Training Aided-System in Senology: Methodologies and techniques
Souad Demigha, Colette Rolland

To cite this version:

HAL Id: hal-00706498
https://hal.archives-ouvertes.fr/hal-00706498
Submitted on 16 Jun 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Training-Aided System in Senology: Methodologies and Techniques

Souad Demigha, Colette Rolland

Université de Paris I Panthéon-Sorbonne – Centre de Recherche en Informatique – Paris, France

Souad.Demigha@malix.univ-paris1.fr
Colette.Rolland@univ-paris1.fr

ABSTRACT

The paper provides a description of methodologies and techniques required for a Training System Development in the field of Senology (TSDS), based on the exploitation of senologic images (primarily mammograms but also echographic images or MRI) and their related clinical files. The aim of such a system is to help breast cancer screening in education. This system will help assist junior radiologists in routine clinical use.

Development of such a TSDS requires understanding of users’ needs (expertise and pedagogy), model design and system implementation.

Specifications have been derived from the experience of the senologists from the Department of Radiology of the Necker Hospital (Paris, France), Department where the training system will be implemented.

To be compliant with commercial systems for digital and CAD mammograms, terminological systems used by the TSDS to describe and index data must be based on DICOM and BI-RADS dictionaries.

A detailed discussion of the choice of such a method and technique is provided and their respective contribution is described.

Keywords: Training System Development, Senology, Breast Cancer, Screening, Education, Mammogram, BI-RADS, DICOM.

1. INTRODUCTION

Early detection of breast cancer is considered as a major public health issue. Breast cancer incidence is the highest among female cancers and the second cause of mortality in Europe. About one in ten woman will develop a breast cancer in her lifetime [1]. To address this problem, it is necessary to create the adequate conditions allowing for the installation of mass detection campaigns, i.e. involving the maximum of women at risk.

Detection is carried out starting from the analysis of breast images, primarily mammograms but also echographic images or MRI, coupled with the exploitation of information derived from the patient’s history, from punctures, etc. Therefore, the clinician grounds his/her diagnosis on the result of image analysis procedures and on the synthesis of various types of information. It requires a significant amount of knowledge and know-how, which can be acquired only through a long practice.

It is thus critical, in order to meet the requirements of mass detection, to have tools that contribute to the training of senologists to acquire and update this knowledge, together with the evolution of imaging systems in senology.

To meet these requirements, we propose to specify and to build a Training System Development in Senology (TSDS) based on the exploitation of senologic images and their related clinical files. TSDS must be simple, interactive, pedagogical and easily implemented. It will promote the evolution of teaching in senology by offering the “junior radiologist” trainees an advanced pedagogical product. Finally, TSDS will permit a strengthening of knowledge together with a very elaborate presentation of results. At last, the know-how will derive from all these factors.

The development of TSDS requires the understanding of users’ needs (expertise and pedagogy), model design and system implementation. The requirements have been analyzed by using the Crews-l’Ecritoire (Cooperative
Requirements With Scenarios) approach [2]. It is based on the “Requirement Engineering” concept. It helps understand the users’ needs via a semi-automatic analysis of textual scenarios, i.e. scenarios written in a natural language. These requirements serve as an input of the design modeling of the TSDS, the next phase of TSDS design. Domain knowledge modeling has been identified as a key-issue during Requirements Engineering (RE) to help specify complete, consistent and accurate requirements [3]. After capturing experts' requirements, we have determined a model including "domain knowledge" experts and beginners.

TSDS includes the “case base” (knowledge on the taught field), “pedagogical module” (pedagogical knowledge), “trainee’s model” (the trainee’s knowledge) and the “interface” (communication with the trainee). In this paper, we have focused on the “case base” design of the TSDS. Other modules like pedagogical module, trainee’s model and interface will be dealt with in an other paper to be issued. Also, implementation and validation of the model will lead to an other work.

The paper is organized as follows:
- Section 2 presents an overview of a medical education and the medical context of senology.
- Section 3 details the material and methods used for the TSDS development, in particular the knowledge model.
- Section 4 provides the adequate language for the TSDS and finally:
- Section 5 is the conclusion with further research works in progress.

2. THE MEDICAL CONTEXT

After providing an overview of medical education, we describe different data and knowledge included in the TSDS.

2.1 MEDICAL EDUCATION

Medical practice requires an aptitude for decision-making, based on a continuously updated knowledge, which amount can only increase. Medical education is therefore aimed at training future doctors. It has to prepare the students to control management technologies and medical data processing, which take into account the making-off, the spreading and the validation of knowledge [4].

