

A Knowledge Representation Model Based on Select and Test Algorithm for Diagnosing Breast Cancer

*¹Olaide N. Oyelade, ²Pam Bulus Dung and ³Najeem Owamoyo

¹Department of Computer Science, Ahmadu Bello University, Zaria, Nigeria

²Department of Computer Science, Federal College of Education, Pankshin, Nigeria

³Department of Computer Science, Federal College of Education, Zaria Nigeria

olaide_oyelade@yahoo.com | victoryagain16@gmail.com | najolp@yahoo.com

Submitted: 11-MAR-2020;

Reviewed: 23-MAR-2020;

Accepted: 28-MAR-2020

Abstract— There exist several terminal diseases whose fatality rate escalates with time of which breast cancer is a frontline disease among such. Computer aided systems have also been well researched through the use intelligent algorithms capable of detecting, diagnosing, and proffering treatment for breast cancer. While good research breakthrough has been attained in terms of algorithmic solution towards diagnosis of breast cancer, however, not much has been done to sufficiently model knowledge frameworks for diagnostic algorithms that are knowledge-based. While Select and Test (ST) algorithm have proven relevant for implementing diagnostic systems, through support for reasoning, however the knowledge representation pattern that enables inference of missing or ambiguous data still limits the effectiveness of ST algorithm. This paper therefore proposes a knowledge representation model to systematically model knowledge to aid the performance of ST algorithm. Our proposal is specifically targeted at developing systematic knowledge representation for breast cancer. The approach uses the ontology web language (OWL) to implement the design of the knowledge model proposed. This study aims at carefully crafting a knowledge model whose implementation seamlessly work with ST algorithm. Furthermore, this study adapted the proposed model into an implementation of ST algorithm an obtained an improved performance compared to the simple knowledge model proposed by the author of ST algorithm. Our knowledge mode resulted in an accuracy gain of 23.5% and obtained and AUC of (0.49, 1.0). This proposed model has therefore shown that combining an inference-oriented knowledge model with an inference-oriented reasoning algorithm improves the performance of computer aided diagnostic (CADx) systems. In future, we intend to enhance the proposed model to support rules.

Keywords— Semantic web, ontology, OWL, breast cancer, Select and Test (ST) algorithm, knowledge representation

1 INTRODUCTION

The World Health Organization's definition of health is not merely the absence of disease but the attainment of a state of physical, mental, emotional and social wellbeing (Omotosho *et al.*, 2005). Electronic health (eHealth), also known as eHealth, is the use of Information Communication Technology (ICT) in the delivery of healthcare services. The gap that exists in the use of technology in healthcare delivery between the developed and developing nations is being gradually closed by technological transfer and globalization of the entire world. This has translated into the development and deployment of varying eHealth applications. Most of these applications have been used in tackling problems of patient record registration and keeping of hospital equipment inventory, financial and billing systems, and pharmacy systems. More so, medical professionals have also employed the use of technology in advancing information sharing across borders (Rwashana and Williams, 2008), thereby having a great impact on the healthcare delivery.

Chronic diseases such as cancer (breast cancer), diabetes, heart disease and chronic obstructive pulmonary disease claimed the lives of 35 million people worldwide in 2006 alone (Johnston *et al.*, 2008), and some 80% of all non-communicable disease (NCD) in both low- and middle-income countries. In the Champlain LHIN region in Canada, 73% of people over the age of 12 already have at least one chronic disease and 89% of people engage in one or more activity that places them at risk of developing a chronic disease (Johnston *et al.*, 2008).

Just a decade after, in 2016 specifically, World Economic Forum (WEF) observed through a report that diabetes alone caused 1.43 million deaths in 2016, an increase of 31.1% since 2006 (Harris, 2017). This paper is particularly focused on a research that is based on Computer Decision Support Systems (CDSS) for diagnosis of breast cancer, a non-contagious disease which has generated much concern from the World Health Organization (WHO). In addition, WHO, World Cancer Report (2014) and de Martel (2008) enumerated some basic facts about cancer.

