A knowledge-based approach to assign breast cancer treatments in oncology units

Teddy G. Miranda-Mena [a,1], Sandra L. Benítez U. [a,1], José Luis Ochoa [a,1],
Rodrigo Martínez-Béjar [a,1], Jesualdo Tomás Fernández-Breis [b,*], Juan Salinas [c,2]

[a] *Departamento de Ingeniería de la Información y las Comunicaciones, Universidad de Murcia, 30071-Espinardo (Murcia), España*
[b] *Departamento de Informática y Sistemas, Universidad de Murcia, 30071-Espinardo (Murcia), España*
[c] *Servicio de Oncología y Radioterapia. Hospital Virgen de la Arrixaca, 30120-El Palmar (Murcia), España*

Abstract

The aim of this work is to present a developed knowledge-based approach for breast cancer treatment. This Decision Support System (DSS) uses an incremental knowledge acquisition technique called Multiple Classification Ripple Down Rules (MCRDR) and a breast cancer treatments knowledge base. This system integrate MCRDR inference engine and use knowledge bases to reach advanced reasoning level during the decision making process. The knowledge base will be built from clinical protocols, and its content will be updated by the doctor using the MCRDR component. This DSS infers a cancer treatment from the (clinical or pathological) input data supplied by the physician and it also allows the maintenance of the knowledge base.
© 2005 Elsevier Ltd. All rights reserved.

Keywords: Knowledge acquisition; Knowledge representation; Decision support system; Breast cancer; Protocols and guidelines; Validation and verification

1. Introduction

The detection and treatment in early stages of breast cancer in women and men are fundamental for the survival of the patient. Given the difficulty of defining early breast cancer, some doctors decided to focus on the diagnosis and treatment of small invasive breast cancer. A small invasive breast cancer is defined as an invasive lesion with a diameter of 10 mm or less-in other words (Silverstein, 1997). The past 20-30 years have seen dramatic changes in the treatment of breast cancer. This has provoked the generation of vast amounts of knowledge about breast cancer treatments, mainly in tacit nature, that is, kept in the mind of oncologists. So, there is a need for mechanisms and systems allowing for making all this knowledge explicit. The explicit knowledge of breast cancer treatments is usually stored and organized as clinical protocols (Amrit, 2002). This is the case of the guide known as 'National

Comprehensive Cancer Network' (NCCN) (http://www.nccn.org/). Many oncologists currently use NCCN as a clinical protocol to learn about cancer treatments, because they act as digital medical guides (explicit knowledge) through which the expert browses through a set of algorithms showing the steps that must be followed in a specific treatment. However, these guides do not permit the experts the extracting knowledge quickly, so becoming not efficient when time is critical. On the other hand, when oncologists assign cancer treatments, they use to complement the knowledge of clinical protocols with tacit knowledge (Amrit, 2002). This tacit knowledge has been accumulated by the oncologist with the experience of historical cases and alternative knowledge that may reinforce the decision made. In this sense, there is a need for mechanisms to manage the (tacit and explicit) knowledge in oncology units when cancer treatments are assigned. An adequate solution to manage oncological knowledge might be the creation of mechanisms, in cooperation with experts, for managing knowledge bases (Loshin, 2001) for different types of cancer treatment (i.e., breast, lung, skin, and so on). The benefits of knowledge-based systems in medical units are very significant, such as: decision making time reduction, historical cases-based support to the decision making, update and maintenance of the specialized knowledge, the reuse and storing of the corporative knowledge of the unit, and tutoring capabilities for non-experts (Hayes & Jacobstein, 1994).

---

* Corresponding author. Tel.: +34 968364613; fax: +34 968364151.
E-mail addresses: tegomime@yahoo.com (T.G. Miranda-Mena), sandrabenitez6@hotmail.com (S.L. Benítez U.), spotsy8a@hotmail.com (J.L. Ochoa), rodrigo@um.es (R. Martínez-Béjar), tegomime@yahoo.com (J.T. Fernández-Breis).
[1] Tel.: +34 968364634; fax: +34 968364151. http://klt.dif.um.es
[2] Tel.: +34 968369502.