One of the major targets of medical education is thus to learn how to learn. This approach assumes an education effort in the following fields:

- To help the trainee (student, then physician) how to define his/her individual objectives and thus his/her needs for education.
- To individualize the relationship with the trainee and to adapt education to each one's level.
- To increase the share of individual work (updating knowledge, practical exercises, self-checking…).
- To think over the process of education, to achieve a critical study of decisional situations, validation of knowledge, etc.).

There are many efforts in tutoring system achievements in medicine. We quote the very famous GUIDON system. GUIDON [5] (expert system for teaching diagnosis and therapeutic rules of meningitis) is a tutoring system based on the expert system MYCIN [6] and its rules strategy.

GUIDON is based on the case method: the trainee is placed in a realistic context of problem solving, i.e. where he/she is confronted with a concrete case in which he/she must explain his/her assumptions. When a case is selected and described by GUIDON, the trainee makes a diagnosis, asks questions when additional information is required. The tutoring takes place in case of the trainee’s explicit request or when the trainee’s answers do not lead to the result. The main aspect of GUIDON is the separation between the knowledge and the teaching parts.

2.2 KNOWLEDGE CAPITALISATION

The development of a quality training system in senology requires that the needs (expertise and pedagogy) of various expert users (experts and beginners) are analyzed. The educational needs for experts and beginners senologists are both based on “expertise knowledge” and “pedagogical knowledge”.
"A knowledge is a description of the world. It determines the competence of a system in problem solving: the depth and breadth of problem solving power is determined by what the system knows" [7]. After capturing knowledge, it is necessary to encode it, i.e. to store, to release and to use it.

In reference [7], a representation of knowledge is defined by: "representation is the way knowledge is encoded. It defines the performance of a system in solving problem: speed and efficiency of problem solving are determined to a significant degree by the choice of representation".

To build tutoring systems (Baron [8]), it is necessary to integrate knowledge in the taught field. Be it "expert knowledge" or "reference knowledge", it is not obvious to be able to collect this knowledge from experts. Twenty years ago, a specific field was developed in AI (Artificial Intelligence) as "acquisition knowledge". Methodologies and tools have been developed to enhance the design of system-based knowledge together with other scientific fields such as cognitive psychology.

For the building of TSDS, we must take into account explicit elements in the context of training and choose them according to the targets of training. Thus two questions can be raised:
- How to organize these elements in a model?, and
- Which models are useful and acceptable in a TSDS?

2.2.1 EXPERTISE KNOWLEDGE
Expertise knowledge is based on the mixture of experience experts acquired from their routine practice and from textbook knowledge.

The senologic process includes four basic phases: **image reading**, **radiological interpretation**, **decision-making** and **follow-up**.

- **Image reading**: it consists in searching and extracting relevant information (imaging data and textual ones).
- **Radiological interpretation**: it is based both on "clinical data" (patient’s history, screening, current health status, and information on previous clinical examination) and "radiological data" (information such as those defined by BI-RADS [9]).
- **Decision-making**: it consists of normal observation, particular observation, and biopsy.
- **Follow-up**: it consists of patient’s short-term follow-up and long-term follow-up.

2.2.2 PEDAGOGICAL KNOWLEDGE
The aim of pedagogical knowledge in the TSDS consists of the 3 points as defined in [10]:

- To organize trainee’s activities around pedagogical targets and programs.
  - "Task" level, choice of examples, exercises, assistance, explanations,…
  - "Learning sequence" level, pedagogical targets.
- To adapt knowledge formulation, initiatives and dialogue between trainees.
  - Presentation of the interface, texts and images.
  - Explanation of statements, rules and procedures.
- To identify errors and solving them.
  - Error catalogues for “false rules” rules.
3. MATERIAL AND METHODS

This section presents the material and the methods required for the building of the TSDS.

3.1 THE TSDS

The design of our system includes four research axes (see Fig.1):

1. Educational contents (*Case Base*);
2. Trainee’s difficulties (*Trainee’s Model*);
3. Educational methods (*Pedagogical Model*);
4. User’s Interface.

![Diagram of TSDS](image)

**Figure 1:** General architecture of the TSDS

- **The Case Base:** includes reference domain from experts as follows:
  1. **Identification characteristics:** they serve to associate mammogram records with other clinical data from patient’s hospital record and among others, demographic data.
  2. **Clinical characteristics:** they include data about patient’s history: screening history, current health status, and previous clinical examination.
  3. **Radiological characteristics:** information such as that defined by BI-RADS [9].
  4. **Histological characteristics:** information from histological examinations procedure, (histopathology).
  5. **Digital image characteristics:** image production characteristics and other technical or administrative information, such as those defined by DICOM [11] and information about the analysis procedure (e.g., CAD system).