Our deep concern is: how can these chronic diseases be mitigated through advancement recorded in computer aided diagnostic solutions? Are there no sufficient knowledge representation model to ensure that knowledge-based algorithms are improved to curtail inadequate computational solution for improving breast cancer detection/diagnosis, treatment and monitoring? Although there are already CDSS applications that are deployed for managing automated healthcare services and their knowledge-base, however, an improvement on the existing CDSS and their knowledge base framework will rekindle hope in employing technology to improved health service delivery. Gerhard's (2004) knowledge management mode highlights solutions it. In addition to all these, the Semantic Web (and some other fields of computing) can be a very promising platform for developing knowledge management systems. This study therefore seeks an efficient knowledge representation model for medical domain, with specific interest in modelling breast cancer clinical guideline protocols.

Knowledge representation models are very important for attaining an optimal performance of algorithms which depends on them. But once a good knowledge representation model is crafted, the choice knowledge

*Corresponding Author

formalism or representation language becomes non trivial. There exist several knowledge modeling languages, query languages, rule encoding languages, and Semantic Web tools such as ontologies and multimedia (OntOMat), OntOMat set up and ontology evolution protocol OntOMat-SOEP, OntOMat-REVERSE (Maedche *et. al.*, 2012), OntOMat-CRAWL (Petridis *et. al.*, 2006), Ontobroker (Ontobroker 2011) and Protégé (Daniel *et. al.*, 2007). All of these can be used to create, integrate and disseminate knowledge among medical personnel, health facilities, the government agencies and patients.

The interest of this paper aimed at adapting our proposed knowledge representation model to use semantic web technologies. The Semantic Web may well provide a solution to how to acquire and manage large volumes of knowledge to develop truly intelligent problem solvers and address the brittleness of traditional Knowledge base systems. Clinical knowledge – Breast Cancer in particular - is characterized by its: Large volume (Miranda-Mena *et al.*, 2006); Incompleteness of data; and need for correct reasoning operations. If a situation necessitating that tacit (oncologist's mind) knowledge must be stored and organized in clinical protocols for use through knowledge elicitation and knowledge representation, then the Semantic Web Ontology languages already provide efficient ways of doing these.

In this paper, we propose a novel knowledge representation model which leverages on Semantic Web to efficiently allow an inference-based select and test (ST) algorithm to adapt easily to its knowledge-base for carrying out medical reasoning. The medical reasoning task the paper is focused on diagnosing breast cancer from signs and symptoms presented by user. The remaining part of this paper is organized as follows: section 2 presents some related works; section 3 presents the proposed knowledge representation model and a brief introduction of ST algorithm. Section 4 presents the application of Semantic Web in implementing our model; section 5 presents results and discussion; and in section 6, we conclude the paper.

2 RELATED WORK

Lanzola and Stefanelli (1991) described a specialized framework for Medical Diagnostic Knowledge Based Systems able to help an expert in the process of building KBSs in a medical domain. The framework is based on an epistemological model of diagnostic reasoning which has proved to be helpful in describing the diagnostic process in terms of the tasks by which it is composed of. Cecilia *et. al.* (2006) proposes a framework to analyse, classify and choose knowledge-based applications for diagnosis. It defines a three-dimensional space in which an application may be represented by a point, whose coordinates are defined on each of the 3 axes corresponding to the conceptual, the functional and the phenomenological dimensions of it.

Jiangbo *et. al.* (2008) presented an ontological knowledge framework that covers healthcare domains that a hospital encompasses from the medical or administrative tasks, to hospital assets, medical insurances, patient records,

drugs, and regulations so as to attain personalized healthcare. Reyes-Ortiz *et. al.* (2013) proposed a computational model of representation of medical knowledge to support decision-making task during a medical consultation in order to reduce the chance of misdiagnosis in general medicine. Although their representation is based on ontologies that provide a mechanism for structuring knowledge to become computer-understandable information, we observed that there was no clear indication of the algorithm used for diagnosis. Fernando and Henskens (2016) designed a simple table-like knowledge representation model for ST algorithm in diagnosing ailment.