In this paper, a Decision Support System (DSS) that uses an incremental knowledge acquisition technique called Multiple Classification Ripple Down Rules (MCRDR) (Kang, 1996), is presented. MCRDR allows for solving multiple classification problems. The knowledge acquisition is based on the justification provided by the oncologist when (s)he makes a diagnosis, but not in the explanation of the steps to draw that diagnosis. MCRDR offers benefits for the knowledge base maintenance, because the correction of a wrong conclusion is carried out by refining the rule that inferred the wrong conclusion. So, the oncologist does not have to modify the MCRDR structure to remove the wrong conclusion and to add the correct one. In this paper, MCRDR is used as a black box whose input is clinical information and whose output is the suggestion of a set of breast cancer treatments. The internal structure of MCRDR falls beyond the scope of this paper. When wrong treatments are inferred or no treatment is suggested, it is necessary to active the knowledge acquisition procedure. The MCRDR knowledge acquisition procedure provides a natural technique through which the oncologist must supply the system with the correct treatment and the rules explaining the decision; the MCRDR corrects the knowledge base, allocating the rules and the treatment in their correct positions. This makes the acquisition of a new treatment trivial. The oncologist decides which treatments assigned by the system are wrong and must be removed. Provided that the oncology treatments are large and explicative texts, the system must allow experts to use this type of explanations.

Finally, this paper is structured as follows. Section 2 presents the concepts modelled in the knowledge-based system. The construction of the breast cancer treatments knowledge based is the target of Section 3. The Decision Support Systems is presented in Section 4. The Section 5 contains the validation of the tool and finally, related work, further work and the conclusions are put forward in Section 6.

2. Modelling the concepts of the knowledge-based system

In oncology, the concept 'patient' has the following attributes: healthcare record number, name(s), surname(s), age, gender, date of birth, and marital status. An episode can be defined as a session in which the oncologist performs these actions: (1) analysis of the patient's records, diagnosis, evaluation of the stage of the cancer tumor and other clinical and/or pathological data; (2) assignment of a primary treatment. The attributes of the concept 'episode' are: healthcare record number, episode number, type of episode, episode date, observations, primary treatment. At this point, other concepts such as 'clinical condition', 'clinical case', and 'treatment' must be defined. A clinical condition may be defined as the pair (description, value), for instance (tumor size, 5 cm), and they are used by the oncologist to provide a partial description of the stage of the disease, for instance{(age,35), (gender, male), (marital status, married), (tumor size, 5 cm),…}. A clinical case may be defined as a set of clinical conditions that provides

a complete description of the disease stage. They are defined by the oncologist during an episode. Clinical cases are the input data for the MCRDR inference engine. Breast cancer treatments are defined by the oncologist as a set of rules. Each rule is defined as a set of clinical conditions, and the rule is fired if and only if all the conditions are met. For instance, the oncologist can assign a primary treatment 'A' to a patient in case the following clinical conditions stand: age $>=50$, gender$=$male, size tumor$>=5$ cm. If the clinical conditions of the patient meet this rule, then the primary treatment A is assigned to the patient. It must be noticed that several treatments can be assigned to the same clinical case.

3. Construction of the breast cancer treatments knowledge base

The breast cancer treatments knowledge base is the source of the knowledge-based system. In this research work, the knowledge is built from widely recognized oncology protocols. In particular, the oncology NCCN clinical protocols are used. There, the cancer treatments are represented by means of algorithms. Through this work, a knowledge base for non-invasive breast cancer treatments (DCIS) was developed. Next, each step of the knowledge base construction is detailed:

(Step 1) Knowledge delimitation: Selection of the DICS algorithm (see Fig. 2) amongst the ones presented in the NCCN clinical protocols. Then, the knowledge to consider is constrained and delimited by following this procedure:

