CADMIUM II: Acquisition and Representation of Radiological Knowledge for Computerized Decision Support in Mammography

Eugenio Alberdi[a], Paul Taylor[a], Richard Lee[a], John Fox[b], Margarita Sordo[b], Andrew Todd-Pokropek[c]
[a]CHIME, University College London, UK
[b]ACL, Imperial Cancer Research Fund, London, UK
[c]Dept. of Medical Physics, University College London, UK

ABSTRACT

CADMIUM II is a system for the interpretation of mammograms. A novel aspect of the system is that it combines symbolic reasoning with image processing, in contrast with most other approaches, which use only image processing and rely on artificial neural networks (ANNs) to classify mammograms. A problem of ANNs is that the advice they give cannot be traced back to communicable diagnostic inferences. Our approach is to provide advice based on explicit knowledge about the diagnostic process. To this end, we have conducted a knowledge elicitation study which looked at the descriptors used by expert radiologists when making diagnostic decisions about mammograms. The analysis of the radiologists' reports yielded a set of salient diagnostic features. These were used to inform the advice provided by the symbolic decision making component of CADMIUM II.

INTRODUCTION

Decision support in mammography

Advances in digital mammography have led to the development of a variety of computer decision aids for mammogram interpretation[1]. The most common approach is the use of image processing algorithms for the detection of abnormalities. Further, many systems are capable of providing a diagnosis for the detected findings by using neural network classifiers[2]. Many such systems obtain remarkable results in terms of diagnostic accuracy. However, a problem of neural networks is that their "reasoning" is implicit. These systems may be able to classify reliably a mammogram as malignant or benign but cannot explain the reasons behind the diagnosis. In our view, for a decision aid to be usable, the basis for its decisions should be communicable to the user.

In previous publications we have described CADMIUM, a system for the diagnosis of mammograms which combines state of the art image processing with symbolic representations of clinical decisions[3,4]. The system uses symbolic reasoning to relate information obtained from image processing to the decisions radiologists take. To our knowledge, no other system has been developed which combines symbolic decision making and imaging. Some decision aids do use symbolic reasoning and can provide useful information about diagnoses[5,6]. However, such systems require the features and measurements of mammographic appearances to be introduced manually by a human expert. In CADMIUM, in contrast, the process is fully automated as such measurements are obtained from the results of image processing operations.

CADMIUM was evaluated with radiographers who were being trained to interpret mammograms and it was found that they performed significantly better when using the system[3].

Up to now, we have concentrated on a single problem: screening for breast cancer. More specifically, our focus has been the differential diagnosis of calcifications. These appear as small white flecks in a mammogram and distinguishing "benign" from "malignant" forms is one of the important skills that a film reader must learn.

The work presented in this paper concerns the development of CADMIUM II, an advanced version of the prototype which refines and extends a number of different components of the system[7]. One improvement in CADMIUM II is the incorporation of a wider selection of image processing operators than were provided by the earlier system. Different approaches to the detection and characterization of calcifications have been implemented[7]. Another advancement of CADMIUM II is the use of PRO*forma* to represent decisions made in radiology. PRO*forma* is a formal specification language with constructs for representing complex tasks like decisions, plans and actions; it is a natural and expressive formalism for specifying clinical processes[8]. We have used PRO*forma* to specify a protocol which represents the sequence of tasks to be performed in breast screening: the decisions, the candidates at each decision and the arguments that would support each candidate.

In CADMIUM II, the key decision is the assessment of calcifications in a mammogram. For this decision, the candidates are terms which are indicative of the risk of malignancy and the arguments are statements about confirmed properties of the observed calcifications. A crucial problem is the acquisition of relevant radiological knowledge about calcifications to be used in the specification of such arguments.

Acquiring radiological knowledge

In the development of the original CADMIUM prototype, an attempt was made to build up a knowledge base of arguments for diagnosis options based on properties of calcifications. This was done by transcribing statements about the differential diagnosis of calcifications that were found in review articles, textbooks and monographs on the topic.

The resulting knowledge base was considered by radiologists to be confusing and in some respects erroneous. One problem was that some of the statements in the knowledge base had been removed from the context in which they were written; for example, some statements appeared in articles as diagnoses which were possible but not typical of calcifications. Respondents disagreed strongly with such statements. Furthermore, clear inter-rater disagreements were found about terminology and medical knowledge.

There is extensive evidence to suggest that radiologists often disagree in their interpretation of mammograms[9]. This has led some authors to conclude that the standardization of descriptive terms for mammogram interpretation may be an unattainable goal[10]. However, various studies of radiologists' judgments have positively contributed to the definition of useful sets of mammographic features for inclusion in decision aids[11,12]. These studies typically start with a set of pre-specified ad-hoc mammographic terms which experts are asked to rate. The terms with highest inter-rater agreement are then extracted and incorporated in the decision aid.

