

Original Research Paper

# ONCODIAG Select and Test (ST) Algorithm: An Approximate Clinical Reasoning Model for Diagnosing and Monitoring Breast Cancer

O.N. Oyelade, A.A. Obiniyi and S.B. Junaidu

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

## Article history

Received: 24-10-2019

Revised: 24-12-2019

Accepted: 01-02-2020

## Corresponding Author:

O.N. Oyelade

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

Email: olaide\_oyelade@yahoo.com

**Abstract:** Clinical reasoning involves coordinated thinking strategies targeted at gathering and analyzing relevant information to arrive at a high-quality clinical solution to patients' medical problems. The correctness and proper choice of clinical reasoning approach chosen by a clinician or machine affect the accuracy and acceptability of the diagnosis; moreover, another challenge that usually surfaces during clinical reasoning is the burden of reasoning out what is known as missing data. This implies that failure to measure-up to the requirement of an acceptable clinical reasoning procedure and inability to gather enough data will have adverse impact on patient's health. A major motivating factor necessitating research of models for automation and formalization of in clinical reasoning is due to its sensitivity and complexity. Although different formalism now exists, such as approximate reasoning and use of models or logical inferences, to curtail the issue of incompleteness of data and as well attain a high-quality clinical problem solving. However most of these formalisms have the drawback associated with their underlying methods of approximation of reasoning structures that usually mar the procedure and result of clinical problem solving process. The Select and Test (ST) algorithm offers close solution to the problems earlier stated. This paper presents an improved model of ST algorithm for clinical diagnosis and monitoring. This paper is focused on redesign of the ST model and the formalization of the approximate clinical reasoning. The redesigned ST model incorporates a monitoring module and mechanism for interacting with an ontology-based knowledge store. Specifically, the ST algorithm was modified to accommodate these additions to the ST model. Data was sourced through a retrospective study on breast cancer patients' records and through the administration of questionnaire at the Ahmadu Bello University Teaching Hospital (ABUTH) Zaria, Nigeria. In addition, Wisconsin datasets were used to test the resulting formalism. Empirical analysis of breast cancer diagnosis using the proposed model revealed that the accuracy of 88.72% was achieved. Similarly, a sensitivity of 1.0, specificity of 0.51 and an ROC point of (0.49, 1) was also attained. This paper presents an inference model and its approximate reasoning representation to diagnose and monitor the presence of a disease.

**Keywords:** Clinical Reasoning, Model, Select and Test Algorithm (ST), Intelligent Agent, Medical Diagnosis, Monitoring Agent, Ontology, Rule Sets, Breast Cancer, Clinical Decision Support Systems, Diagnosis Decision Support System

## Introduction

Clinical reasoning which is often associated with machine reasoning is also considered a branch of artificial intelligence (Kishan *et al.*, 2012). Human aided

clinical reasoning is a routine and delicate task that requires clinicians to reason out likely diagnosis base on some facts provided by patient and/or other sources. Usually, while the older clinicians rely on experience built over years to deal with the challenge of clinical

reasoning, the younger ones are constrained to the large medical literature and theory taught in medical school. In fact, clinical teachers are often unfamiliar with the complex body of literature; and they tend to rely mainly on their practical knowledge as clinicians (Audétat, 2011). While a blend of experience and medical literature or theory may help to sustain medical diagnoses based on human reasoning, the need for supportive roles of machines through machine reasoning *is* becoming glaring. Aiding clinical diagnosis through machine reasoning is characterized by unavailability of optimal and accurate reasoning models and formalism for approximate reasoning techniques. To tackle the problem of inaccurate models, designing models that can bridge the gap between theory and the reality of clinical practice will help clinicians to gain a better understanding of problems (Coiera, 2003). Application of such models is known as Clinical Decision Support Systems (CDSS), whose roles include: Information management, diagnoses and patient-specific consultation (Kawamoto *et al.*, 2005), have improved clinical practice by 69% of trials (Shortliffe, 1987). Though these models can be categorized into fuzzy logic, statistical, data-driven, mathematical, rule-base, inference model, case-based, knowledge-based problem-solving models, structural (symbolic) model, probabilistic, statistical pattern classifier, production model and prognostic models (Abu-Hanna and Lucas, 2001). This paper is an improvement on an inference model known Select and Test (ST) (Fernando and Henskens, 2013; Fernando and Henskens, 2016a; Ramoni and Stefaneli, 1992), similar to clinical diagnostic reasoning models are characterized by dual processing (Monteiro and Norman, 2013).

Approximate clinical reasoning is necessitated by the *presence* of uncertainty and missing data which usually characterizes medical reasoning, clinical data and diagnosis procedure. Different techniques such as probability theory and multivalued logic have been used for approximate reasoning. These techniques are further classified into formalism such as fuzzy logic or argumentation systems, probabilistic reasoning, belief functions (Dempster-Shafer theory), possibility theory, certainty factor and Coherence inductive probabilities and Bayes theorem. (Fernando *et al.*, 2013). Most of these techniques are limited by their different approaches of approximate reasoning. Fuzzy logic is limited by its implication operators which does not sufficiently model medical diagnoses procedure; belief functions (Dempster-Shafer theory) and Coherence inductive probabilities are limited by their inability to capture the positive and negative relationship exists between manifesting symptoms *on a* diagnosis; though Certainty factor tackles this limitation, it is however, unable to deeply describe clinical diagnoses procedure; though Bayesian networks are a powerful and sound formalism that allows reasoning under uncertainty (Bayesian networks), however, such models are limited by their

probabilistic disposition (assumption of conditional dependence and approximation of probabilities to one).

The focus of this paper is to bridge the gap between clinical reasoning solutions designed from the perspective of approximate reasoning and those which adopted design of reasoning models. The approach adopted in this paper is first to design an inference model which illustrates the flow of the proposed clinical reasoning procedure. Secondly, algorithms were written to further describe the applicability to the proposed inference model. Thirdly, our proposed inference model was further described using mathematical notations to present the relevance and possibility of achieving approximate reasoning based on the constraint associated with clinical reasoning. Meanwhile, this paper also included a monitoring module and an opening to relating the proposed model with an ontology knowledge base. The remaining part of the paper is organized as follows: Review of related literature was carried out and the limitations of such literature highlighted; this was followed by the improved ST inference model and the approximate clinical reasoning formalism; Furthermore, sections on the monitoring module and the proposed improve ST algorithm were presented; this was followed by the implementation section; and finally the result and evaluation cum discussion sections were outlined. We concluded the paper by restating its aim and our results.

## Relate Work

The use of multiple models or process, as proposed in this paper, for managing clinical reasoning process is already gaining research interests. The research in (Hosseinzadeh and Hosseini, 2017) identified six models of clinical reasoning which included hypothetic-deductive model, pattern recognition, a dual process diagnostic reasoning model, pathway for clinical reasoning, an integrative model of clinical reasoning and model of diagnostic reasoning strategies in primary care. They observed that only one model had specifically focused on general practitioners reasoning and then suggested that there is need for model of clinical reasoning that included specific features for scaling difficulties of clinical reasoning. However, going by the models they reviewed, one model may not be sufficient to meet this need, hence the necessity for intelligent hybridization of models. CDR (2018), the authors presented a dual-process model system and aimed at curtailing the limitation of one-process model discovered in (Hosseinzadeh and Hosseini, 2017). The dual-process model consists of Type 1 (Intuitive) processes are fast – used by experts most of the time. Type 2 (Rational) processes are slower, deliberate and more reliable and focus more on a hypothesis and deductive clinical reasoning (Hypothetical- Deductive Reasoning). They further stated that repetitive operation of Type 2 leads to Type 1. Type 2 processing can override Type 1 and Type 1 processing can override Type 2. This was

further reinforced by (Croskerry, 2009) when it stated that dual-process theory had emerged as the predominant approach, positing two systems of decision making, System 1 (heuristic, intuitive) and System 2 (systematic, analytical). The author proposes a schematic model that uses the theory to develop a universal approach toward clinical decision making. Properties of the model explain many of the observed characteristics of physicians' performance.