Cedeno-Moreno and Vargas-Lombardo (2018) designed a model, which was validated with a data corpus of approximately 200 patient records, and targeted at extracting necessary elements in a text written in NL using NLP tools and with them create a knowledge base represented by one domain ontology and extract knowledge to help medical specialists. Boshnak *et. al.* (2019) proposed a knowledge modelling methodology to develop Patient Clinical Data (PCD) ontology. The PCD ontology model identifies the concepts and semantic types in the Electronic Health Records (EHR) clinical data. PCD ontology aims to represent clinical data to be accessible and usable by researchers and healthcare facilities users. The development of PCD ontology and the main components are formalized in the Ontology Web Language (OWL).

3 METHODOLOGY

3.1 THE REASONING STRUCTURES OF ST ALGORITHM

The ST Model (Oyelade and Adewuyi, 2019) 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). The model adopted a two-layered entity mapping in order to model a simplified knowledge base representation of diagnosis and symptoms. The four modules in ST algorithms are listed as follows;

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 is stored in a data structure as diagnoses to be elicited. This list of diagnoses elicited is passed on to the deduction module.

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 threshold 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.

Abstraction: The process of mapping descriptive terms that are understood by patients onto well-defined symptom entities used in the knowledge-base 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.

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.

3.2 BREAST CANCER KNOWLEDGE REPRESENTATION MODEL

Figure 1 illustrates the knowledge representation model proposed in this paper. The first layer in the stack is the knowledge representation for the abstraction module. In this layer, there are three nodes of knowledge representation: the thesaurus, the user profile file. The second layer is the knowledge representation for the abduction layer which captures the diagnosis-symptoms ontology file. The deduction module’s knowledge is represented in the third layer and represents the diagnosis ontology and the operator-reductor rule set (discussed in section C). The last layer models the knowledge base of the induction module which contains the induction ontology file.

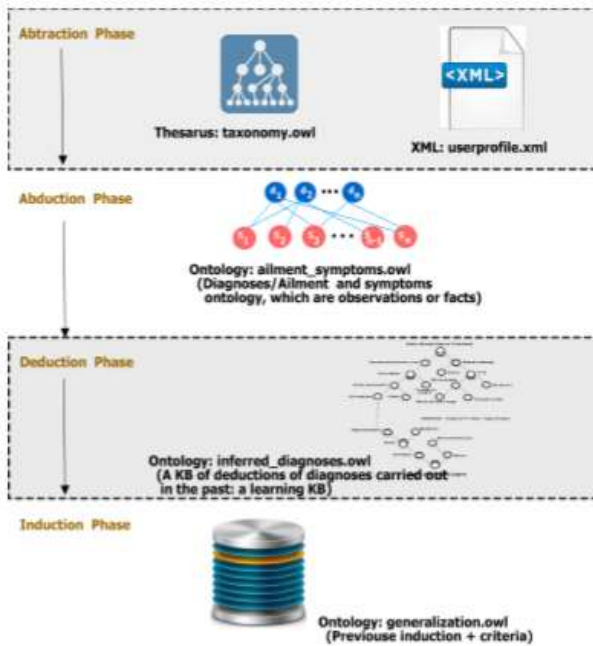


Fig. 1: The proposed knowledge representation model for supporting ST algorithm in diagnosing breast cancer

3.3 THE REDUCTION-OPERATOR RULE APPLIED ON THE DEDUCTION LAYER

Figure 2 captures the illustration of rule used to describe the effect of a Reduction-Operator pattern for reasoning on the knowledge modelled on the deduction layer. Recall that the aim of the deduction layer in ST algorithm is to get all the symptoms related to some given ailment or diagnoses (or types of cancer in the case of this research). Therefore, the reduction-operator rule is aimed at creating a connection between the goal and fact tree depicted in Figure 2. The basic form of an OPERATOR (extend the fact tree) is $A \Rightarrow EXP$. The basic form of a

REDUCER (extend the goal tree) is $EXP \Rightarrow A$. Steps for deduction:

- Initialize the goal and fact trees to the given expressions.
- If the termination check succeeds, exit.
- Use the domain-specific control strategy to select one of the literal nodes and an OPERATOR or REDUCER whose pattern matches this literal node.
- Apply the selected rule, extend the goal or fact tree, and go to (2).