Select the breast cancer stage: This stage is determined by physicians gathering information from examinations and tests on the tumor, lymph nodes, and distant organs. The TNM staging system, also known as the American Joint Committee on Cancer (AJCC) system, is the most widespread to describe the growth and spread of breast cancer. Information about the tumor, nearby lymph nodes, and distant organ metastases is combined and a stage is assigned to specific TNM groupings, from I to IV. The clinical stage is determined by what the doctor learns from the physical examination and tests. The pathologic stage includes the findings of the pathologist after surgery. Most of the time, the pathologic stage is the most important one because it is not usually known whether the cancer has spread to lymph nodes until the pathologist examines them under the microscope. TNM includes different categories with associated possible conditions. The T category is based on the size and the spread to nearby tissue. The N category is based on which of the lymph nodes near the breast, if any, are affected by the cancer, for instance, the cancer has not spread to lymph nodes or the cancer has spread to lymph nodes under the arm on the same side as the breast cancer. Lymph nodes have not yet attached to one another or to the surrounding tissue. Finally, the M category depends on the cancer spread to any distant tissues and organs, such as, no distant cancer spread or the cancer has spread to distant organs. Once the T, N, and M categories have been assigned, this information is combined to assign an overall stage of 0, I, II, III, or IV. For instance, the stage 0 corresponding to Ductal
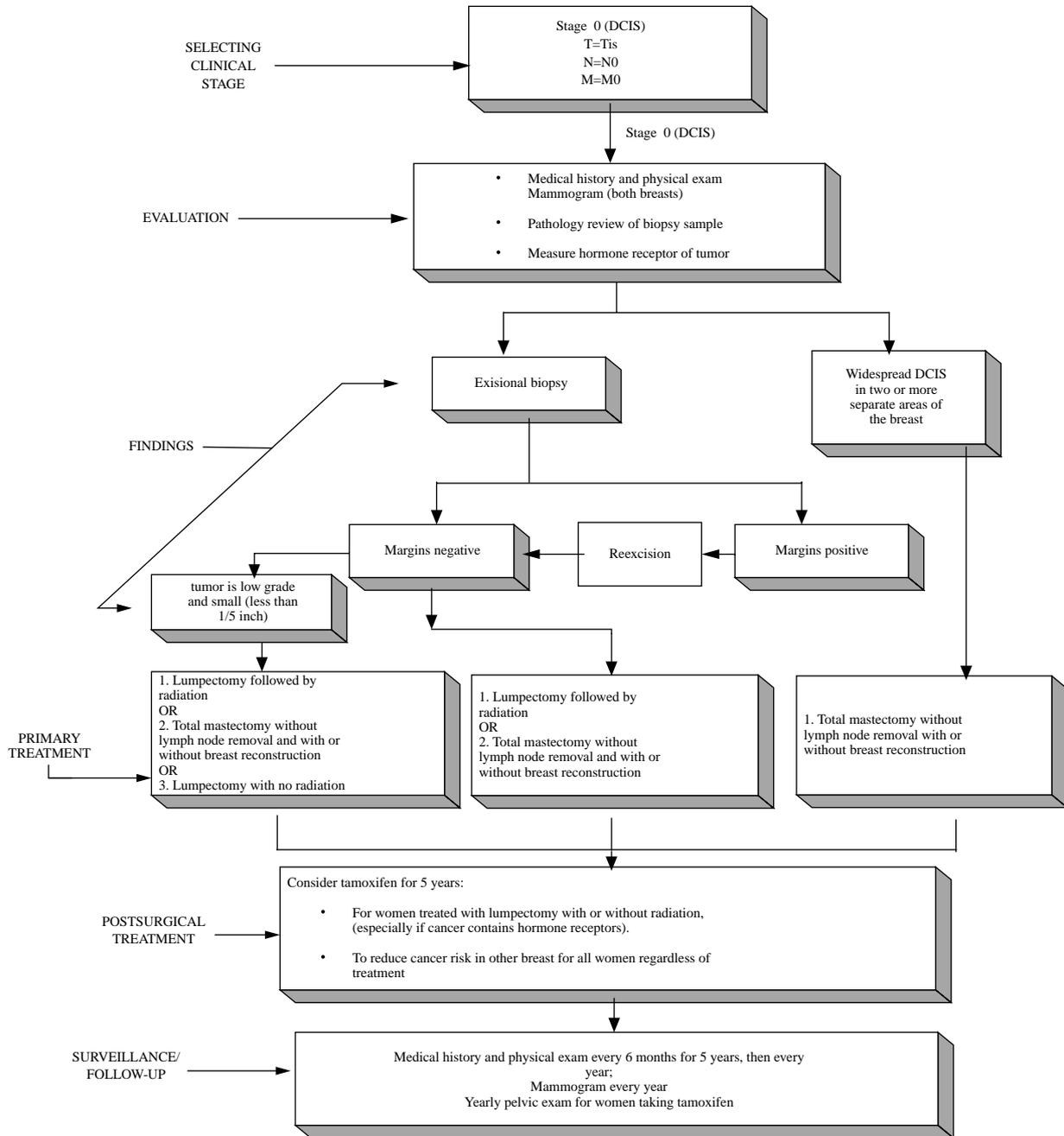
Fig. 1. Algorithm of treatments for DCIS in Non-invasive breast cancer.