- **The Trainee’s Model:** this model takes into account the trainee’s capabilities by proposing several levels of exercises. It elaborates a feedback adapted for each type of error.

- **The Pedagogical Module:** thanks to the knowledge base of the field, the pedagogical module develops a reasoning which allows to evaluate the trainees, to guide them with a pedagogical strategy adapted to the trainee’s model. Misdiagnosis made by the trainee allows more relevant and more effective interventions of the system. It is aimed at helping the trainee use the knowledge necessary and to neglect the non relevant one.

- **The User’s Interface:** serves to communicate with the trainee.
A case description:

1. The system interrogates the “case base”.
2. The system retrieves data from the “case base”.
3. The system chooses a case in the “case base”.
4. The system presents the case to the trainee.
5. The system asks the trainee to describe his/her case.
6. The trainee answers the questions of the system.
7. The system compares the “trainee’s answer” with the “case base’s answer”.
8. If the “trainee’s answer” = “case base’s answer” then
   9. (Cases are similar)
      10. The system calculates similarity measures.
      11. The system determines evolution rules.

19. Else (Cases are not similar),
   20. The system determines errors.
   21. The system determines errors step by step.
   22. The system asks the trainee to find the error

Figure 2: Extract of a case description in senology

We have focused on the knowledge base “case base”, which is the basic component of our system. Other modules (pedagogical module, trainee's model and interface) will be dealt with later on.

3.2 DOMAIN KNOWLEDGE: THE TSĐS MODEL

The most commonly used modes in medical education consist in teaching trainee’s experiments, called clinical cases. These cases learned individually or in groups are examples resulting from real situations. The case-based reasoning (CBR) [12] is an approach which permits to consider expert knowledge as a set of cases. The expert relies on this set of cases, experience in decision-making and diagnosis.

3.2.1 THE CASE-BASED REASONING (CBR)

Case-based reasoning (CBR) is an Artificial Intelligence approach to learning and problem solving based on past experience. A past experience is stored under the form of solved problems (“cases”) in a so-called “case base”. A new problem is solved grounded on adapting solutions to similar problems (see Fig.3) to this new problem.

Case–based reasoning is a cyclic, four-phased process [13] (from [14]):

1. REtrieve: the aim of this phase is the selection of one (or several) case(s) which solve(s) a problem similar to that of the new case (also called the target).
2. REuse (adaptation): the target and the retrieval case (source) are combined to reach a solution. The solution of the source is adapted to account for the differences between the target and the source.
3. REvise: the purpose of this phase is to make sure that the proposed solution is correct and shall lead to success if applied.
4. RETain: the new case and its solution are stored into the case base. Thanks to this learning phase, the system requires new knowledge at each reasoning cycle.
3.2.2 MEDICAL TUTORING SYSTEMS WITH CBR

In medicine, CBR has mainly been applied for diagnosis and tutoring. One of the earliest medical expert systems that use CBR techniques is CASEY.

- CASEY [16] (from [17]) is a system that diagnoses heart failure. It uses as the input the patient’s symptoms and produces a causal network of possible internal states that could lead to those symptoms. When a new case arises, CASEY tries to find out cases of patients with similar but not necessarily identical symptoms. If the new case matches, then CASEY adapts the retrieved diagnosis by considering differences in symptoms between the old and the new cases.

- PROTOS [18] (from [17]), was developed in the domain of clinical audiology. It learned to classify hearing disorders from descriptions of patients’ symptoms, previous history, and test results. PROTOS was trained with 200 cases in 24 categories from a speech and hearing clinics. After training, PROTOS had an absolute accuracy of 100%.

There are also many works in medical imaging using CBR. We quote the very well-known:

- PROTOISIS [19] is a case-based system based on the PROTOS [18] learning and reasoning from experience. It has shown a great potential for use in decision support systems. A prototype was developed and tested to explore the applicability of this technique to the selection of diagnostic imaging procedures.

- MACRAD [20] also permits the retrieval of reference radiologic images (standards X-rays, scanners...). The 300 cases illustrated by 3000 images are stored in a relational database and indexed with their content.