These steps were adopted from Nilsson (2010). In a way to demonstrate how the connection is done, Figure 2 captures a simple rule: the OPERATOR, which states that mutations \Rightarrow lumps ^ nipple invasion (that is, if mutation can lead to lumps in breast and nipple discharge), then connect the mutation node from the fact tree to the lump and nipple discharge node from the goal tree. Once reasonable connections are created between the fact and goal tree, then the reasoner may draw conclusion by returning a true and a list of symptoms reasoned to be related to the supplied ailment or diagnoses (or a type of cancer).

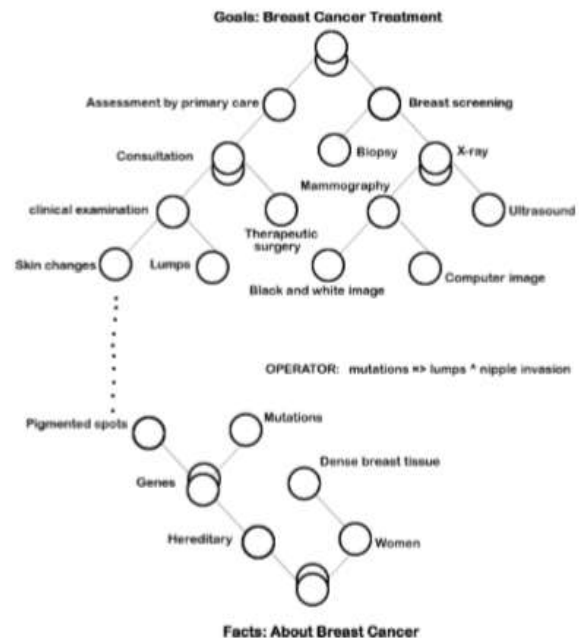


Fig. 2: A simple goal and fact trees for deducing breast cancer.

4 APPLICATION OF ONTO-KB ALGORITHM IN BREAST CANCER KNOWLEDGE-BASE

In this section, a detail of the proposed model presented in Figure 1 is given. Also, an implementation of the knowledge model using web ontology language (OWL) and Protégé is shown. In addition, curated clinical guidelines on breast cancer through the assistance of an Oncologist at the Ahmadu Bello University Teaching Hospital, Shika-Nigeria, was carried out. This became necessary to enable us capture all necessary concepts and because domain specific ontology development is usually done with the assistance of experts from that field (breast cancer for example). Hence, there are four sub-sections detailing the knowledge model for each layer of ST algorithm.

4.1 ABSTRACT LAYER KNOWLEDGE REPRESENTATION (THESAURUS/LEXICON)

In Figure 3, 4 and 5, we capture a thesaurus/lexicon of acceptable terms used by oncologist as clinical guidelines protocol for breast cancer. Observe that the words or terms are arranged in a hierarchical pattern denoting a semantic relationship existing among them. This forms a vocabulary that inputs into the ST algorithm are matched with to ensure that such inputs align with the terms required for the diagnostic system to effectively reason. The purpose of the knowledge represented at the abstraction layer is to formalize inputs to the ST algorithm.