Carcinoma In Situ (DCIS) is chosen, so, the values for clinical stage are T=Tis, N='the cancer has not spread to lymph nodes' and M='no distant cancer spread'.

Identifying the knowledge fractions to the breast cancer stage: For DCIS cancer, the next six knowledge fractions are identified (see Fig. 1): a) clinical stage, b) evaluation, c) findings, d) primary treatments, e) post surgical treatment and f) surveillance/follow-up.

Knowledge representation. According to the given definition for treatment rules (in Section 2), the knowledge of clinical conditions and treatments for DCIS cancer is represented as follows:

(1) Clinical Conditions = < clinical stage + evaluation + findings + age + gender >
(2) Treatment = < Primary Treatment + Post surgical Treatment + Surveillance/Follow-up >

(Step 2) Knowledge codification: Each clinical condition and treatment for DCIS is identified from the NCCN algorithm and assigned a code; their possible values are determined. These are used to build the explanation tables (code, values), which are used by the explanation module of the knowledge-based system. Now, the steps to codify knowledge for DCIS cancer are described.

Table 1
Examples of knowledge codification for DCIS

Knowledge Fraction	Codes	Values	Observations and comments
Evaluation	E1	Medical history and physical exam	Patient's previous examination to evaluate if other abnormal areas in the breast exist.
	E2	Mammogram (both breasts)	
Primary Treatment	PT1	Observation	The treatments against the breast cancer he includes surgery and radiation therapy of the breast, as well as medicaments for the cancerous cells that they had contaminated to another organs.
	PT2	Total mastectomy without extirpation of the lymphatic ganglions with or without reconstruction of the breast	
	PT3	Tumorectomy with or without radiation.	

For each identified knowledge fragment, a code is assigned. Table 1 contains examples of these codes for evaluation and primary treatment. For instance, the code 'E1' corresponds to Evaluation knowledge fraction and its value is 'Medical history and physical exam'.

Grouping codes for construction of clinical cases: Codes of 'clinical conditions' and 'treatments' are grouped for an easy knowledge base's construction. Table 2 presents a summary of these codes.

(Step 3) Construction of Clinical Cases: The knowledge is consistently grouped and contextualized in concepts to create the clinical cases with their respective clinical conditions and treatments. Here, some examples of the possible clinical conditions and treatments are identified for each clinical cases appearing in the algorithm, shown in Fig. 2. An example for the triple $<$clinical case number, clinical conditions, treatments$>$ might be <1,{ Stage$=0$, T$=$Tis, N$=$N0, M$=$M0, Age$=$Age0, Eval$=$Eval1,H$=$H1, S$=$S1}, {PT2, PostT1, CST}$>$

(Step 4) Use of a knowledge acquisition tool: The knowledge base is created using the knowledge acquisition module of the knowledge based system and the cases, which have previously been identified and coded. The result of this process is a decision tree like the one shown in Fig. 2.