In the study described in this paper, we have followed an alternative approach. Rather than testing a predefined set of descriptive terms, our goal was to obtain a new set directly from radiologists. The purpose was to identify a core set of salient features that are actually used by radiologists in making the distinction between benign and malignant calcifications.

METHOD

Radiologists

Eleven consultant radiologists from 6 different hospitals and screening centres in Southeast England volunteered to take part in the study. The radiologists had an average of 13 years of experience in reading mammograms with an average of 9,050 mammograms read per year.

Material

Mammograms from 20 symptomatic patients evaluated at the Whittington Hospital (North London) in 1997-98 were used as stimuli. The set consisted of mammograms from: 4 patients with no reported calcifications or abnormalities, 8 patients with reported malignant appearances, and 8 patients with reported benign appearances. Four films were used from each patient, namely, the standard mediolateral-oblique and craniocaudal views of the left and right breasts.

Data acquisition

The participants were asked to read the mammograms as they would in a normal clinical situation and to "think aloud", reporting everything that went through their mind. They were instructed to describe in detail what they saw in the images, especially the calcifications, if present. They were also asked to provide a tentative diagnosis and suggest a course of action for each patient. Their verbal reports were recorded on audio-tape.

Data Analysis

The resulting audio-recorded think-aloud reports were transcribed. A record was then kept of all the different findings reported by each radiologist. The resulting records were compared with each other to identify those findings on which more than one of the radiologists agreed.

The main focus of the analyses was a detailed study of the terms used by the radiologists to describe the reported calcifications. Firstly, a record was kept of all the terms used by the radiologists when describing calcifications. The resulting descriptors were then grouped according to the different dimensions or properties of the calcifications to which the descriptors referred (e.g., shape, size, density, distribution, etc.). With the aid of two consultant radiologists, synonymies were established. As a result, a reduced list was generated with the properties and property values which covered all the different calcification appearances described by the participants but excluded redundant terms.

RESULTS

A total 159 different terms were used by the participants to describe the calcifications in the study. These terms were found to refer to twelve different properties: 1) shape of the flecks; 2) contour of the flecks; 3) size of the flecks; 4) density of the flecks; 5) distribution of the flecks; 6) number of flecks in a group; 7) variability amongst the flecks in a group; 8) cluster or area size; 9) location of the flecks; 10) orientation; 11) associated findings; and 12) differences in calcification appearances between the mediolateral and craniocaudal views.

Seven of these properties were found to provide discriminating evidence. These are shown in Table 1. The properties and values in the table are those which were used by radiologists to describe calcifications that they considered benign but never to describe calcifications that they considered malignant, and *vice versa*. From the table are excluded all those descriptors which were noted by only one of the participants and a few which, in consultation with a consultant radiologist, were considered to be of dubious discriminating value. The table shows the maximum number of radiologists who noted, at least once, each property value when describing a calcification in each diagnostic category. Two of the descriptors in the benign category were also used when describing indeterminate calcifications thus their discriminatory value is less obvious. On the other hand, the value pleomorphic was noted for both malignant and indeterminate calcifications. This suggests that, at least in some of the participants' opinion, this appearance is not an unambiguous sign of malignancy but, nevertheless, indicates moderate to high level of suspicion.

Part of the analysis focused on the 45 calcifications which were reported by more than half of the radiologists (about 50% of the 91 calcifications reported in the study). These are shown in Table 2. The groupings indicate whether the calcifications were diagnosed by most of the radiologists who reported them as benign, malignant or indeterminate; the group "discrepant diagnoses" includes the calcifications for which there was no diagnosis on which the majority of the participants agreed. Further, the benign calcifications could be classified in 10 distinct groups, as a function of the descriptors most commonly associated with them and, similarly, the malignant ones were classified into two distinct groups (see Table 2). The first five groups of calcifications in the benign group were diagnosed as benign by all the radiologists who reported them. For the rest of the benign calcifications and for the malignant and indeterminate ones there was no unanimity but agreement was always over 75%.

Table 1. Discriminating Descriptors

Properties and Values	Benign N=11	Indeterminate N=11	Malignant N=11
Size			
large	11	--	--
Density			
lucent centre	11	--	--
Distribution			
vascular	11	--	--
isolated	2	--	--
Variability			
pleomorphic	--	6	10
homogeneous	4	1	--
Shape			
branching	--	--	9
curvilinear	4	--	--
Contour			
well defined	6	2	--
rim	3	--	--
ill-defined	--	--	1
Orientation			
towards nipple	--	--	2

The descriptors presented in Table 2 are those which were used for all the calcifications in each group by two or more of the radiologists.

