expert Systems

An expert system is a computer application intended to make reasoned judgments or give assistance in a complex area in which human skills are fallible or scarce. In [8], the author cited that they are computer systems that operate by applying an inference mechanism to a body of specialist expertise represented in the form of 'knowledge'. They are employed as decision support systems, and have some approaches to implementation includes; rule-base (MYCIN and PROSPECTOR), data-base approach, descriptive method (INTERNIST and CADUCEUS), and Causal Network method. Authors in [9] developed a model for expert systems for the diagnosis of human diseases. The expert system carries out its diagnoses by organizing symptoms into three groups namely Key group(Kg), Sub group(Sg) and Unexpected(Ue). Kg is a group of symptoms whose presence is necessary and sufficient to confirm the diseases whereas the presence of Sg is not sufficient and it is a subset of Kg. [10] designed ASTHMA, an expert system for the diagnoses of asthma. They combined some machine learning algorithms such as Context sensitive auto-associative memory neural network model (CSAMM), Back-propagation model, C4.5 algorithm, Bayesian Network, Particle Swarm Optimization to realize their design. Ex-Dr Verdis is an integrated expert system that combines an advanced medical information system containing various medical services supported by information technologies, with ES capabilities in a single system [11]. Heart Disease Program (HDP) is a medical expert system that enables physicians to enter patient's symptoms, laboratory tests, and physical examination. It then generates clinical data that support the diagnoses of heart disease [12]. These are but few out of dozens of medical expert systems that abounds.

B. Medical Diagnostic Algorithms

Different models (algorithms) and approaches are being used for diagnostic problems, though depending on the kind of ailment to be diagnosed. A particular algorithm might be suitable for one ailment and may not suitable for another. Hence, in this section, we enumerate some of these algorithms and their weaknesses. Scheme inductive reasoning (also known as forward thinking) is based on adding characteristics of the syndrome to narrow the list of potential diagnoses. In scheme inductive reasoning, schemes are drawn to resemble that of road maps. It helps clinicians break down information into chunks, storing them in their memory and then retrieving them subsequently for problem solving task [13]. Pattern recognition is employed in machine learning for assigning some outputs to some inputs base on the coordination of a given algorithm [14]. Hypothetico-deductive reasoning involves the self-reflection and informed clinical decision making process of generating and testing hypotheses in association with the patient's presenting symptoms and signs [15]. Forward chaining system, includes writing rules to manage sub goals. Whereas, backward chaining systems automatically manage sub goals [16]. Forward reasoning is efficient and fast, backward reasoning can be employed to resolve the conflict between two competing hypotheses. A combination of the two reasoning method – backward and forward – with increased experience leads to increased coordination of hypothesis and evidence [17].
Parsimonious Covering Theory (PCT) works on the basis of associating a disorder to a set of manifestations. It uses

two finite sets (disorders & manifestations) to define the scope of diagnostic problems [18]. Certainty Factor (CF) model is used for managing uncertainty cases in a rule based system [19] and can be interpreted as measures of change in belief within the theory of probability. Bayesian networks are oriented acyclic graphs consisting of nodes (circles), which represent random variables; arcs (arrows), which represent probabilistic relationships among these variables [20] and this helps in dealing with uncertainties. However, Bayesian medical reasoning depends on utilization of conditional probabilities as a priori probability function and possibilities.

In fuzzy logic, linguistic variables are used to represent operating parameters in order to apply a more human-like way of thinking [21]. One of the main factors affecting fuzzy logic model performance is data clustering for membership function generation. The last of these algorithms is the ST Model adjured to be the most approximate [22]. We dedicate the next section to discuss it; for we seek to improve on it.

C. *The ST Model*

In [22], the authors described the approach for medical diagnostic reasoning based on ST Algorithm model which was earlier introduced by [23]. In their work, they adduced the fact that most of the algorithms discussed in the previous paragraph are lacking accuracy in their diagnostic approximation result. Hence, they showed that their approach of using ST algorithm in medical diagnostic reasoning yields an approximate reasoning model.