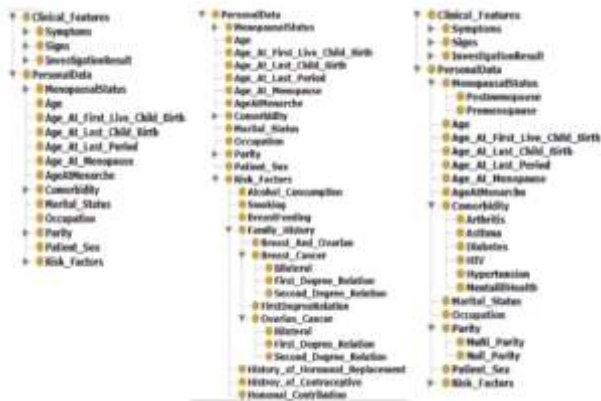


Fig. 3: Breast Cancer Thesaurus (expanded view of patient data that increases risk factors)



Fig. 4: Breast Cancer Thesaurus (Expanded view of signs or manifestation)

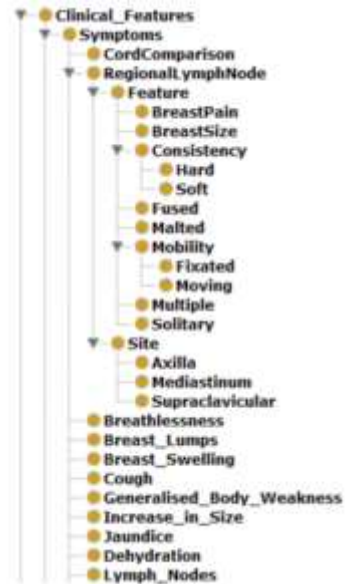


Fig. 5: Breast Cancer Thesaurus (expanded view of symptoms)

4.2 KNOWLEDGE REPRESENTATION OF ABDUCTION LAYER

The abduction module of ST algorithm is aimed at determining all likely diagnoses related to the reported symptoms. Hence it requires the knowledge representation of the abduction layer in Figure 1 to achieve this. In Figure 6, a model of all the necessary facts or information needed to aid the reasoning of the abduction module of ST. Naturally patients are often expected to present symptoms to medical experts, while the expert intuitively come up with signs and even manifestations patient might not have noticed. Figure 4 therefore was modelled to capture this event by classifying the symptoms and signs of breast cancer on the left and right respectively. Under symptoms (OWL class) there are some examples of symptoms, while under signs examples are also stated.



Fig. 6: Knowledge representation for the deduction layer

4.3 DEDUCTION LAYER KNOWLEDGE REPRESENTATION

The deduction knowledge representation discussed in this section works with the deduction algorithm of ST described in section 3. This section of ST algorithm aims at connecting a goal to a fact to infer that a disease is diagnosed. Hence the illustration of Figure 7 for the deduction layer of our knowledge representation model captures this. The connection of both the goal and fact segments largely depends on algorithmic solution.



Fig. 7: Knowledge representation of the deduction layer

4.4 INDUCTION MODULE KNOWLEDGE REPRESENTATION

The knowledge representation for the induction layer is aimed at induction module of ST algorithm which checks if each of the likely diagnoses correlates with the diagnostic criteria model in induction knowledge layer shown in Figure 8.

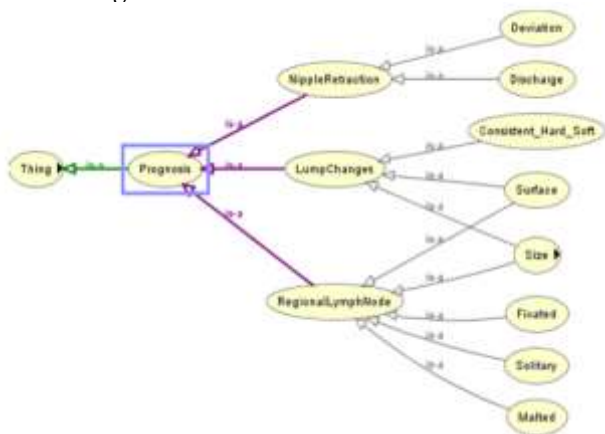


Fig. 8: The induction module knowledge representation

5 IMPLEMENTATION, RESULTS AND DISCUSSION

The implemented of ST algorithm as described in Oyelade *et. al.* (2018) and Oyelade & Adewuyi (2019) was adapted for use in this paper to verify the usability and test the effectiveness of our proposed knowledge representation model. Figure 10 captures an interface of the implementation where the first box on the left hand side prompts user to respond to some questions. Then based on the input by the user, the knowledge representation of the abstraction layer (lexicon) is then used to derive a list of acceptable tokens to be passed into ST algorithm. The abduction, deduction and induction sub-modules of ST algorithm in turns access their knowledge representation layer to complete their tasks.