A similar argument was made in (Babette *et al.*, 2013) when the authors opined that system medicine as a specialized aspect of systems biology combines in an interdisciplinary approach all expertise necessary to decipher the human body in all its complexity. The new initiative aims for the integration of molecular data, anatomical, physiological, environmental and lifestyle data in a predictive model approach called the 'virtual patient'. The benefit of this multiple process model is that it will allow the clinician to predict and anticipate the optimal treatment for the individual patient. Application of the virtual patient model will allow truly personalized medicine. In a separate study, Daniel *et al.* (2016) hybridized models to form a cognitive engineering technique known as work domain analysis which was implemented to provide a framework for uncovering the relationship between diagnosis, complex health systems and theories of cognition and reasoning. The resulting model of diagnosis provides a comprehensive, novel perspective of the diagnostic process that offers a new foundation to formulate empirical inquiries about diagnosis and provides new avenues for the design and development of health information technologies, assessment strategies and diagnosis-centered simulation paradigms.

Stausberg and Person (1999), the paper presented a model-based approach to diagnostic reasoning in medicine. A process model is defined on the levels of static elements, dynamic elements and reasoning control. Static elements, facts, hypotheses and different types of disease knowledge, are identified and variations relevant for hypotheses generation are described. Dynamic elements correspond to actions, which in turn modify static elements, but are also controlled and started by the expressions of the static elements. The presented model could serve as a basis for an implementation in a model-based and process-oriented decision-support system. While the work in (Stausberg and Person, 1999) might demonstrate a form of multiple-process model, we observed that it is not. Hence, this paper is focused on improving the multi-inference (model) clinical diagnostic model in (Fernando and Henskens, 2013) named Select and Test (ST).

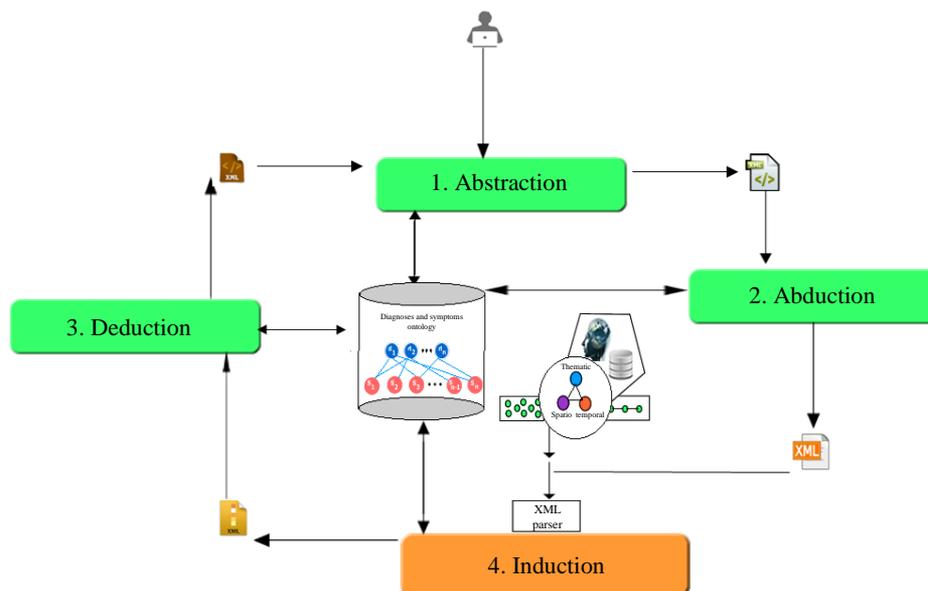
Building medical reasoning algorithms also requires a supportive knowledge base to aid the reasoning process. Formalizing such knowledge base in structured and machine meaningful ontology formalisms (Simple Knowledge Organization System [SKOS]) remains a research interest in artificial intelligence. Oftentimes, information that may aid reasoning may be silo in different databases in an unstructured format. One

relevant knowledge silo is the social media which users (and even patients) openly discuss relevant issues. Practical approaches like (Mike *et al.*, 2017; 2018) have used Patient Authored Text, statistical and linguistics methods to build a French Consumer Health Vocabulary on breast cancer to aid medical diagnostic reasoning. Similarly, authors in (Alsane *et al.*, 2018) formulated an approach for mining and further representation of medical knowledge from existing studies into formalism that will assist Clinical Decision Support System (CDSS) usage in-flight medical emergency management.

The fuzzy nature of reasoning and knowledge representation in medicine has raised the need for more research in development of approximate reasoning algorithms. For example, Rakus-Andersson (2009) argued that if the biological index of a patient has risen to risk-level, there may be no need for surgery. Hence, the authors proposed a means for evaluating such biological index through the incorporation of fuzzy set in the approximate algorithms. This, they argued, will detect patient's clinical symptom levels, pathologically heightened levels that indicate the presence of a disease possible to recover by surgery. Similarly, fuzzy-like approximate medical reasoning algorithm which was based on the logical inferences of (Ramoni and Stefanelli, 1992), was presented by Fernando *et al.* (2013). The approximate reasoning algorithm used the abduction, deduction and induction logical inferences in creating an inference model. Though the inference model was argued to achieve a desirable accuracy in clinical reasoning, it is however, limited by its approach to approximate reasoning and knowledge representation. This paper therefore seeks to improve on inference model proposed by Fernando *et al.* (2013).

## The Improved Select and Test (ST) Inference Model

In this section, the improved ST model, based on models of (Fernando and Henskens, 2013; Oyelade *et al.*, 2017a) and the supportive models of the intelligent personal agents with that of the monitoring agents are discussed. Meanwhile, a model typical to the application of the resulting improved ST model is also discussed in this section. Fernando and Henskens (2016b), the authors designed the ST model which consists of three inferences (abduction, deduction and induction) and input collection sub-module (abstraction). In this paper, we build on the model by enhancing its reasoning process through semantic web inference making tools and models. Furthermore, a monitoring agent was built into the modified model, with the redesigned model enabled for portability of knowledge representation in ontology forms. This model is illustrated in Fig. 1 and consists of five major components. These are the abstraction, Abduction, deduction, induction, monitoring agent and the ontology knowledge base.



**Fig. 1:** Modified ST model

The following are further explanation of each component of Figure 1.

### Abstraction

This sub-module is an input model that interfaces between the patient/user and the system. Whatever we can classify as symptoms, patient profile and manifestations are elicited, abstracted and stored as input source. The output of this module serves as input to the abduction module.

### Abduction

The overall aim of this logical inference making is to get all the diagnosis related to some given symptoms. It involves determining all likely diagnoses related to the reported symptoms. These differential diagnoses are generated by finding all related diagnosis with respect to each symptom in a set of **symptoms found**. The output of this module is becomes the input to the deduction sub-module.

### Deduction

For each likely diagnosis from the previous stage, all the expected symptoms of such diagnosis are drawn out based on the result of a logical inference process. The output of this module is sent as input to the abduction module, while receiving more inputs from the abstraction module.

### Induction

Abduction-deduction-abstraction forms a cyclic pattern which performs the process of clinical findings

– it carries out differential diagnoses. The result of the cyclic refinement of likely diagnoses is stored. The aim of inference by induction involves matching the acceptable criteria of each diagnosis already elicited for consideration, with their corresponding criteria according to standards or clinical protocols. This will enable the reasoning process isolate diagnoses with most likelihood of existence in patient, based on manifestations and symptoms presented – a process called clinical decision making. Meanwhile, it is at the induction sub-module that data gathered by the monitoring module is used to aid the process of clinical decision making.

### Monitoring Agent

The monitoring module is encased within the ST model. It monitors some activities/events in the patient and logs information gathered to support the task of the induction sub-module.

### Diagnoses and Symptoms Mapping

While the knowledge base of (Fernando and Henskens, 2016b) was modeled in tabular form, that of (Oyelade *et al.*, 2017b) was modeled as bipartite graph. This paper promotes enrichment of inference making in during the clinical reasoning process. Hence, an ontology approach for knowledge representation was crafted into the model in Fig. 1 that is improvement on the original model presented in (Fernando and Henskens, 2016b). Similarly, in Fig. 2 is a block diagram representing Fig. 1 and it demonstrates an outlined description and flow of the prosed ST model.