The ST Model describes a cyclical process which uses the logical inferences of abduction, deduction, and induction procedures in arriving at its reasoning task. The algorithm involves a bottom-up and recursive process using its four stages of logical inferences (abduction, deduction, and induction). Figure 1 shows the cyclic flow of these four stages of the ST model. The model adopted a two-layered entity mapping in order to model a simplified knowledgebase representation of diagnosis and symptoms. Given the high precision power in diagnostic reasoning of the ST model, this positions it for consideration in our research, as the medical diagnostic algorithm to be used in diagnosing cancer. The four modules in ST algorithms are listed as follows;

a) **Abduction**: Abduction is often described as inference to the best explanation. It involves determining all likely diagnoses related to the reported symptoms. The overall aim of this module is to get all the diagnosis related to some given symptoms. And all diagnosis gotten are stored in a data structure as diagnoses to be elicited. This list of diagnoses elicited is passed on to the deduction module.

b) **Deduction**: In this stage and for each likely diagnosis, all the expected symptoms of the diagnosis are drawn out based on a logical inference means. In addition, each of the known symptom of a diagnosis is assigned a thresh hold value, of which each expected symptom must be equal to or greater than it before it is included in the list of accepted symptoms of the diagnosis.

c) **Abstraction**: The process of mapping descriptive terms that are understood by patients onto well-defined symptom entities used in the knowledgebase is known as abstraction. No logical inference is done here, except for the elicitation of information from the patient. In a cyclic manner, this list is then passed back to the abduction and deduction stages for further refinement until the list of possible diagnosis are reduced to the minimum.

d) **Induction**: Induction involves matching the elicited symptoms with the expected symptoms for each likely diagnosis. At this stage, each of the likely diagnosis passed down from the cyclic process in steps 1-3 are then checked to see if they meet their diagnostic criteria.

III. THE PROPOSED MODIFIED ST MODEL

In this section, we present and anatomize our enhanced ST model. The modified model consists of the Abstract module and the three logical inference modules namely Abduction, Deduction and Induction. More so, the existing ST model data is not temporal, we add a monitoring module to the ST model so as to make it data-gathering procedure alive and as well make its data temporal – assigning symptoms, signs and testing to their timing during the stages of the ailment. Contrary to the ST model by [22], we model our data using ontological approach. The concept of semantic web rule language is employed for implementing our rule systems. Each of the rules in the coordinated rule system is annotated with a Certainty Factor (CF) value. This enables the assignment of weights to each of the rules and thereby determining the order of selection when fired.

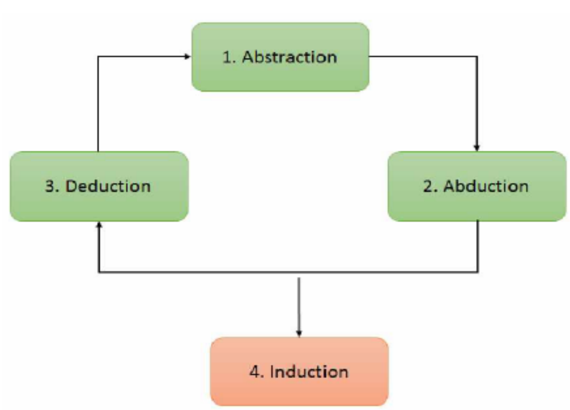

Figure 1: ST Model [22]

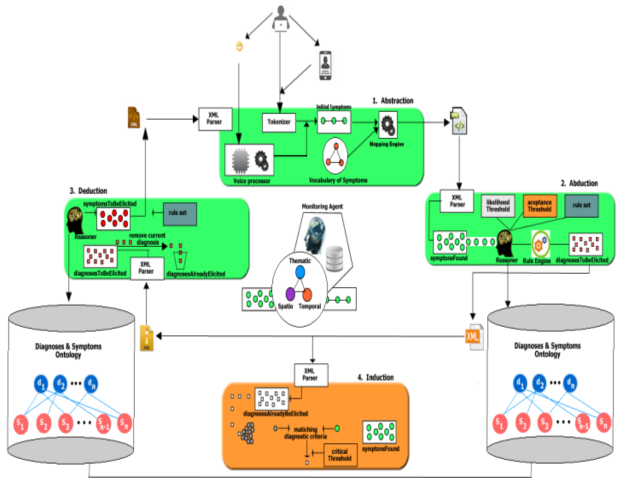

Figure 2: Modified ST Model

Figure 2 is an improvement on the ST model shown in Figure 1. The following subsections give a breakdown of each of the consisting components.