4. The Oncology Decision Support Tool

A software tool called 'Oncology Decision Support Tool' was developed by following the principles described in previous sections. This system is comprised of the three modules shown in Fig. 3. The first module implements the MCRDR engine to extract the knowledge stored in cases in clinical protocols. The second module allows for inferring clinical treatments. The third module is used to format the explanation of the inferred treatments. The three modules were developed by using Java and MySQL. The programming tool and the database management system were connected through the JBDC interface.

4.1. Patients.

The tool offers the oncologist the option to assign treatments to the patient. For this purpose, the oncologist has to select the type of cancer, then the types of cancer related to (invasive and non-invasive) breast cancer, and finally the oncologist select the stage LCIS or DCIS. Then, the tool displays a list of patients to treat, and if the patient no exist the oncologist can add new patients.

4.2. Inference engine

The implementation of the inference engine is based on MCRDR. The algorithm of the inference module is: 1) Capturing data from the clinical records of the patient; 2) Capturing the clinical conditions of the clinical case of the patient; 3) Formatting the clinical case and 4) inferring the treatment by using MCRDR. Fig. 4 shows an example of use of the tool for infer a treatment of a breast cancer clinical case, having three main areas: 1) the clinical conditions according to NCCN; 2) the values of the conditions; and 3) the clinical conditions of this clinical case, which will be the input clinical case for the knowledge-based system. The button 'Treatment Inference' allows the oncologist for inferring the treatment that will be assigned to the input clinical case. The result of the inference is processed by the explanation module described next.

4.3. Knowledge acquisition and maintenance

The knowledge acquisition module is based on MCRDR and it allows oncologists to maintain the case-based knowledge base. Fig. 5 shows a working session with the tool. There, a knowledge base of breast cancer cases and treatments is created. The background window contains the input case {T$=$Tis, N$=$N0, M$=$MO, Age$=$Age0, H$=$H1, Est$=$Est0, Eval$=$Eval1}. The inference results are modified in the intermediate window ('PosT1' and 'PT2'). The foreground window shows the cases that are involved in the result ('CST'). The tool also allows for loading and saving files from and in different knowledge bases, so the oncologist can work with different knowledge bases.

Table 2
Summary of codes for clinical conditions and treatments (some examples)

Code	Description
Stage	0, I, II, III and IV
T	Primary Tumor
N	Regional Lymph Nodes

Rules Created

```
IF  THEN -------
   IF M==M0 & H==H1 & Est==Est0 & T==Tis & N==N0 & Eval==Eval1 & G==GF & Edad==Edad0 THEN PT2
   IF M==M0 & Est==Est0 & H==H1 & T==Tis & N==N0 & Eval==Eval1 & Edad==Edad0 & G==GF THEN PostT1
   IF M==M0 & Est==Est0 & H==H1 & T==Tis & N==N0 & Eval==Eval1 & Edad==Edad0 & G==GF THEN CST
   IF M==M0 & H==H2 & Est==Est0 & N==N0 & Eval==Eval1 & G==GF & Edad==Edad0 THEN PT2
   IF M==M0 & Est==Est0 & H==H2 & N==N0 & Eval==Eval1 & Edad==Edad0 & G==GF THEN PostT1
   IF M==M0 & Est==Est0 & H==H2 & N==N0 & Eval==Eval1 & Edad==Edad0 & G==GF THEN CST
```
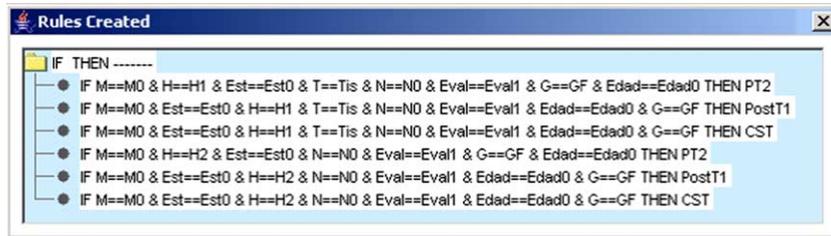
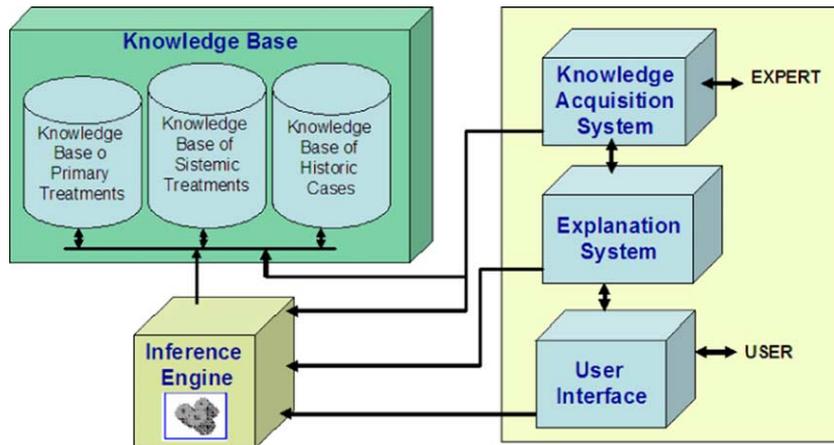Fig. 2. Decision tree of treatments.