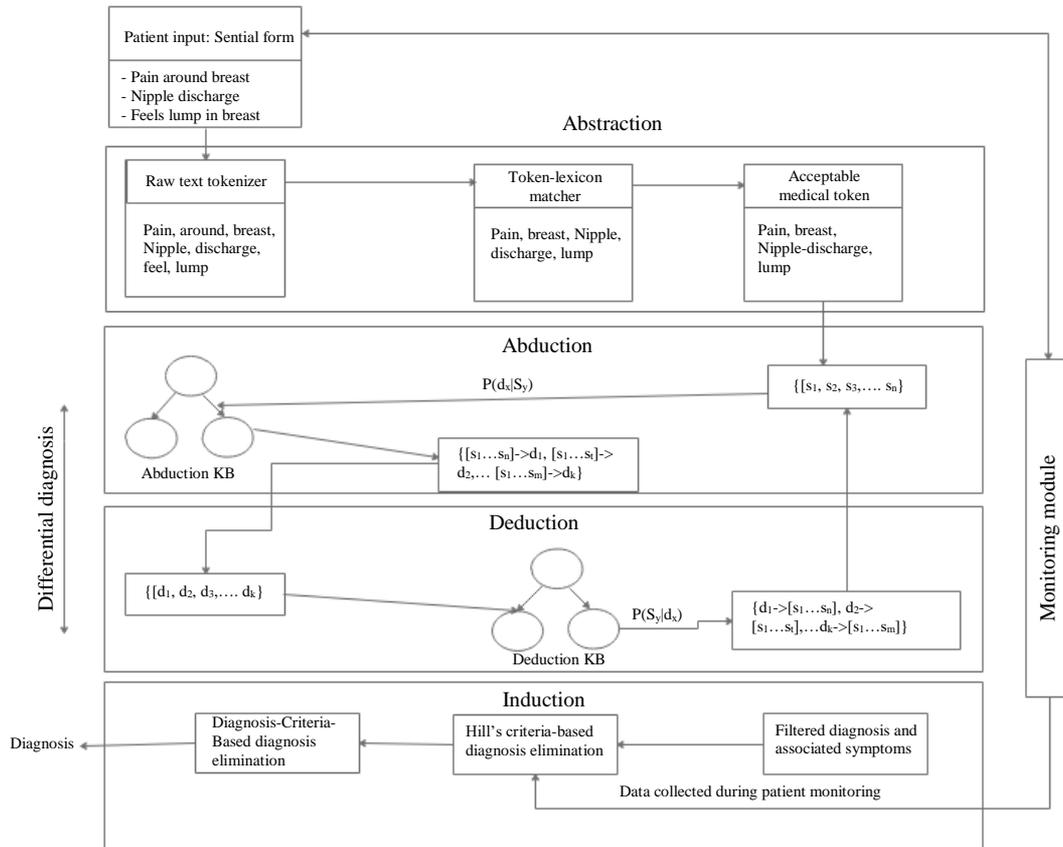
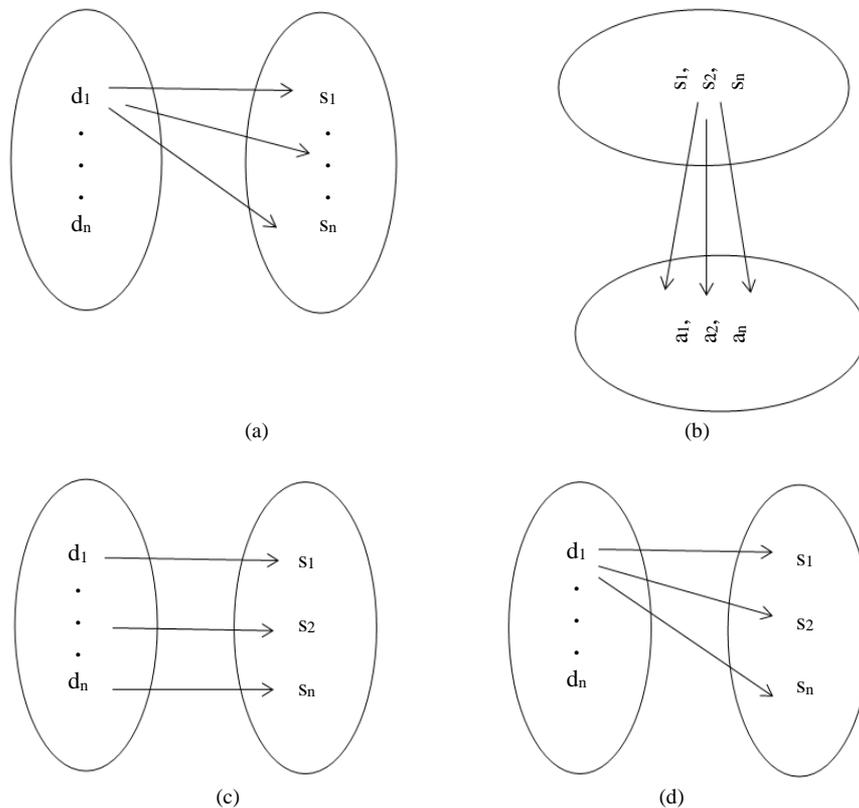


Fig. 2: Modified ST model

## Formalism for Clinical Reasoning for the Improved ST Algorithm

In this section, we present the formalism for approximation of the clinical reasoning process of the improved Select and Test (ST) algorithm. Our approach for approximate reasoning improves on that presented by (Fernando *et al.*, 2013). Note that relation differs from functional relations by their mapping strategy. While the later maps elements of one set ( $D$ ) onto exactly one element in the other ( $S$ ), the later permits multiple mapping onto the other set ( $B$ ). Contrary to (Fernando *et al.*, 2013) which simply model functional relationship, we introduced relation of sets  $D$  and  $S$  which denotes sets of Diagnosis and Symptoms respectively. Meanwhile, let's assume we have a set of diagnosis  $D = \{d_1, \dots, d_n\}$  and set of symptoms  $S = \{s_1, \dots, s_m\}$ . The clinical relationship that usually exists between set of  $D$  and  $S$  is that of dependent entities in  $D$  relating with independent entities in  $S$  respectively. Recall that the element of  $S$  may also relate with another  $d_i$  even though it relate with some  $d_k$ . Hence, we termed the each element of  $D$  as dependent variables while the elements in  $S$  are independent variables. Note that only a composition of some  $s_x$  will confirm the existence/occurrence of a particular  $d_y$ .

Furthermore, we introduce another set  $A$  which describes the likely attributes of each  $s_x$  in  $S$ , that relates to a particular  $d_y$  in  $D$ . Consider a  $d_y$  (breast cancer) and an attribute  $s_x$  (nipple discharge). The likely attributes of  $s_x$  relating to  $d_y$  are frequency of nipple discharge, duration of nipple discharge, associated manifestation of nipple discharge, commencement of nipple discharge. Figure 3a presents the illustration between sets  $D$ ,  $S$  and Fig. 3b captures the mapping of elements in  $S$  to  $A$ . Considering the forgoing, there exist a relation between  $D$  and  $S$  which we simply call a relation rather than functional relationship as claimed in (Fernando *et al.*, 2013). The justification for this renaming follows from definition of the two concepts as it applies to clinical reasoning. Functional relationship is considered to hold between  $D$  and  $S$  if  $d_y$  maps/relates with only one  $s_x$ , while a relation will require that  $d_y$  may map unto some  $s_x$ . Clinically speaking, if we have a set of all known symptoms and diagnosis, we will not obtain a functional relationship between  $D$  and  $S$ , but simply a relation. However, conjugations of all known symptoms, say  $KS$ , with respect to a diagnosis in  $D$  will result into a functional relationship between  $D$  and  $KS$ . Figures 3c and 3d illustrates the total relation  $D \mapsto S$  and partial function  $D \mapsto KS$ .



**Fig. 3:** (a) Mapping diagnosis to symptoms; (b) Mapping symptoms to attributes of symptoms; (c) Illustrating  $D \rightarrow S$ ; (d) Illustrating  $D \rightarrow KS$