A. *Abduction*

Our abduction stage improves on the existing abduction module. Except that we enabled a semantic web based reasoning operation in the module. We propose the use of rule engine for this reasoning task. Both the abduction and the deduction stages here use this rule engines. And compose a rule system for aiding diagnostic reasoning task. A new parameter, ***acceptanceThreshold*** is added to the existing ***likelihoodThreshood*** parameter. This is to check if every deduction task passes a given acceptance value before we can conclude that it is correct. The abduction modules gets all diagnoses related to symptoms found, and reasons by hypothesis, studying facts and devising theory to explain it. The process of abduction: The whole process of abduction includes generation, criticism and acceptance of explanatory hypotheses.

B. *Deduction*

Deductive Reasoning is a process in which general premises are used to obtain a specific inference. A form of logic that identifies a particular item by its resemblance to a set of accepted facts. Deductive reasoning moves from general principle to a specific conclusion. It is inference by reasoning from generals to particulars. Deductions support their conclusions with TRUE result. They compute their results using heuristics. We modify the existing deduction module to be a rule-based deductive reasoning task. Hence, a coordinated rule system and a reasoner are added to semantically realize the deductive reasoning.

C. *Abstraction*

The process of mapping descriptive terms understood by patient onto a well-defined symptom entities modeled in the knowledgebase is known as abstraction. In this proposal, we seek to provide patients with a textbox for entering their entering descriptive terms of how they feel. Our natural language NL-query to Semantic Web SW-query model, then semantically matches their inputs against ontology of vocabularies in the knowledgebase. The modified abstraction module allows input to be in speech or textual. Patients may voice in their symptoms and this data will be processed by the voice processor.

D. *Induction*

It entails reasoning from the particular to the general. This may or may not be true. But it provides a useful generalization. At the induction state, we check if likely diagnosis meets diagnostic criteria. While Abduction and Deduction are termed clinical reasoning, Induction is termed clinical decision making. The induction stage in this modified ST model builds on the existing features of the existing ST model. Except that we develop a mathematical model for computing the ***criticalThreshold*** parameters, which is now added to calibrate and alert patient on the status of the ailment.

E. *Monitoring Agent*

The monitoring agent works continuously in the system. It is more like a daemon which logs events into a Spatial-Temporal-Thematic (STT) ontological database. The essence of agent is to be able to monitor development of the ailment in the patient's body, and then adequately signal the needed alert or logs necessary information that the diagnostic algorithm will mine data from it. Temporal information gathered is a clinical data that helps in tracking the progression of a disease in a patient with respect to time. Spatial information models data that relates with patient and its environment. Thematic data models concepts and terms used in clinical operations.

The monitoring agent consists of the following components: Event Selector, Event Monitor/Trigger, Data Gathering and Reasoner, STT Data Modeler, Storing STT, and STT Ontology.

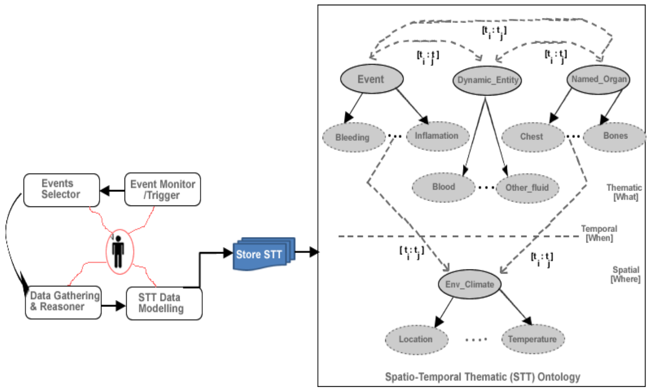

Figure 3 shows the interactivity among these components.