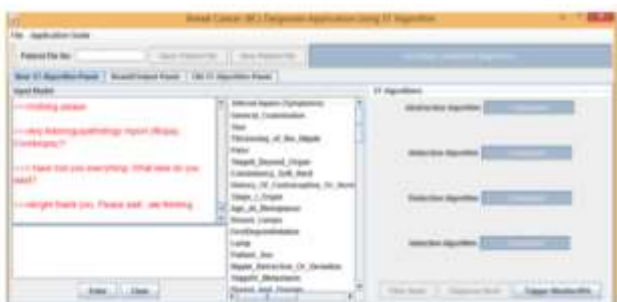


Fig. 10: Application of knowledge representation model to ST-algorithm for diagnosing breast cancer

Selected performance metrics were measured against the ST algorithm for comparing improvement in performance obtained by our proposed knowledge representation model. Results obtained showed that the performance gain in accuracy of the enhanced ST due to the knowledge representation model attained 23% while area under curve (AUC) shows that the point falls close to the top-leftmost section of the curve of (0, 0.42). The sensitivity and specificity values of 1 and 0.54 are indications of how the enhanced ST algorithm has the tendency to detect cases with breast cancer from those without it. Table 1 summarizes our observations:

Table 1. Breakdown of results of applying the proposed knowledge representation model to ST and enhanced ST algorithms

Parameters	ST algorithm implementation	Enhanced ST algorithm implementation
Result of Diagnoses	Breast Cancer	Breast Cancer
Accuracy of Diagnoses:	59.45%	67.36%
Precision of Diagnoses:	0.031	0.5
Time of Execution:	0:0:187	0:5:310:310
No. KB Items Used:	10 Items	5745 items
Sensitivity	0.42	1.0
Specificity	1.0	0.5125
AUC	0, 0.42	0.49, 1.0
Symptom Weight:	Not applicable	0.54347

6 CONCLUSION

This paper proposes a knowledge representation model for improvement performance on ST algorithm. The proposed model became necessary to augment for need for formalism of knowledge for medical reasoning algorithms. However, the proposed model is limited the manual approach for ontology building. In future, we intend to adapt rules and ontology learning to improve this study

REFERENCES

Boshnak H., Abdelgaber S., Yehia E., and Abdo A. (2019). Ontology-Based Knowledge Modelling for Clinical Data Representation in Electronic Health Records. *International Journal of Computer Science and Information Security*, 16(10):68-86 · April 2019

Cedeno-Moreno D., Vargas-Lombardo M. (2018). An Ontology-Based Knowledge Methodology in the Medical Domain in the Latin America: the Study Case of Republic of Panama. *Acta Inform Med.* 2018 Jun;26(2):98-101. doi: 10.5455/aim.2018.26.98-101.

Dang J., Hedayati A., Hampel K., and Toklu C. (2008). An ontological knowledge framework for adaptive medical workflow. *Journal of Biomedical Informatics* 41(5), 829-836

de Martel C, Ferlay J, Franceschi S. (2008). Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *The Lancet Oncology*, 13(6) pp 607-615