The last paragraph described the set comprehension of elements that constitutes sets  $D$ ,  $S$  and  $A$ . However, it is clinically established that  $D$ ,  $S$  and  $A$  can be described in a sense of degree of occurrence or manifestation. For example, consider any arbitrary  $d_x$  in  $D$ , say  $d_x$  is breast cancer. Such  $d_x$  can be described as an ailment that has advanced to a particular critically (just as the advancement of breast cancer is usually described in stages, e.g., Stage1, Stage2a and so on). Similarly, it is necessary to quantify any symptom  $s_y$  observed to have been an effect to the cause  $d_x$ . Hence, we shall describe the specification of degree/quantity of  $d_x$  and  $s_y$  by  $Q(d_x)$  and  $Q(s_y)$  respectively. Furthermore, we describe the possible relationship that could exist between the quantity of  $Q(s_y)$  required to result into quantity of  $Q(d_x)$  to be  $Q(d_x|s_y)$ . Also the attributes of  $s_y$  may be described by some quantification that may be denoted by  $Q(s_y | [a_1 \dots a_n])$ . The outcome of  $Q$  with respect to any of its argument is bounded from 0 to 1, i.e.,  $[0, 1]$ :

$$Q(d_x|s_y) \wedge Q(s_y | [a_1 \dots a_n]) \leq Q(d_x) \quad (1)$$

The absence of any  $s_y$ , that is if  $Q(s_y) = 0$ , then there cannot be  $Q(d_x|s_y)$  and  $Q(s_y|[a_1 \dots a_n])$ . This then

necessitates equation 2 which establishes the presence of  $s_y$  for the confirmation of the diagnosis of  $d_x$  to be reinforced. So,  $d_x$  can be ruled out or confirm by:

$$P(d_x|[s_1 \dots s_m]) \in \{0,1\} \quad (2)$$

If Equation 2 evaluates to 0, then  $d_x$  is completely ruled out while its evaluation to 1 confirms the diagnosis. We observe that even if some  $s_y$  can evaluate  $P(\dots)$  into 1, Equation 3 must be established:

$$Q(d_x|[s_1 \dots s_m]) \geq 0.4 \ \&\& \ Q(d_x|[s_1 \dots s_m]) \leq 1.0 \quad (9)$$

Equation 3 approximates the clinical reasoning for confirming  $d_x$ . We define a relation  $\text{diagR}: D \rightarrow S$  or a functional relation  $\text{diagF}: D \rightarrow KS$ , where a  $ks_y$  in  $KS$  is an element which is a conjugation/agglomeration of some  $s_y$  that relates with a  $d_x$  in  $D$ . In equation 4, we present a mathematical function based  $\text{diagF}$  given some observed symptoms  $s_1 \dots s_n$  conjugated in  $S$ :

$$\text{diagF}(S) = d_x \quad (4)$$

But each  $s_y$  in  $S$  must first be computed using  $Q(s_y)$  which is interpretable in equation 5:

$$Q(s_y) = w_y \cdot c_y \quad (5)$$

Note that  $w_y$  denote the weight of  $s_y$  in  $d_x$ , while  $c_y$  denotes the observed severity of  $s_y$ . While severity  $c_y$  may be defined/quantified by a clinically acceptable scaling system in the range  $[0, 1]$ , weight  $w_y$  is given by Equation 6, where  $BI(d_x)$  is the biological impact of  $s_y$  with respect  $d_x$ . In addition,  $BI(d_x)$  may be defined as the degree of biological impact of  $s_y$  in aiding the existence or spread of  $d_x$ :

$$w(s_y) = BI(d_x) \quad (6)$$

Considering Equations 1-6 and some definitions made in this section, the function  $diagF(S)$  can further be defined by Equation 7:

$$diagF(S) = \sum_{y=0}^m Q(s_y) \quad (7)$$

The three (3) logical inferences earlier listed shall now be employed to describe the procedure for arriving at  $diagF(S)$ :

- i. **Abduction:** The aim of this logical inference is to generate all possible diagnosis  $d_x$  which may be clinically related with the  $s_y$  in  $S$ . The generated diagnoses are stored in  $D$ :

$$\begin{aligned} &\forall y, x: y \in \mathbb{N} \cdot (y \leq m) \text{ and } s_y \in S, \quad x \in \mathbb{N} \cdot (x \leq n) \text{ and } d_x \in D \\ &\text{if } (P(s_1) \wedge P(s_2) \dots \wedge P(s_m)) \cong P(d_x) \text{ then} \quad \text{rule 1} \\ &P([s_1 \dots s_m] | d_x) \end{aligned}$$

- ii. **Deduction:** contrary to the aim of abduction, the logical inference of deduction generates the lists of all possible symptoms that relates with each  $d_x$  in list  $D$ :

$$\begin{aligned} &\forall y, x: y \in \mathbb{N} \cdot (y \leq m) \text{ and } s_y \in S, \quad x \in \mathbb{N} \cdot (x \leq n) \text{ and } d_x \in D \\ &\text{if } (P(s_1) \wedge P(s_2) \dots \wedge P(s_m)) \triangleq P(d_x) \text{ then} \quad \text{rule 2} \\ &P(d_x | [s_1 \dots s_m]) \end{aligned}$$

- iii. **Induction:** Here, some criteria are used to narrow out unlikely diagnosis. Consider that  $Q(d_x)$  can be

calibrated along 0.0-1.0 such that the following thresholds of  $d_x$   $T(d_x)$  holds:

$$\begin{aligned} \text{Min}(T(d_x)) &= 0.4 && \text{Mild (early stage)} \\ \text{Avg}(T(d_x)) &= 0.6 && \text{Manageable (acute/non-} \\ &&& \text{aggressive stage)} \\ \text{Max}(T(d_x)) &= 1.0 && \text{Severe (chronic/advance stage)} \end{aligned}$$

We can then eliminate unlikely diagnosis in  $D$  through the rule:

$$\text{if } \left( \sum_{y=0}^m Q(s_y) \geq \text{Min}(T(d_x)) \right) \text{ then } d_x \quad \text{rule 3}$$

In conclusion, observe that it is not every  $s_y$  in  $S$  that may have a positive functional relationship with the  $d_x$  concluded in rule 3. Such  $s_y$  with negative functional relationship with  $d_x$  are implicitly excluded while those with positive functional relationship are absorbed in rule 3.

## Patient Monitoring module

Recall that the aim of this paper is to improve the ST model through the redesign of the model and addition of personalized diagnosis and monitoring. Therefore, this section presents two critical components for monitoring patients and gathering of information to aid the personalization of diagnoses process of the application of the expert system to the patient. First, we present the monitoring model, targeted for capturing relevant events related to patients' lifestyle, in Fig. 4 which has the knowledge representation schema to the right and the monitoring process to the left. Monitoring in this context means a procedure for making findings which could augment the plausibility of the diagnoses process of ST model. Basically, four sub-processes which includes event monitoring, event selector, data gathering and data formalizer, all controls the monitoring process proposed in this paper. Furthermore, the output data/knowledge of the four sub-process represented in Spatial-Temporal-Thematic (STT) format in ontology file. The event monitor receives information from the intelligent personal agent (discussed later) and then sends it to the event selector which appropriately identifies some characteristics of information desired. The data gathering and reasoning faculty conceptualizes the required data base on the event being captured. The last component then formalizes the data in a STT representation.

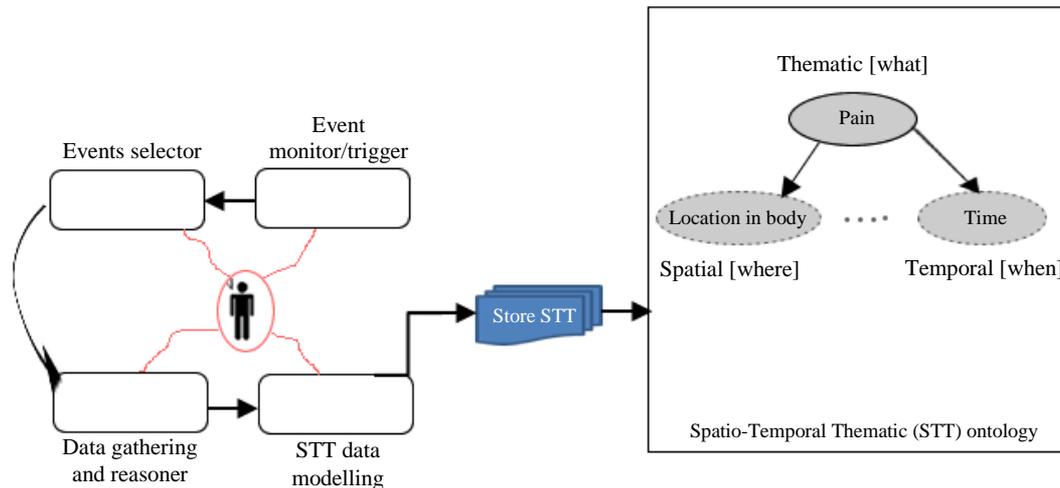


Fig. 4: Monitoring model

### ONCODIAG ST Algorithms

In this section, an improved version of algorithm in (Oyelade *et al.*, 2018) is presented. Algorithm 1 contains the demonstration of the models in Fig. 4; algorithm of the proposed model is captured in algorithms 2 through 5. The ONCObc-ST algorithm in Algorithm 6 combines the five algorithms described in Algorithms 2-5 and also lists the expected input and output of the algorithm. Within the body of the algorithm are sets initially declared as empty. Meanwhile, before the call to the *Abduction()*, *Deduction()* and *Abstraction()* sub-modules, the *SymptomsAlreadyElicited* and the *SymptomsFound* sets are populated by the *PresentingSymptoms* set which itself derives its elements through the input model in (Bayesian networks).