IV. The Knowledge Representation Model of the Medical Expert System

The proposed expert system discussed above is a rule base expert system. Though case based and model based expert systems are being promoted in other literatures, this research however seeks to exploit the use of predefined rules (rules jointly crafted by specialist medical expert in breast cancer). Figure 4 is a structural representation of the coordinated rules systems. The structure consists of four layers, and each layer comprising of facts – model with ontological language – and rules – modeled with choice semantic web based rule languages. Appended to these four strata is the monitoring agent knowledge model.

The first layer models the knowledge of the abstraction layer. Basically, there are three modularized data representation. These are a thesaurus modeled with OWL, patient's personal profile modeled with XML, and a rule for mapping descriptive terms of patients into a well-defined symptoms entities, modeled with RuleML. The second layer is a knowledge representation for the abduction phase. Knowledge representation at this phase comprises of the facts, modeled with OWL, and rule set for carrying out abduction, modeled with semantic web rule language (SWRL). The deduction module is the next phase for knowledge modeling representation. This phase has a rule set modeled in Jess rule language, and the fact modeled with OWL also.

Similarly, the induction phase also comprises of an ontological knowledge base and a rule set for induction, with the rule modeled with the Jena rules. The last component in this structured knowledge model is spatial-temporal-thematic ontological representation of the data generated during the monitoring process.

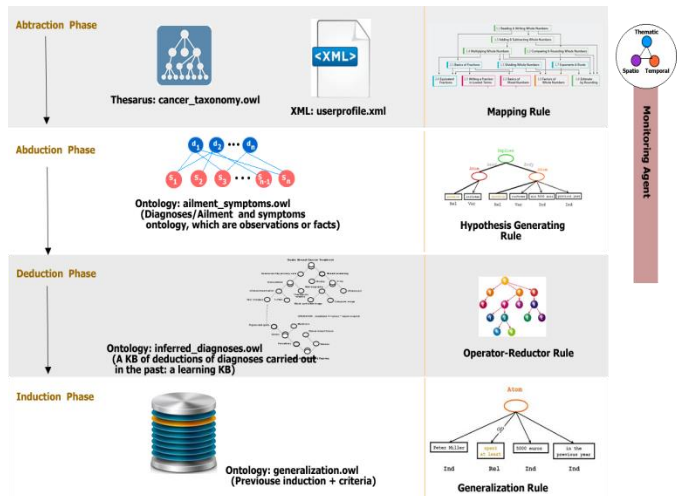

Figure 4: Knowledge Representation Model for the Modified ST Algorithm

V. CONCLUSION

In other to show a proof of the concept being argued in this proposal, some metrics are considered for testing the framework against the existing ST model. This will enable the result of this research to be placed side by side with the existing ST model for the purpose of result presentation. Our metrics include; we measure the accuracy of diagnosis process of the improved ST framework against the existing ST framework. It also assigns assurance/certainty value to every diagnosis listed in the output. The third metric computes the weight of monitory logs on overall diagnosis process, finally, the fourth metric implements original ST algorithm against modified ST algorithm.

References

[1] Malathi, A. and Santra, A. K. (2009). *Diagnosis of Lung Cancer Disease using Neuro-Fuzzy Logic*. CARE Journal of Applied Research (ISSN 2321-4090), pp. 6-9.

[2] Cassileth, B.R. and Deng G. (2004). Complementary and alternative therapies for cancer. Oncologist. Vol. 9, pp. 80-89.

[3] Smith, A. (2006). Proton therapy. Phys. Med. Biol., Vol. 51, pp. 491-504

[4] Vickers, A. (2004). Alternative cancer cures: 'unproven' or ,disproven'? CA. Cancer J. Clin. Vol. 54, pp. 110-118.