Fig. 3. Architecture of the System.

4.4. Explaining conclusions

The explanation module processes the conclusions produced by the knowledge-based system inference engine. Each conclusion is an abstraction of the inferred treatment. This module builds the explanations for the inferred treatments in natural language, which is understandable by the expert and uses oncology terms. So, for each conclusion, the following operations are performed: (1) retrieval of the explanation from the base of treatments; (2) generation of the natural language explanation; (3) display to the oncologist.

Clinical Diagnosis Conditions

Clinical Diagnosis Conditions
- E: Evaluacion
- Edad: Edad
- Est: Estadio
- Eval: Evaluaciones
- G: Genero
- H: Hallazgos
- M: Metastasis
- N: Ganglios
- T: Tumor ①

Values of Clinical Diagnosis Conditions
- E1: Historia médica y examen físico
- ② E10: Tomografía computarizada, ecografía abd…
- E2: Mamogramas diagnósticos (ambos senos)
- E3: Segunda opinión de la evaluación patológic…
- E4: Recuentos sanguíneos y pruebas químicas
- E5: Radiografía de tórax
- E6: Ecografía e imágenes por resonancia mag…
- E7: Evaluación patológica de la muestra de la b…
- E8: Prueba del tejido del receptor de estrógeno…
- E9: Gammagrafía ósea (sólo si los síntomas o l…

Gammagrafía ósea (sólo si los síntomas o las pruebas sugieren que el cáncer se ha propagado los huesos)

= ▼   Add Condition

Case for Evaluation
- G: Genero = GF: Femenino
- H: Hallazgos = H1: Propagación del DCIS en dos o más áreas separadas del seno ③
- M: Metastasis = M0: No hay evidencia de metástasis a distancia
- N: Ganglios = N0: No hay evidencia de metástasis ganglionares regionales
- T: Tumor = Tis: Carcinoma in situ: Tis (DCIS), Tis (LCIS), Tis (Piaget)