#### Algorithm 1: Monitoring algorithm

```

IPA_monitoring() {
    KB ← read IPA knowledge base
    Q ← ∅
    BodyFeelingsToMonitor ← questions to aid in
    gathering symptoms/feelings
    lifeStyleToCheck ← lifestyles related to disease
    considered
    while user is in session
        for lifeStyleToCheck as lifestyle{
            inquiry ← make inquiry from patient
            base on lifestyle
            reason on inquiry with respect to KB
            if reasoning is suggestive of disease
        then
            data1 ← ask(Q) write to
            monitoring module KB
        }
    end-while
    for each qi in BodyFeelingsToMonitor{
    
```

```

        add qi to Q
    }
    data2 ← ask(Q)
    data ← eventSelector(data1, data2)
    P ← ∅
    P ← conceptualizeData(data)
    STT ← formalizeData(P) // encode data as
    spatial-temporal-thematic
    Store STT in ontology
}

Ask(Q){
    R ← ∅
    for each q in Q
        r ← response to q
        add r to R
    return R
}
    
```

#### Algorithm 2: Modified abstraction module of ST algorithm

```

Abstract(){
    PatientProfile ← SymtomsFound
    For each s ∈ SymptomsTobeElicited {
        if s ∈ PatientProfile, add s to SymptomsFound.
        remove s from SymptomsToBeElicited
        Add s to SymptomsAlreadyElicited
    }
}
    
```

#### Algorithm 3: Modified abduction module of ST algorithm

```

Abduct(){
    ar ← AbductionRuleSet
    akb ← AbductionKB
    diagnoses ← ∅
    for each s ∈ P(s|d) in akb{
    
```

```

isCasual ← diagnosticReasoner(d, ar, s)
if isCadual
    diagnoses ← d
}
for each di in diagnoses
    where p (di | S) ≥ tD {
    if di E DiagnosesBeElicited AND
    di E DiagnosesAlreadyElicited,
    add di to DiagnosesToBeElicited
    }
}
}

diagnosticReasoner(d, ar, s){
    for each rulr in ar
    diagnoses ← invoke rule using s ad parameter
    if d in diagnoses
    return true
else
    return false
}
    
```

**Algorithm 4:** Modified deduction module of ST algorithm

```

Deduct () {
    for each d ∈ DiagnosesToBeElicited {
        dr ← DeductionRuleSet
        dbk ←DeductionKB
        symptom ← ∅
        if deductionReasoner(d, dr, dkb) {
            symptoms ← get all
            symptoms from dkb relting to d
            for each sj symptoms where
            (sj|d) > tD {
                if sj E
                symtomsFound AND sj E
                SymptomsAreadyElicited
                Add sj to
                symptomsToBeElicited.
            }
            Remove d from
            DiagnosesToBeElicited
            Add D to
            DiagmosesAlreadyElicited
        }
    }
}

deductionReasoner(d, dr, dkb){
    load reasoner pellet
    reason over dkb
    if inferred instance in dkb
        return true
    return false
}
    
```

**Algorithm 5:** Modified induction module of ST algorithm

```

Induct() {
    for each d ∈ DiagnoseAlreadyElicited{
        criteria ← computeHillsCriteria(d)
        SymptomsFoundWeight ← ∅
        SymptomsWeight ← ∅
        Weight←
        Symptoms_standard_weight
        Add pair(s1, weight) to
        symptomsFoundWeight
        Add pair(s1,
        symptom_standard_weight) to
        symptomsWeight
        If weight <=
        symptoms_standrdd_weight
        Criticalityd ←weight/n
        If criteria
            Add d to DiagnosesIncluded
            and save Criticilatyd
        Else
            Add d and criticalityd to
            DiasnosesExcluded
    }
}

computeHillsCriterias(d){
    ir ← InductionRuleSet
    ikb ← InductionKB
    temporalCriteria ← prompt user if exposure precedes
    the diseases
    bioPlausibleCriteria ← does disease cause relate to lab
    findings
    coherenceCriteria ← does ikb relate to lab findings
    consistencyCriteria ← does ikb relate to lab findings
    if temporalCriteria AND bioPlausibleCriteria
    AND coherenceCriteria AND
    consistencyCriteria
    return true
return false
}
    
```

**Algorithm 6:** Combined ST Algorithm

```

Inputs:
A special graph of ontologies P (sj|sj) and P (d1|sj):
AbductionKB
A special graph of ontologies P (sj|sj) and P (d1|sj):
AbductionKB
A special graph of ontologies P (sj|sj) and P (d1|sj):
AbductionKB
A presenting symptoms set PresentingSymptoms
Profile of the patient PatientProfile
Threshold for symptoms ts
Threshold for diagnoses tc
A set of rules for reasoning at abduction:
AbductionRuleSet
A set of rules for reasoning at deduction:
DeductionRuleSet
    
```

A set of rules for reasoning at induction:  
 InductionRuleSet

Output:

Set of likely diagnoses DiagnosesIncluded  
 Set of diagnoses excluded, DiagnosesExcluded  
 Set of symptoms that were found in patient SymptomsFound  
 Set of symptoms that were not found in patient SymptomsNotFound

Begin

SymptomsFound =  $\Phi$   
 SymptomsToBeElicited =  $\Phi$   
 SymptomsAlreadyElicited =  $\Phi$   
 DiagnosesToBeElicited =  $\Phi$   
 DiagnosesAlreadyElicited =  $\Phi$

For each  $s \in$  PresentingSymptoms {  
     Add  $s$  to SymptomsAlreadyElicited;  
     Add  $s$  to SymptomsFound;  
     For each  $d_i$  where  $P(d_i|s) \geq t_D$  OR  $C(d_i) \geq t_c$  {  
         Add  $d_i$  to DiagnosesToBeElicited;  
     }  
 }  
 While (DiagnosesToBeElicited  $\neq \Phi$  AND SymptomsToBeElicited  $\neq \Phi$ ) {  
     Abduct();  
     Deduct();  
     Abstract();  
 }

Induct();  
 End

## Implementation of the Inference Model for Clinical Reasoning

In this section, we used the breast cancer knowledge base developed in (Oyelade *et al.*, 2017b; 2018) in the implementation of the algorithms described in section 6. In the meantime, we first present the criteria for deciding the criticality (diseases staging) of the diagnosis (breast cancer) base on Fig. 2 (induction section) and algorithm 5. This criterion is computed using equation 8 and described by Table 1 which lists the values of likelihood ratio (LR) for each basic parameter considered during diagnosis elimination stage at the induction logical inference:

$$LR(\text{parameter}) = \frac{\text{Probability of parameter manifesting in stage } N}{\text{Probability of parameter not manifesting in stage } N} \quad (8)$$

Note that stage  $N$  ranges from breast cancer stages 0, 1A, 1B, 2A, 2B, 3A, 3B, 3C and 4. The probability of an event is given by equation 9:

$$\text{Probability} = \frac{\text{No. of favorable outcome}(\text{parameter manifesting in stage } N)}{\text{Total No. of possible outcomes}[\text{parameter risk factors}(\text{category})]} \quad (9)$$