Most of the calcifications in the table (especially those unanimously diagnosed as benign) were described with at least one property value which belonged to the set of discriminating descriptors shown in Table 1. The exceptions were: the "smudgy" and "punctate" calcifications, the benign and malignant calcifications associated with an opacity, and the calcifications in the indeterminate and "discrepant diagnoses" groups.

KNOWLEDGE REPRESENTATION

The findings reported in the previous section have been used to specify the arguments for the candidates (diagnosis options) considered in the decision "Interpret calcification" implemented in the PRO*forma* specification in CADMIUM II. In the current implementation, two candidates are being considered: benign and malignant.

Arguments in PRO*forma* can be represented in different ways depending on the type of support they provide. For example, an argument can be represented as "confirming" or "excluding" a

Table 2. Most commonly reported types of calcification

BENIGN calcifications
vascular (Σ= 6) vascular/vessel related (100%)
ring-like (Σ= 8) round (56%), lucent centre (41%), rim (33%)
curvilinear (Σ= 1) curvilinear (63%), large (38%), dense (38%), clustered (38%)
round & large (Σ= 2) round (73%), large (56%)
coarse (Σ= 2) dense (52%), large (35%)
smudgy (Σ= 2) low density (100%), small (45%), number (23%), scattered (23%)
linear (Σ= 2) linear/ductal (90%), dense (40%), well defined (30%), large (30%)
popcorn/almost popcorn (Σ= 3) large (41%), dense (38%), clustered (22%)
punctate & scattered (Σ= 2) small (80%), scattered (66%), round (45%), low density (32%)
associated with opacity (Σ= 4) assoc. with opacity (22%), 1-5 flecks (40%)
MALIGNANT calcifications
DCIS (Σ= 4) small (54%), branching (51%), pleomorphic (46%), linear (34%), low density (21%), variable size (21%), irregular (21%)
associated with opacity (Σ= 2) assoc. with opacity (95%), small (77%), low density (45%), pleomorphic (23%)
INDETERMINATE calcifications
(Σ= 3) assoc. with opacity (70%), small (62%),
DISCREPANT DIAGNOSES
(Σ=4) assoc. with opacity (81%), small (50%)

NOTE: "Σ" indicates the number of different calcifications in each group. The percentages next to each descriptor indicate the average proportion of radiologists who used the term to describe the calcifications in each group.

candidate; if any such argument is proved true, it is considered to be sufficient evidence to either approve or eliminate a candidate. Alternatively, arguments can be represented as providing support "for" or "against" a candidate; such arguments add or remove support for a candidate but they are not sufficient individually to approve or eliminate the candidate. Further alternative ways of representing the strength of each argument (or the certainty associated with it) are currently being investigated.

In the current implementation, those descriptors which were found to be discriminating between benign and malignant calcifications and were never used for indeterminate ones (see Table 1) have been represented in PRO*forma* as "confirming" arguments for either candidate. Those descriptors which discriminate between benign and malignant but are also used for indeterminate calcifications are represented as arguments providing support "for" (but not "confirming") either candidate. Some combinations of descriptors were used for calcifications diagnosed by most of the radiologists

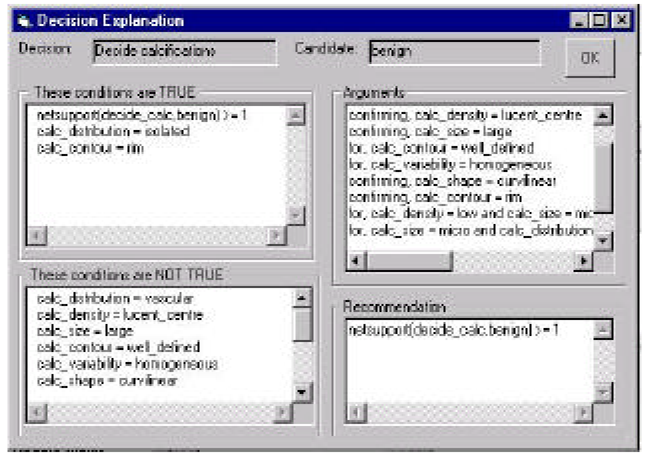

Figure 1. Screenshot of a decision explanation generated by the PRO*forma* Engine (in the Arezzo® development environment).

as either benign or malignant even if the individual descriptors were not found to be discriminating (see e.g. "smudgy" and "punctate" Table 2). These combinations are represented as arguments supporting (though not "confirming") either candidate.

More specifically, the following descriptors were represented as "confirming" arguments for the candidate benign: vascular distribution, curvilinear shape, large size, contour with a rim, lucent density centre, and isolated distribution. The descriptors represented as "confirming" arguments for the candidate malignant were: branching shape, ill-defined contour and orientation towards nipple. The appearances represented as arguments providing support "for" the candidate benign were: well defined contour, homogeneous variation and the combinations of descriptors "small + scattered + round" and "associated opacity + few (1-5) flecks". The arguments providing support "for" the candidate malignant were: pleomorphism and the combination descriptors "small + low density + associated opacity".