Eraser Condition   New Inference   Treatment Inference   Close
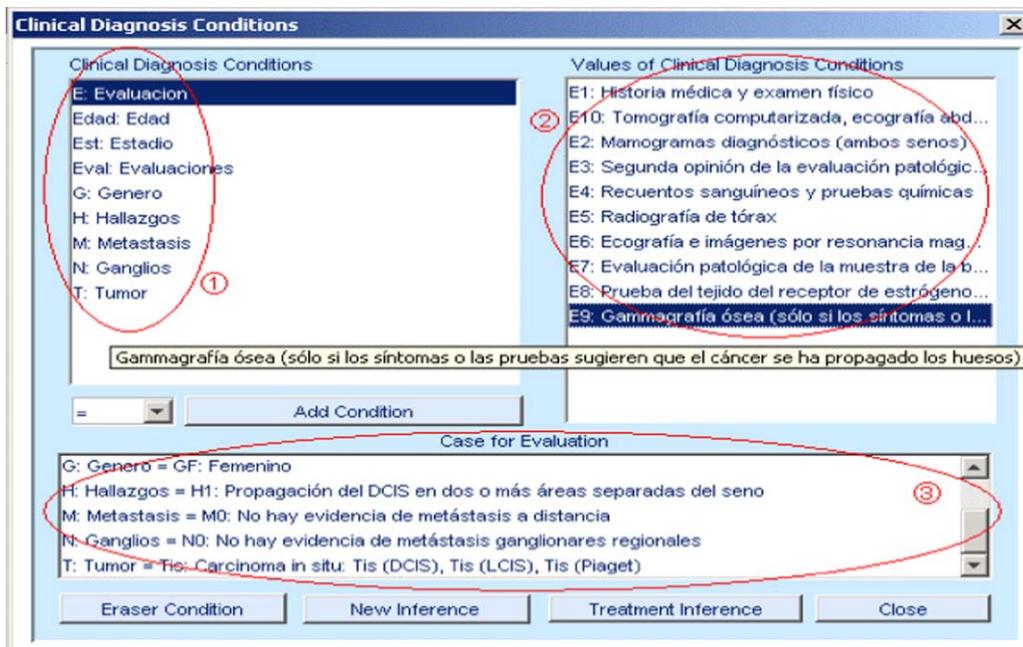
Fig. 4. A particular case for the Inference: 1) the clinical conditions according to NCCN; 2) the values of the conditions; and 3) the clinical conditions of this clinical case, which will be the input clinical case for the knowledge-based system.
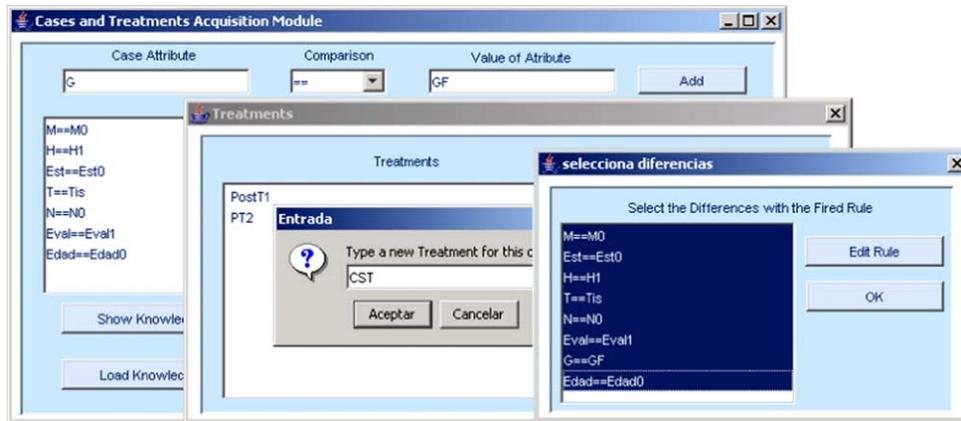
Fig. 5. Knowledge acquisition session.

5. Validation

The goal of the tool is the generation of a correct treatment for the patient. In particular, given a set of patient's clinical conditions (a clinical case), the system has to assign the correct treatment. Determining the appropriate validation criteria is an important consideration (Tsai, 1999). In our tool, the proposed treatment by the system implies a critical decision, because this treatment must be assigned to a patient. Consequently, it is only possible to validate the performance of the knowledge-based system against the opinion of oncologists. So, reliability (Guida and Mauri, 1993) has been the validation criterion in this case. Reliability tests checks whether the system conforms to its specifications and if all expected requirements by oncologists are satisfied by the system. In our case, the system was installed in the Radiotherapy service at the Hospital 'Virgen de la Arrixaca' in Murcia (Spain), and it has been used by oncologists. For the validation, clinical test cases were generated extracting information from healthcare records from n patients. Each clinical test case was structured according to the description given in Section 2. For each patient, m clinical conditions and their k corresponding treatments were generated. Hence, for each clinical test case, he following operations were performed: (1) the oncologist inputs the clinical test case to the system; (2) a set of treatments for the clinical test case are proposed; (3) the oncologist revises the set of treatments; (4) if the oncologist finds wrong treatments then, he/she corrects the error; (5) the decisions of the oncologist are documented. After some running time, some feedback was obtained from the oncologists. Apart from the comments about the usability and friendliness of the user interface, they stated that the systems performs correctly according to the reference protocol, but they do not generally follow the protocol, so they would be happier with a system that combines the protocol and their proper knowledge to suggest treatments. And this is exactly what allows MCRDR to do. The system is still under validation before being incorporated to daily practice but the degree of satisfaction of the oncologist increases as the system is capable of incorporating their knowledge.