[5] Kramer, B. (2015). Retrieved on 19th June 2015, from http://www.cancer.gov/news-events/cancer-currents-blog/2015/cancer-prevention-research-importance

[6] Carla Pollastro, Carmela Ziviello, Valerio Costa, Alfredo Ciccodicola (2015). Pharmacogenomics of Drug Response in Type 2 Diabetes: Toward the Definition of Tailored Therapies? PPAR Res. 2015; 2015: 415149. Published online 2015 Jun 15. doi: 10.1155/2015/415149

[7] Mustafa Jarrar and Robert Meersman: Ontology Engineering -The DOGMA Approach. Book Chapter (Chapter 3). In Advances in Web Semantics I. Volume LNCS 4891, Springer. 2008

[8] Todd, B. S. (1992). An Introduction to Expert Systems. Technical Monograph PRG·95 ISBN 0-902928-73-2, Oxford University Computing Laboratory Programming Research Group 11 Keble Road Oxford OXI3QD England. Pp. 3.

[9] Patra, P.S. K., Sahu, D. P. and Mandal, I. (2010). An Expert System for Diagnosis of Human Diseases. *International Journal of Computer Applications*, Volume 1 – No. 13, pp. 71-73.

[10] Prasadl, BDCN, Prasad, K. P. E. and Sagar, Y. (2011). An Approach to Develop Expert Systems in Medical Diagnosis using Machine Learning Algorithms (ASTHMA) and a Performance Study. *International Journal on Soft Computing (IJSC), Vol.2, No.1 pp. 26-33.*

[11] Keles A. (2014). Expert Doctor Verdis: Integrated medical expert system, Turkish *Journal of Electrical Engineering & Computer Sciences, pp. 1032-1043.*

[12] Al-Ajlan, A. (2007). Medical Expert Systems HDP and PUFF. King Saud University College of Computer & Information Sciences Department of Computer Science, pp. 3

[13] Anderson, K. J. (2006). Factors Affecting the Development of Undergraduate Medical Students' Clinical Reasoning Ability. Medicine Learning and Teaching Unit, Faculty of Health Sciences, University of Adelaide, pp. 18.

[14] Umoh, U. A., Umoh, A. A., James, G. G., Oton, U. U., and Udoudo, J. J. (2012). Design of Pattern Recognition System for the Diagnosis of Gonorrhea Disease. *International Journal of Scientific & Technology Research (IJSTR) Volume 1, Issue 5*, pp. 74-79

[15] Kumar, S. P., Kumar, A. and Sisodia, V. (2013). *Clinical Reasoning and Sports Medicine-Application of Hypothetico- Deductive Model.* J Sports Med Doping Stud ISSN: 2161-0673 JSMDS, Volume 3 Issue 1

[16] Sharma, T, Tiwari, N. and Kelkar, D. (2012). Study of Difference between Forward and Backward Reasoning International Journal of Emerging Technology and Advanced Engineering (IJETAE), 2(10), pp

[17] Hardin L.E. (2002). Research in Medical Problem Solving: A Review. JVME 30(3), pp 227-232.

[18] Wainer, J. and Rezender, A. M. (1996). A Temporal Extension to the Parsimonious Covering Theory

[19] Heckerman, D. E. and Shortliffe, E.H. (1992) From Certainty Factors to Belief Networks. *Artificial Intelligence in Medicine*

[20] Gadewadikar, J., Kuljaca, O., Agyepong, K., Sarigul, E., Zheng, Y. and Zhang, P. (2010) Exploring Bayesian Networks for Medical Decision Support in Breast Cancer Detection. *African Journal of Mathematics and Computer Science Research Vol. 3(10), pp. 225-231, October 2010.*

[21] Torshabi, A. E., Riboldi, M., Pella, A., Negarestani, A., Rahnema, M. and Baroni, G. (2013). *A Clinical Application of Fuzzy Logic*. Fuzzy Logic – Emerging Technologies and Applications, pp. 1-18.

[22] Fernando, I. and Henskens, F. A. (2013). ST Algorithm for Medical Diagnostic Reasoning. Polibits pp. 23-29.

[23] M. Ramoni, M. Stefanelli, L. Magnani, and G. Barosi "An epistemological framework for medical knowledge-based systems" *IEEE Transactions on Systems, Man and Cybernetics, vol. 22, pp. 1361-1375, 1992.*