**Table 1:** Parameters likelihood ratio

Parameters	Stage 0	Stage 1A	Stage 1B	Stage 2A	Stage 2B	Stage 3A	Stage 3B	Stage 3C	Stage 4
<b>Risk Factors</b>									
Older	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16
Genetic Mutations	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16
Dense Breasts	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16
Hormone Therapy	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16
Family History	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16
Drinking Alcohol	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16
	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16
<b>Symptoms and Signs</b>									
Hardening	0.43	0.43	0.43	0.67	0.67	1.5	1.5	4.0	4.0
Invisible Lump	0.43	0.43	0.43	0.67	0.67	1.5	1.5	4.0	4.0
Skin Erosion	0.43	0.43	1.0	1.0	1.0	1.0	4.0	4.0	4.0
Fluid Discharge	0.25	0.25	0.67	0.67	0.67	4.0	4.0	4.0	4.0
Redness	0.11	0.11	0.67	0.67	0.67	0.67	0.67	1.0	1.0
Growing Vein	0.11	0.11	0.43	0.43	0.67	0.67	1.5	1.5	1.5
Nipple Retraction	0.43	0.43	0.43	1.0	1.0	1.0	2.3	2.3	2.3
New Asymmetry	0.11	0.11	0.11	0.25	1.0	1.0	2.3	2.3	2.3
Orange Peel Skin	0.11	0.43	0.43	0.43	1.5	1.5	1.5	1.5	1.5
Breast Pain	0.11	0.25	0.25	1.0	1.0	1.0	4.0	4.0	4.0
<b>Laboratory Test</b>									
Biopsy	0.19	1.0	1.0	1.0	2.0	5.0	5.0	5.0	5.0
Mammogram	0.19	1.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
Ultrasound	0.19	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
Size of Tumor	0.19	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
Lymph Node	0.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
Metastasized	0.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0

```

private void algorithm(){
    ipaa=new IPAAAlgorithm();
    ipaa.start(); //CALL SUB-PROGRAM->startIPA()
    mma=new MonitoringModuleAlgorithm(ipaa);
    mma.start(); //CALL SUB-PROGRAM->startMonitoringModule()
    symptomsFound=AppContainer.userInputModelsymptomsFound;//CALL SUB-PROGR
    AbstractionAlgorithm abstractalgot=new AbstractionAlgorithm(this);
    DeductiveAlgorithm deductalgot=new DeductiveAlgorithm(this);
    AbductiveAlgorithm abductalgot=new AbductiveAlgorithm(this);
    QueueAsLinkedList tmpSymptomsAlreadyElicited, tmpDiagnosesToBeElicited,
    while(!symptomsFound.isEmpty()){ //WHILE we are still treating initial :
        tmpDiagnosesToBeElicited=abductalgot.algorithm(); //CALL SUB-PROGRAM->
        showstore.updateAbductionOutput(tmpDiagnosesToBeElicited);
        tmpDpdiagnosesAlreadyElicited=deductalgot.algorithm();//CALL SUB-PROGR
        showstore.updateDeductionOutput(tmpDpdiagnosesAlreadyElicited);
        tmpSymptomsAlreadyElicited=abstractalgot.algorithm();//CALL SUB-PROGR
        showstore.updateAbstractionOutput(tmpSymptomsAlreadyElicited);
        symptomsFound.dequeue();
    } //END-WHILE
    boolean ismmaipaSession=true; //get info from monitoring agent and ipa
    while(ismmaipaSession){
        if(!mma.isAlive() && !ipaa.isAlive()){
            ismmaipaSession=false;
        }
    }
    InductiveAlgorithm inductivealgot=new InductiveAlgorithm(this);//CALL SU
    inductivealgot.algorithm();
    showstore.setDiagnosesResult(alignedResult);//show output
}
    
```

Fig. 5: Code snippet from implementation of Algorithm 6

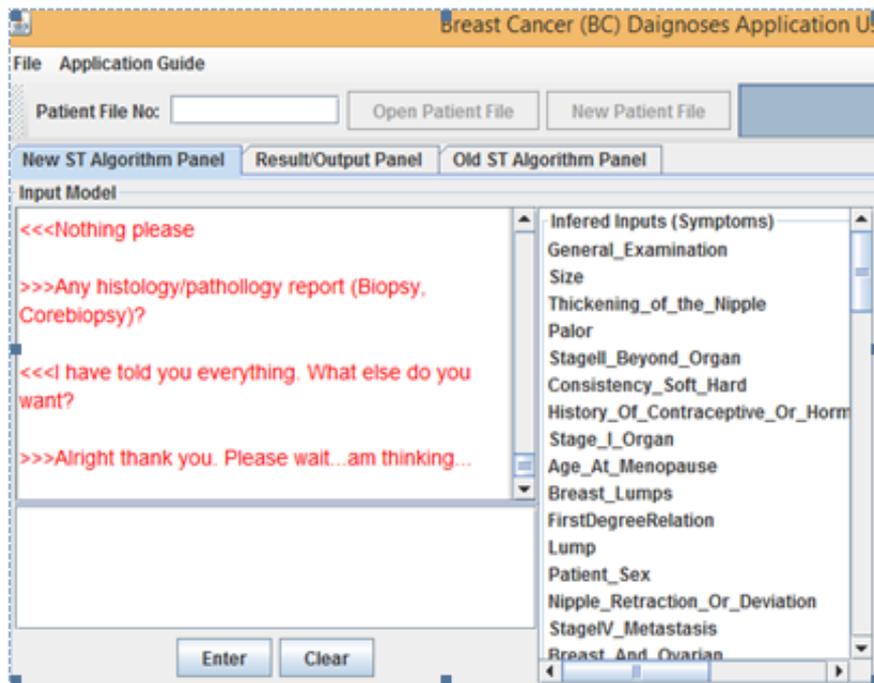


Fig. 6: Enhanced ST application diagnosing breast cancer

For each of the parameter in Table 1, the patient is expected to enter the likelihood value or *Symptom Weight*: to describe approximation of each parameter as felt by the patient. Algorithm 6 then compares the summed *Symptom Weight* of patient with that of the summed *Symptom Weight* for each stage on Table 1 and computes the *Criticality* or staging of diagnosis using the summed *Symptom Weight* collected. Java programming language was used for the implementation of the Algorithm 6 as shown by a code snippet in Fig. 5 and the application interface in Fig. 6.

## Results and Discussion

This research considered it appropriate to employ the medical diagnostic metrics state by Šimundić (2008). The author stated that diagnostic accuracy of any diagnosis test gives us an answer to the following question: How well this test discriminates between certain two conditions of interest (health and disease)? It is this discriminative ability that this paper measures by measuring diagnostic accuracy: True Positive (TP), True Negative (TN), False Negative (FN), False Positive (FP), Receiver Operating Characteristics (ROC), True Positive Rate (TPR) and False Positive Rate (FPR). Equations 10 through 16 are expressions for computing some metrics chosen for the presentation of our results in this section:

$$Sensitivity = \frac{TP}{(TP + FN)} \quad (10)$$

$$Specificity = \frac{TN}{(TN + FP)} \quad (11)$$

$$Accuracy = \frac{(TN + TP)}{(TN + TP + FN + FP)} \quad (12)$$

$$TPR = \frac{TP}{(TP + FN)} \quad (13)$$

$$FPR = \frac{FP}{(FP + TN)} \quad (14)$$

$$LR+ = \frac{sensitivity}{(1 - specificity)} \quad (15)$$

$$LR- = \frac{(1 - sensitivity)}{specificity} \quad (16)$$

The area under the ROC curve (AUC): By both FPF and TPR yields a coordinate (x, y), where TPR = TP/(TP+FN) and FPR = FP/(FP+TN). Hence ROC point is given by (FPF, TPR). Consider two tests/diagnosis X and Y with false positive rates (FPR) 0.95 and 0.88 respectively. If the region that is clinically relevant in the AUC is at low FPRs, then diagnosis Y is preferable to diagnosis X even though the ROC area of X is greater than Y.

Sensitivity, specificity and accuracy are described in term of number of true positive assessment (TP), Number of true negative assessment (TN), Number of all false negative assessment (FN) and Number of all False Positive assessment (FP). A good diagnostic test has LR+ > 10 and LR-< 0.1. By the standard of accessing the values of AUC with respect to diagnostic accuracy, values ranging between 0.9-1.0 are judged to excellent, 0.8-0.9 is very good, 0.7-0.8 is good, 0.6-0.7 is sufficient, 0.5-0.6 is bad and anything less than 0.5 is considered a diagnosis that is not useful. Data collected from the Ahmadu Bello University Teaching Hospital (ABUTH) as shown in Table 5, was used to draw a comparison between the improved ST algorithm (ONCODIAG) with that of (Fernando and Henskens, 2013; Fernando *et al.*, 2013) to obtain the result in Table 2. Similarly, in Table 3, the performance of our proposed algorithm is compared with the Wisconsin Breast Cancer Database (WBCD) and Wisconsin Diagnostic Breast Cancer (WDBC) datasets.