In CADMIUM II, symbolic processes are partly executed by the "PRO*forma* Engine", an interpreter which takes a PRO*forma* specification as input and identifies the appropriate task to be performed according to scheduling and other constraints, requesting actions and acquiring information. The Engine identifies the advice which is relevant to how the patient should be managed and presents it in the form of arguments for the different candidates for the current decision. Figure 1 shows an example of the output of the Engine when running the protocol implemented in CADMIUM II. It shows the explanation provided by the system for choosing the candidate benign when deciding the diagnosis of a particular calcification. The arguments for the candidate benign appear on the right of the figure. On the left are listed the conditions which are or are not known to be true about the calcification. The Engine obtains the evidence used in the argumentation by calling the image processing algorithms in CADMIUM II.

When the evidence available to the system is not sufficient to unambiguously diagnose a calcification as either benign or malignant using the noted arguments, the calcification will be interpreted as indeterminate. The user will then be presented with the arguments, if any, which support a “benign” decision and those which support a “malignant” decision.

CONCLUSIONS

An empirical study of the descriptive terms that expert radiologists use when making decisions about calcifications has yielded a set of salient features with a potential value for discriminating between malignant and benign mammographic appearances. These features have been used to inform the argumentation used in the decision support component of CADMIUM II. An advantage of this approach is that the advice provided will be presented at a level of description which is relevant and informative for the user.

An issue currently being explored is the generalizability of our results as both the data and subject samples in the study were fairly limited. One way in which we are addressing the issue is by a systematic review of published evidence relating the features of calcifications to the risk of malignancy. We have identified over a hundred papers dealing with this topic, which are currently being analyzed. We intend to complement the arguments so far implemented in the system with the information obtained from the literature and thus allow the system to provide evidence-based decision support. This paper is a description of work in progress. We anticipate that by combining robust image processing algorithms, a powerful formalism for the representation of medical decision making (PRO*forma*), a careful study of the domain literature and a clear understanding of the mammographic information radiologists regard as important, CADMIUM II will be able to provide a useful tool for assisting in the interpretation of mammograms.

Acknowledgments

This work is funded by the Engineering and Physical Sciences Research Council through the Health Informatics programme and by the Imperial Cancer Research Fund. The co-operation of Drs R. Given-Wilson, J. Davies, A. Schneider, J. Cooke, S. Rankin, I. Nockler and their colleagues is gratefully acknowledged.

References

1. Taylor P. Computer aids for decision making in diagnostic radiology – a literature review. Brit J Radiol 1995; 68: 245-957,.
2. Wu et al. Artificial Neural Networks in Mammography: Application to Decision Making in the Diagnosis of Breast Cancer. Radiology 1993; 187: 81-87.
3. Taylor P, Fox J, Todd-Pokropek A. The development and evaluation of CADMIUM: a prototype system to assist in the interpretation of mammograms. Medical Image Analysis 1999; 3: 321-337.
4. Taylor P, Fox J, Todd-Pokropek A. A model for integrating image processing into decision aids for diagnostic radiology. AI in Medicine 1997; 9: 205-225.
5. Cook HM and Fox MC. Application of expert systems to mammographic image analysis. Am J Physiol Imag 1989; 4: 16-22.
6. Stewart BD, Cairns AY, Ricketts IW.The analysis of breast abnormalities by computer. In: Gale A, Astley S, Dance D, Cairns A, editors. Digital Mammography. Amsterdam: Elsevier Science; 1994. p. 251-261.
7. Taylor P, Alberdi E, Lee R, et al. CADMIUM II: combining image processing and symbolic reasoning for computer-aided diagnosis. In: Proceedings of SPIE-2000. In press 2000.
8. Fox J, Johns N, Rahmanzadeh A. Dissemination of Medical Knowlege: the PRO*forma* approach. AI in Medicine 1998; 14: 157-181.
9. Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretations of mammograms. The New England J of Med 1994; 331(22): 1493-1499.
10. Simpson W, Neilson F, Kelly PJ. Descriptive Terms for mammographic abnormalities: Observer variation in application. Clin Radiol 1996; 51: 709-713.
11. D'Orsi CJ, Getty DJ, Swets JA, et al. Reading and decision aids for improved accuracy and standardization of mammographic diagnosis. Radiology 1992; 184: 619-622.
12. Whatmough P, Gale A, Wilson ARM. Do radiologists agree on the importance of mammographic features? In: Doi K, Giger ML, Nishikawa RM, Schmidt RA, editors. Digital Mammography'96. Amsterdam: Elsevier Science; 1996. p. 111-116.