- Fernando D.A.I.P. and Henskens F.A. (2016). ST Algorithm for Medical Diagnostic Reasoning. *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, Studies in Computational Intelligence*, 653(1), 75-84
- Gerhard F. (2004). Knowledge Creation, Integration, and Dissemination in Design Communities. Center for LifeLong Learning & Design (L3D) Retrieved on 21st October 2017 from <http://www.cs.colorado.edu/~l3d/>
- Harris B. (2017). The world's biggest killers? Diseases linked to human behaviour. Retrieved on 13th April 2018, from <https://www.weforum.org/agenda/2017/10/the-world-s-biggest-killers-diseaseslinked-to-human-behaviour/>
- Johnston S, Liddy C., Ives M. S., and Soto E. (2008). Literature Review on Chronic Disease Self-Management. Retrieved on January 29, 2016 from <https://www.livinghealthynortheast.ca/Portals/0/Documents/Literature%20Review%20on%20Chronic%20Disease%20Self%20Management.pdf>
- Lanzola G and Stefanelli M (1991). A specialized framework for Medical Diagnostic Knowledge Based Systems. *Proc Annu Symp Comput Appl Med Care*. 1991:1025-31.
- Maedche A., Motik B., Stojanovic L., Studer R., and Volz R. (2012). Managing Multiple Ontologies and Ontology Evolution in Ontologging. *Intelligent Information Processing: IFIP 17th World Computer Congress*
- Miranda-Mena T. G, Benitez U. S. L., Ochoa L. J, Martinez-Bejar R., Fernandez-Breis J. T., and Salinas J. (2006). A knowledge-based approach to assign breast cancer treatments in oncology units . *Elsevier: Expert Systems with Applications*, 31(3), pp. 451–457.
- Nilsson, N. J. (2010). *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. Published by Cambridge University Press, New York, pp. 580.
- Omotosho, A., Emuoyibofarhe. O.J., Adegbola, O. (2011). " ICTs in Health Care Delivery Systems – A Framework for Developing Nations" .In proceedings of International Conference on ICT for Africa (ICT4A), Covenant University and Bells University of Technology, Ota , Nigeria , 23- 26, March 2011. Vol III. Pp 95 - 100
- OntoBroker (2011). Retrieved from <https://www.w3.org/2001/sw/wiki/OntoBroker> on 26th March 2020
- Oyelade O. N. and Adewuyi S. A. (2019). ONCObc-ST: An Improved Clinical Reasoning Algorithm Based on Select and Test (ST) Algorithm for Diagnosing Breast Cancer. *Current Research in Bioinformatics*, Volume 8, 2019, DOI : 10.3844/ajbsp.2019.1.13
- Oyelade O.N., Obiniyi A.A., Junaidu S.B., and Adewuyi S.A. (2018). ST-ONCODIAG: A semantic rule-base approach to diagnosing breast cancer base on Wisconsin datasets. *Informatics in Medicine Unlocked*, Volume 10, 2018, Pages 117-125
- Petridis K., Anastasopoulos D., Saathoff C., and Timmermann N. (2006). M-OntoMat-Annotizer (Eds.): Image Annotation Linking Ontologies and Multimedia Low-Level Features. *Knowledge-based Intelligent Information and Engineering Systems: 10th International Conference, KES 2006, Bournemouth, UK, October 9-11, 2006: Proceedings* (pp.633-640)
- Reyes-Ortiz, J.A., Jiménez, A.L., Cater, J., Meléndez, C.A., Márquez, P.B., & García, M. (2013). Ontology-based Knowledge Representation for Supporting Medical Decisions. *Res. Comput. Sci.*, 68, 127-136.
- Rubin D.L., Noy N. F., and Musen M.A. (2007). Protégé: A Tool for Managing and Using Terminology in Radiology Applications. *J Digit Imaging*. 2007 Nov; 20(Suppl 1): 34–46.
- Rwashana, S. A., and Williams, W. D. (2008). Enhancing healthcare delivery through ICTs: A case study of the Ugandan immunization system. *International Journal of Education and Development using Information and Communication Technology (IJEDICT)*, 4(2):144-158.
- World Cancer Report (2014). IARC Nonserial Publication, Albert Thomas, France, pp. 630.
- Zanni-Merk, C., LeGoc M., Frydman C. (2006). A conceptual framework for the analysis, classification and choice of knowledge-based diagnosis systems. *International Journal of Knowledge-Based and Intelligent Engineering Systems* 10(2),113-138