3.2.3 THE TSDS WITH CBR

The previous steps show the level of accuracy when using the CBR approach in medical diagnosis and training. It contributes to the acquisition and dissipation of clinical expertise: (1) while the trainee gets familiarized with a rich empirical content, often unavailable in individual clinics, by relating (2) this content to the theoretical aspects of the specific cases, and (3) by revealing the “diagnostic feeling” involved in diagnosis and treatment. This supporting role and, especially, the important function that case-based reasoning systems should have in the structure, the methods and
the content of medical education, underscore the need for further research in theoretical aspects and actual development of such systems [21].

The CBR approach is rather appropriate in the medical field and thus in medical imaging. We adopt CBR as knowledge representation for the case base design of the TSDS. The purpose in senology is to store all the necessary features (texts and images with various modalities) and to use the previous experience for diagnosing current cases. The radiologist can benefit from prior experience and cases.

3.3 THE CASE BASE
The mammogram records used for the making-off of the case were provided by Doctor Corinne Balleyguier, from the Department of Radiology of the Necker Hospital. One hundred out of 500 records were selected and digitized with the agreement of Doctor Balleyguier. Let us recall that the films are digitized with their overall surface, with a resolution of 42\(\mu\)m/pixel and a dynamics of 12 bits/pixels.

3.3.1 KNOWLEDGE INCLUDED INTO THE CASE
A case is a contextualized piece of knowledge representing an experience. It contains the past lesson that is the content of the case and the context in which the lesson can be used [22].

A case includes empirical data describing experience acquired in the solving of an accurate situation. It has two parts: problem description -the case we try to diagnose- and solution description -the diagnosis- [23].

It is mainly standardized information included into the reporting by BI-RADS and DICOM, which has served for the case design (see Fig.4).

<table>
<thead>
<tr>
<th>Patient’s Name :</th>
<th>Patient’s Surname :</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient’s Birth date:</td>
<td></td>
</tr>
<tr>
<td>Referent doctor :</td>
<td></td>
</tr>
</tbody>
</table>

Bilateral screening mammography.
March 1rst, 1998.

Clinical history : history of family breast cancer screening.
There are bilateral, disseminated fibro-glandular opacities. The mammograms are compared with the preceding ones (Wichita Clinic), dated May 6, 1995.

SYNTHESIS
Incomplete examination. Mass circumscribed to center-left part. An echographic examination is recommended. The patient must get an appointment for it.

Further investigation is needed.

BI-RADS CATEGORIE 0

Doctor Jessica Taylor.
Radiologist.

Figure 4: Example of a case as BI-RADS [24]

Conversion of the reporting into cases has required the design of knowledge model.

3.3.2 REPRESENTATION OF KNOWLEDGE DESIGN
Let us assume a context of product and process as they are defined in NATURE project [25]. While a product “is the result to be achieved”, a process is “the way which allows to achieve this result”.

Doctor Jessica Taylor.
Knowledge design is stored in *product traces* and in *process traces*. A *product trace* represents the set of successive problems in processes used in senology. The *process trace* represents the set of successive steps leading to the solution as opposed to the solution itself, i.e. the design artifact (*product*) [26].

For instance, for the TSDS, we consider the four phases of the senologic process: *image reading*, *radiological interpretation*, *decision-making* and *follow-up* as a *product* and the "*how*" to obtain each phase as a *process* (see Fig.5).

The *product traces* and the *process traces* constitute simultaneously the cases, i.e. the reusable problem and the reusable solution. Given a new problem (trace), a similar trace is retrieved from memory, adapted to the new situation and replayed using a derivational analogy similarity approach to PRODIGY [27] (from [14]).

A trace is a case. A case corresponds to a goal that the deciding agent (designer) is trying to reach. A case is itself split up into an *interface* part (the problem, i.e. the "*what*") and a *specification* part (the solution, i.e. the "*how*") [26].

The main component of the *interface* is its *interface context*. The *interface context* represents the intention (the goal) the agent is trying to achieve as well as the *situation* (the initial state) in which he/she is trying to achieve this intention [26].

A *context* associates an intention of the agent with the *situation* in which this intention appears. It is represented as a `<situation, intention>` couple, for example `<TSDS system “TSDS system of the Department of Radiology of the Necker Hospital”, Develop (the TSDS system)>`. A *situation* is built upon one or several product parts, each product part being an instance of a product part type, i.e. an instance of a product part type, i.e. an instance of a concept of the product model. For example, in the above mentioned context, the situation (TSDS system “TSDS system of the Department of Radiology of the Necker Hospital”) is built on the product part “TSDS system of the Department of Radiology of the Necker Hospital”.

![Figure 5: Extract of model trace in senology](image)

According to our goal representation formalism similar to the Crews-l’Ecrtoire approach [28], an *intention* is composed of a *verb* and one or *several parameters*.