**Table 2:** Breakdown of results of ST and improved ST models

Parameters	ST model implementation	Enhanced ST model 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:	187ms	2101ms
Criticality:	Not applicable	Stage 1B
Treatment Plan:	Not applicable	Chemotherapy and Radiotherapy
Symptom Weight:	Not applicable	0.54347
Sensitivity	0.42	1.0
Specificity	1.0	0.5125
AUC	(0, 0.42)	0.49, 1.0

**Table 3:** The comparison of the results of WBCD and the improved ST algorithm

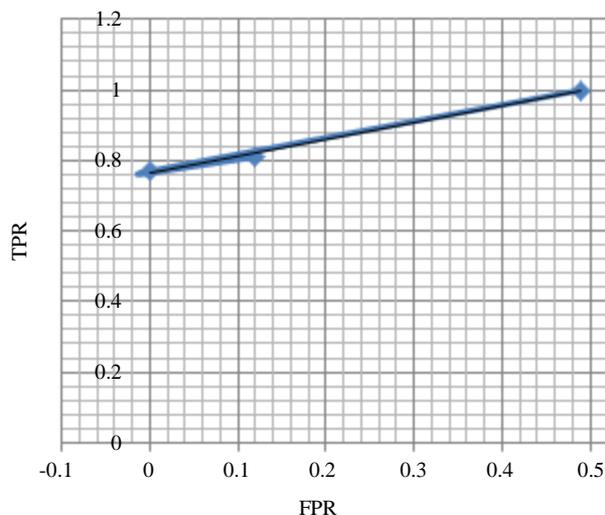
No. Instances	WBCD		ONCODIAG-ST	
	Benign	Malignant	Not-Breast Cancer	Breast Cancer
100	56	44	67	33
200	116	84	139	61
300	163	137	191	109
400	229	171	266	134
500	303	197	343	157
600	380	220	424	176
699	458	241	513	186

**Table 4:** The comparison of sensitivity, specificity and accuracy of WBCS and the improved ST algorithm

Metrics	ONCODIAG-ST	WBCS
Sensitivity	0.81	1.0
Specificity	0.89	1.0
Accuracy	86%	100%

**Table 5:** Sample patient data from Ahmadu Bello University Teaching Hospital (ABUTH) Shika-Zaria Nigeria.

Age	Parity	Breast examination and patient supplied information	Investigation	Staging	Treatment
42	6+0	Left breast was removed, Post surgery, nipple inverted, measured 2.5 by 1 cm solid masses. Lump in breast, Painful, Metastasis from beginning	Axillary lymph node present	Stage IV	Surgery, chemotherapy, radiotherapy
42	No data	Painful RB Lump, Axillary and supraclavicular, edema lymph node, needle biopsy. right breast, Metastasis from beginning	CT, Spinal Metastasis	Stage IV	Chemotherapy and radiotherapy for spine
44	4+0	Symmetry, upward nipple deviation, scar on upper part of axillary, Menarche at 15, age at 1stchild 25, last child 11yrs b4 presentation		Stage IV	Chemotherapy and radiotherapy for spine
35	6+0	Surgery, scar, right axillary lymph node measured 2 by 2 cm. Last menses Aug 26	Right Breast	Stage III	Chemotherapy and Radiotherapy
45	0+3	Surgery, noodle reoccurring. Size of tumor 6.8 by 6.5 cm,	Left Breast lump, no pain, no discharge	Stage III	Chemotherapy and Radiotherapy
38	0+1	Diagnosed 7 month, surgical scar Right Breast, Axillary node, lymph node	Right Breast, No Pain	Stage III	Surgery, Chemotherapy, Radiation
36	4+0	Mastectomy scar, huge lump, Age at first pregnancy 26	Right Breast	Stage III	Surgery, Chemotherapy, Radiation
35	1+0	Lump, mobile lump, 5 by 4cm, no lymph nodes, Right Breast involved	Lump 5 by 4 cm, Painful, no Nipple Discharge	Stage III	Surgery, Chemotherapy, Radiation
41	7+1	Surgery, locally advanced,	First child at 27, last at 38,	Stage IV	Surgery, Chemotherapy, Radiation
49	5+2	Surgery.	Age at first 28	Stage III	Surgery, Chemotherapy, Radiation



**Fig. 7:** ROC of the improved algorithm when tested on ABUTH, WBCD and WDBC datasets

The Area Under Curve (AUC) in Figure 7 shows that the point falls close to the top-leftmost section of the curve, which is an indication that the performance of the algorithm is good. Also, the sensitivity and specificity values of 1 and 0.54 are indications of how the improved ST model demonstrates the ability to detect presence of disease (breast cancer) and to rule out the presence of disease (breast cancer). Base on Table 4, the WBCD datasets has 699 instances and our algorithm achieved an accuracy of 86%, sensitivity and specificity of 0.81 and 0.89 respectively.

## Evaluation of the Proposed Algorithm and Discussion

In this section, we present an evaluation and its discussion in comparison of similar literatures of the result shown in the last section. The evaluation is focused on the accuracy measure of the related systems. We have chosen to compare this system (proposed in this paper) with machine learning, machine reasoning and knowledge representation techniques. Our comparison in this section is done base on the result of the comparison in Table 3 using Wisconsin repository.

In Table 6, eight (8) related works were compared with the proposed algorithm (ONCObc-ST). Mike *et al.* (2017; 2018; Alsane *et al.*, 2018) works which were focused on formalization of knowledge base for medical

diagnoses and were not based on medical diagnoses, accuracy measure was not necessary. Those of (Aloraini, 2012; Agarap, 2018) which are Machine Learning (ML) based techniques have accuracy of 95.6%, 90.0% and 97.36% respectively. Furthermore, the works of (Alharbi and Tchier, 2016) is a fuzzy-genetic based algorithms attained accuracy of 97.33%. However, fuzzy-genetic algorithms are limited by difficulty of genetic algorithm to guarantee optimality and solution weakens with increased size of the problem. Similarly ML algorithms may suffer the problem of acquisition of relevant knowledge or data which usually have impact on the performance of ML. This is contrary to MR which adapts to more flexible adaptation even in big data investigations. We conclude that though our proposed algorithm (ONCObc-ST) attained an accuracy of 88.72%, its technique was rigorously verified and is proven to be acceptable. Furthermore, semantic reasoning (which is the approach of this paper) permits the representation of knowledge in a very deep and meaningfully structured form which yields high inference power. In addition, semantic reasoning approach excels in provability through formal logic proofs to explain the result gotten by the system. Lastly, in domain like medicine, deep knowledge representation with complex rules is required and which semantic reasoning appropriately solves.

**Table 6:** Overall performance measure of breast cancer diagnostic algorithms

Authors/Year	Algorithms used	Category of Machine Intelligence	Purpose of research	Accuracy
Aloraini (2012)	Bayesian Network, Naïve Bayes, Decision trees J4.8, ADTree and Multi-layer Neural Network	Machine learning (ML)	Breast cancer detection	95.6%
Sene (2018)	Evidence theory and data mining	Data mining and Semantic clarification through ontology	In-flight formalism of medical knowledge representation for CDSS	N/A
Tapi Nzali (2018)	Linguistic and statistical approaches	Data mining	Formalism of breast cancer knowledge representation	N/A
Tapi Nzali (2017)	Classic text mining technique, latent Dirichlet allocation (LDA)	Data mining	Formalism of breast cancer knowledge representation	N/A
Agarap (2018)	GRU-SVM, R, MLP, NN, SR and SVM	Machine learning (ML)	Breast cancer detection	>90.0%
Seising (2006)	Fuzzy logic	Machine reasoning	Linkage to foundation of application of fuzzy logic to medical diagnosis	Not stated
Rakus-Andersson and Jain (2009)	Approximate reasoning algorithm and Fuzzy sets	Computational intelligence	Evaluation of a risk in the situation considering patient for surgery	Not stated
Alharbi and Tchier (2016)	Fuzzy-Genetic Algorithm Method	Machine reasoning (MR)	Breast cancer detection	97.33%
WBCS	Medical experts	Human intelligence	Breast cancer diagnoses	100%
Fernando <i>et al.</i> (2013)	Select and Test (ST) algorithm	Approximate Reasoning (Abduction, deduction and induction inference)		59.45%
This paper (ONCObc-ST)	Select and Test (ST) algorithm	Ontology learning and Semantic Machine Reasoning (MR)	Breast cancer diagnoses and formalism of breast cancer knowledge representation	88.72%

## Conclusion

In conclusion, this paper presents an improved Select and Test (ST) model for performing clinical reasoning. In addition, a three-fold (abduction, deduction and induction) logical inferences based algorithm of the improved model was also outlined and discussed. The formalism for carrying out approximate clinical reasoning was *modeled* and discussed using mathematical concepts such as relations and functions. The ST model we designed, which was augmented by a block diagram of the formalism for approximate reasoning, was meant to serve as an illustration of our proposal of using logical inferences in clinical solving problems. Furthermore, this paper gave a hint to the possibility of the interaction these logical inference methods to knowledge base that is modeled as ontology file. Moreover, we have argued that the peculiarity of clinical data and its reasoning process is characterized by some form of uncertainty or missing data, which necessitated the need for a formalized approximation method of reasoning with data. We addressed this challenge through the use of our formalism to prove that our logical inference methods are relevant in handling such peculiarity of clinical problem solving. Meanwhile, we have added a monitoring module into the improved model earlier discussed. Finally, the implementation of the model and its associated algorithms was done and case-study to breast cancer diagnoses. Results showed that the performance of the improved ST model compared to the existing ST model was better. We however note that the model presented in this paper is limited by lack of an explanation facility – a means for justification of how and why clinical reasoning solutions were executed to increase confidence for acceptability of result of diagnosis.