6. Conclusions

In this paper, a knowledge-based system to assign breast cancer treatments is described. The system makes use of the incremental knowledge acquisition capabilities of MCRDR to transform the oncological clinical protocols in a rules-based knowledge base. This knowledge base is used to infer the treatment, so using the MCRDR inference capabilities. The tool developed has been validated in an oncological domain for breast cancer treatments. For this purpose, the NCCN clinical protocol was used. The algorithm corresponding to the non-invasive breast cancer was specifically used. The knowledge-based system intends to facilitate oncologists the maintenance of the knowledge base, thus guaranteeing that its content is updated and efficient decisions can be made. Some decision support systems (DSS) based on artificial intelligent techniques such as neural networks (Zhou et al., 2002), probabilistic events networks (Galán, Aguado, Diéz & Mira, 2002) or hybrid systems (Papadopoulos, Fotiadisb & Likas, 2002) have been put into oncology clinical practice. The main advantage of our knowledge-based system proposed in this paper with respect with those is the use of technologies that facilitate an easy maintenance of the knowledge base. As further work, the following issues should be pointed out: automatic generation of cases and treatments from clinical protocols stored in pdf; knowledge acquisition and inference from cases with alternative treatments and allowing oncologists to select one according to effectiveness degrees; and multiple treatments inference from a concrete historical cases base.

References

Amrit, T. (2002). *The knowledge management toolkit: Orchestrating IT, strategy, and knowledge platforms* (2nd ed.). Upper Saddle River, NJ: Prentice Hall PTR (pp. 35–63).

Galán, S. F., Aguado, F., Diéz, F. J., & Mira, J. (2002). NasoNet, modeling the spread of nasopharyngeal cancer with networks of probabilistic events in discrete time. *Artificial Intelligence in Medicine, 25*, 247–264.

Guida, G., & Mauri, G. (1993). Evaluating performance and quality of knowledge-based systems: Foundation and methodology. *IEEE Transactions on Knowledge and Data Engineering* , 204–224.

Hayes Roth, F., & Jacobstein, N. (1994). The state of knowledge-based systems. *Communications of the ACM, 3*, 27–39.

Kang, B. H. (1996). Validating knowledge acquisition: Multiple classification ripple down rules. School of Computer Science and Engineering University of New South Wales, PhD thesis.

Loshin, D. (2001). *Enterprise knowledge management: The data quality approach*. Los Altos, CA: Morgan Kaufmann pp. 253–277.

Papadopoulos, P., Fotiadisb, D. I., & Likas, A. (2002). An automatic microcalcification detection system based on a hybrid neural network classifier. *Artificial Intelligence in Medicine, 25*, 149–167.

Silverstein, M. (1997). Recent advance: Diagnosis and treatment of early breast cancer. *BMJ, 314*, 1736.

Tsai, W. T., Vishnuvajjala, R., & Zhang, D. (1999). Verification and Validation of knowledge-based systems. *IEEE Transactions on Knowledge and Data Engineering* , 202–212.

Zhou, Z. H., Jiang, Y., Yang, Y. B., & Chen, S. F. (2002). Lung cancer cell identification based on artificial neural network ensembles. *Artificial Intelligence in Medicine, 24*, 25–36.