For instance, the goal of our project can be expressed by the following sentence:
The parameters are:

1. **The target**: which indicates the entity concerned by the goal. Crews makes a distinction between an “object”, which exists before the goal completion, and a “result” which proceeds from the goal achievement. For instance, in the goal “validate the biopsy request”, the biopsy request is the target object because it exists before the course of the scenario. On the contrary, “obtain the agreement of the second radiologist”, the agreement is considered as the target result of the goal.

2. **The direction**: which represents an oriented relationship (arc) between two Crews concepts.

3. **The way of goal realization**: which is a two fold concept including "Means" and “Manner” for achieving the goal. The “Mean” indicates the support used for achieving the goal whereas the “Manner” provides information on the way the goal is achieved. For example, for the goal “communicate patient examination results thanks to a vocal record”, vocal record is the support of the goal and is identified as a “Mean” to “facilitate breast cancer screening using mammogram”. The “Manner” is expressed by “using mammogram”. A “mean” can be formulated as a new goal. In the previous example, after a new formulation of “using mammogram”, we obtain a new goal which is “performing mammogram”.

4. **The beneficiary**: concerns the “agent” taking advantage of the goal achievement.

5. **The referent**: specifies the entity defining the goal. For instance, “perform a second interpretation for assessing the primary reading” plays the role of referent.

6. **The quality**: qualifies the goal in terms of qualitative features.

7. **The place**: locates the goal in space (e.g., in the Department of Radiology of the Necker Hospital).

8. **The time**: indicates temporal constraints on goal achievement or on the scenario.

### 3.3.3 INDEXING DATA

One may query data by index, feature, content or knowledge. One characteristic of medical images is that they contain information that often lacks accuracy. But for research purposes, we need an accurate and standardized description of features on mammograms to be able to query the case base. That is why we specified a complete description of relevant features we are interested in. This description is completed for each mammogram interpretation, and will serve as a basis for the first indexing system we will set. This textual indexing will be used, in a first step, and will help to tune image analysis procedures to offer, in a second step, indexing by image analysis content [29].

Textual indexing assumes the inclusion of all items useful to describe image features. This might be burdensome for radiologists when interpreting images. But up to now, this is the only way to get accurate enough information for research purpose. This indexing will allow research teams to develop image analysis procedures, to permit indexing by content, which is not enough advanced to be routinely used in digital mammography.

To be compliant with commercial systems for digital mammography and CAD mammography, terminological systems used by the TSDS to describe and index data must be based on DICOM and BI-RADS dictionaries.

### 3.3.4 CHOICE OF IMPLANTATION

We initially planned to implement the “case base”. The knowledge of radiologists is stored as “cases” and we must be able to extract it via the Web. We have chosen the Java language for language implementation.
4. DISCUSSION AND CONCLUSION

In this work, we have attempted to determine methods and tools required for the implementation of Training System Development in Senology (TSDS). We have achieved a real case study from the experience of radiologists from the Department of Radiology of the Necker Hospital. The paper has provided discussions and definitions of medical training systems.

The TSDS is aimed at helping acquisition of knowledge by trainees. This acquisition is enabled by adapting the tutoring/trainee's interaction, on the basis of each student's capabilities, which are inherent in each of them. Information may come from various sources, but, most of the time, it deals with knowledge and know-how attributed by the system to the trainee, according to his/her behavior. Knowledge may come from a knowledge base shared by all the trainees. This is why we have thought about the way to memorize this knowledge and be able to retrieve it according to the users’ needs (tutoring/trainee). We have mainly focused our work on the knowledge base of the training system. The approach used was the case-based reasoning for knowledge memorization.

We have provided the rationale for adopting this approach thanks to the support of existing literature.

This approach was as follows: all the radiologists’ knowledge is presented under the form of cases stocked as product and process traces in the knowledge base labeled as "case base". We define a product as the result and the process as the way in which we obtain this result. A trace is a case. Given a new problem (trace), a similar trace is retrieved from memory, adapted to the new situation and replayed using a derivational analogy approach.

We are currently in the process of conceiving trainee’s model, pedagogical module and interface. Future work shall concentrate on the further development and implementation of our TSDS particular of the case base.

ACKNOWLEDGMENTS

I wish to thank Doctor Corinne Balleyguier, MD, Radiologist, from the Department of the Necker Hospital, for her knowledge and information in medical senology and Mrs Françoise Marchand, from the INSERM (National Institute of Health and Medical Research) for her manuscript reading.

REFERENCES