In future, we plan to employ the use of machine learning to carry out feature extraction from mammograms. The current approach presented in this paper depends on Radiologists to study the image and pass necessary input. However, we desire to completely automated the readership of all images: MRI, mammograms tomosyntheses and ultrasounds.

## Acknowledgment

Authors acknowledge the intellectual support provided by other researchers in the department.

## Author's Contributions

All authors equally contributed in this work.

## Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of

the other authors have read and approved the manuscript and no ethical issues involved.

## References

- Abu-Hanna, A. and P.J. Lucas, 2001. Prognostic models in medicine. AI and statistical approaches. *Meth. Inform. Med.*, 40: 1-5. PMID: 11310153
- Agarap, A.F.M., 2018. On breast cancer detection: An application of machine learning algorithms on the wisconsin diagnostic dataset. *Proceedings of the 2nd International Conference on Machine Learning and Soft Computing, (LSC' 18)*, ACM, USA, pp: 5-9. DOI: 10.1145/3184066.318408
- Alharbi, A. and F. Tchier, 2016. A fuzzy-genetic algorithm method for the breast cancer diagnosis problem. *Proceedings of the 9th International Conference on Advanced Engineering Computing and Applications in Sciences*, Jul. 19-24, IARIA, Nice, France, pp: 122-127.
- Aloraini, A., 2012. Different machine learning algorithms for breast cancer diagnosis. *Int. J. Artificial Intell. Applic.*, 3: 21-30. DOI: 10.5121/ijaia.2012.3603
- Alsane, S., K. Bernard and R. Pierre, 2018. Discovering frequent patterns for in-flight incidents. *Cognitive Syst. Res.*, 49: 97-113.
- Audétat, B.C., 2011. Retrieved from [http://www.royalcollege.ca/portal/pls/portal/!PWEB\\_PORTAL.wwpob\\_page.show?\\_docname=547651.PDF](http://www.royalcollege.ca/portal/pls/portal/!PWEB_PORTAL.wwpob_page.show?_docname=547651.PDF)
- Babette, R., V. Zazzu, R. Sudbrak, A. Kühn and H. Lehrach, 2013. Future of medicine: Models in predictive diagnostics and personalized medicine. *Adv. Biochem. Eng. Biotechnol.*, 133: 15-33. DOI: 10.1007/10\_2012\_176
- Bayesian networks. Retrieved from <https://pdfs.semanticscholar.org/a6ff/8be189d4d32fdbd8cc170dbbf15612d4de05.pdf>
- CDR, 2018. Clinical and Diagnostic Reasoning. <https://medicine.uiowa.edu/internalmedicine/education/master-clinician-program/information-students/clinical-and-diagnostic-reasoning>
- Coiera, E., 2003. *Guide to Health Informatics*. 2nd Edn., Arnold Press, London, ISBN-10: 144411400X, pp: 320.
- Croskerry, P., 2009. *Academia medical*. ACM, 84: 1022-1028. DOI: 10.1097/ACM.0b013e3181ace703
- Daniel, T.N., L. Williams, D.E. Paull, M.L. Graber and R.R.A. Hemphill, 2016. Theory-integrated model of medical diagnosis. *J. Cognitive Eng. Dec. Mak.*, 10: 14-35. DOI: 10.1177/1555343415618965
- Fernando, D.A.I.P. and F.A. Henskens, 2016a. The Select and Test (ST) algorithm and Drill-Locate-Drill (DLD) algorithm for medical diagnostic reasoning. *AIMSA LNAI*, 9883: 356-359. DOI: 10.1007/978-3-319-44748-3\_36

- Fernando, D.A.I.P. and F.A. Henskens, 2016b. Select and Test (ST) Algorithm for Medical Diagnostic Reasoning. In: Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, Lee, R. (Ed.), Studies in Computational Intelligence, Springer, Cham, pp: 75-84.
- Fernando, I. and F. Henskens, 2013. ST algorithm for diagnostic reasoning in psychiatry. *Polibits*, 48: 23-29.
- Fernando, I., F. Henskens and M. Cohen, 2013. An Approximate Reasoning Model for Medical Diagnosis. *Stud. Comput. Intell.*, 492: 11-24.  
DOI: 10.1007/978-3-319-00738-0-2
- Hosseinzadeh, Y.S. and F. Hosseini, 2017. Models of clinical reasoning with a focus on general practice: A critical review. *J. Adv. Med. Educ. Profess.*, 5: 177-184.
- Kawamoto, K., C.A. Houlihan, E.A. Balas and D.F. Lobach, 2005. Improving clinical practice using clinical decision support systems: A systematic review of trials to identify features critical to success. *BMJ*, 330: 765-765.  
DOI: 10.1136/bmj.38398.500764.8F
- Kishan, B., V. Chadha and C. Maini, 2012. A review of development and applications of expert system. *Int. J. Adv. Res. Comput. Sci. Software Eng.*, 2: 319-325.
- Mike, D.T.N., B. Sandra, L. Christian, M. Caroline and O. Thomas, 2017. What patients can tell us: Topic analysis for socialmedia on breast cancer. *J. Med. Internet Res.*, 5: e23-e23.  
DOI: 10.2196/medinform.7779
- Mike, D.T.N., B. Sandra, L. Christian, O. Thomas and A. Jérôme *et al.*, 2018. Reconciliation of patient/doctor vocabulary in a structured resource. *Health Inform. J.* DOI: 10.1177/1460458217751014
- Monteiro, S.M. and G. Norman, 2013. Diagnostic reasoning: Where we've been, where we're going. *Teach. Learn. Med.*, 25: 26-32.  
DOI: 10.1080/10401334.2013.842911
- Oyelade, O.N., A.A. Obiniyi and S.B. Junaidu, 2017b. An ontology-based knowledge framework for diagnosing breast cancer. Proceedings of the 13th International Conference Proceedings of Nigerian Computer Society, (NCS' 17).
- Oyelade, O.N., A.A. Obiniyi, S.B. Junaidu and S.A. Adewuyi, 2017a. Patient symptoms elicitation process for breast cancer medical expert systems: A semantic web and natural language parsing approach. *Future Comput. Inform. J.*, 3: 72-81.  
DOI: 10.1016/j.fcij.2017.11.003
- Oyelade, O.N., A.A. Obiniyi, S.B. Junaidu and S.A. Adewuyi, 2018. ST-ONCODIAG: A semantic rule-based approach to diagnosing breast cancer base on Wisconsin datasets. *Inform. Med. Unlocked*, 10: 117-125. DOI: 10.1016/j.imu.2017.12.008
- Rakus-Andersson, E., 2009. Computational intelligence in medical decisions making. *Recent Adv. Dec. Mak.*, 222: 145-159.  
DOI: 10.1007/978-3-642-02187-9\_9. 222: 145-159
- Ramoni, M. and M. Stefanelli, 1992. An epistemological framework for medical knowledge-based systems. *IEEE Trans. Syst. Man Cybernet.*, 22: 1361-1375.
- Seising, R., 2006. From vagueness in medical thought to the foundations of fuzzy reasoning in medical diagnosis. *Artificial Intell. Med.*, 38: 237-256.  
DOI: 10.1016/j.artmed.2006.06.004
- Shortliffe, E.H., 1987. Computer programs to support clinical decision making. *JAMA*, 258: 1-6.
- Šimundić, A., 2008. Measures of diagnostic accuracy: Basic definitions. Department of Molecular Diagnostics University Department of Chemistry, Sestre milosrdnice University Hospital, Zagreb, Croatia.
- Stausberg, J. and M. Person, 1999. A process model of diagnostic reasoning in medicine. *Int. J. Med. Inform.*, 54: 9-23.  
DOI: 10.1016/S1386-5056(98)00166-X