A Decision Aid for Diagnosis of Liver Lesions on MRI

Rhea Tombropoulos, Smadar Shiffman, and Chris Davidson
Section on Medical Informatics, Stanford University School of Medicine

Abdominal magnetic resonance imaging (MRI) plays an important role in the evaluation of liver abnormalities. The interpretation of MR images requires expert training in a rapidly changing field. DAFODILL (Decision Aid for Diagnosing Liver Lesions) is a decision-support tool designed to aid radiologists in the diagnosis of hepatic lesions seen on MRI. DAFODILL uses a knowledge base of MRI findings and a belief-network inference engine to generate probabilistic differential diagnoses of the most commonly encountered hepatic lesions. DAFODILL performs limited image processing to identify clinically relevant features, which are presented to the user for confirmation before they are used by the network. Preliminary evaluation of an initial version of the system suggests that DAFODILL may be a useful tool for radiology residents and nonexpert radiologists in interpreting MR images of the liver.

INTRODUCTION

The evaluation of liver abnormalities is an important application of abdominal magnetic resonance imaging (MRI). The nature and origin of hepatic lesions are variable, ranging from insignificant incidental findings to malignant and life-threatening diseases. MRI provides a noninvasive method for visualizing the liver without risk to the patient, and plays an important role in the workup of hepatic lesions [1, 2, 3]. The interpretation of MR images of the liver is a relatively new area in the field of radiology, and requires special training that is not available to many radiologists. The two groups of users who stand to benefit the most from a decision-support tool for diagnosing hepatic lesions are (1) radiology residents who are actively training in MRI and (2) practicing radiologists who have some training in MRI but do not have broad experience in the subspecialty of abdominal MRI. For residents, the expert system would serve as an educational tool, as well as a decision-support tool. For practicing radiologists, the expert system would provide first-line decision support for consultation.

BACKGROUND

The primary task of medical image evaluation is to distinguish those aspects of an image that are normal or insignificant from those that indicate underlying abnormalities.

Expert Systems and Medical Imaging

Image-based diagnostic tools differ in the way they use images. Reported uses include image reference libraries [4], decision-support tools that use images as sources of findings [5, 6, 7], tools that perform image processing on manually marked regions of interest (ROIs) [8], and tools that recognize anatomical structures and classify tissues [9].

Our previous related work on probabilistic methods of image interpretation dealt with the evaluation of computed tomography (CT) and MR images of patients who had pituitary tumors [7]. This work focused on the probabilistic interpretation of reported features of diagnostic images without direct image analysis. We have extended this work to combine elements of image processing with diagnostic decision support.

Belief Networks

Belief networks, also referred to as *Bayesian networks* or *probabilistic networks,* provide a means of representing uncertain variables and their relationships graphically, where each variable is represented as a node in a directed graph. The edges between nodes are used to represent assertions of conditional independence. The assessment of conditional probabilities is based on the relationships among nodes as expressed in the network [10, 11].

Currently, the use of belief networks constitutes a small proportion of the work done in probabilistic reasoning in medical image analysis, although probabilistic networks have been used in general studies of visual recognition [12]. Belief networks have been applied successfully in the evaluation of microscopic pathology specimens [13].

DESIGN CONSIDERATIONS

The primary purpose of our work is to make expertise on evaluation of hepatic lesions readily available to practicing physicians.

Knowledge Representation

We found belief networks to be well suited for the task of representing relationships between diseases and features that are typically seen in MR images of hepatic lesions, particularly because of the coherent

0195-4210/92/$5.00

way in which the networks represent conditional dependencies. In addition to defining relationships between diseases and image features, the networks are also useful in identifying the conditional dependencies among features [10, 11].

The network that we designed initially included many dependent relationships between nodes in the network; we assessed many conditional probabilities at a level of granularity that made the network overly complicated. At a later stage, we eliminated certain relationships from the network by combining features into pairs that are meaningful for diagnosis. For example, the features *intensity* and *homogeneity* were combined into a single node, *intensity–homogeneity*, which decreased the complexity of the network for these features by a factor of 3.

Image Processing

We developed several automatic procedures for analyzing input images so as to increase the speed and accuracy of image-feature acquisition by the program. The parameters inferred by these procedures were used as the default values of the system that were presented to the user for verification and alteration. We hoped that allowing the user to override system computations would encourage user acceptance.

An important task for image processing is *segmentation*, which is the identification of a relevant ROI within an image. Reliable segmentation requires contextual knowledge and common sense, and is generally performed much more easily by humans than by computers. We simplified the image-analysis tasks by having the user perform segmentation.

User Interface

We developed the program in SuperCard [14], a HyperTalk-based authoring environment that allows for flexibility in the design of the interface and that also supports display of grayscale images. After considering several possibilities, we chose a screen layout that displays sets of related data and images in separate windows, thus allowing the user to maintain control of the amount of information that is visible at any given time.

KNOWLEDGE ACQUISITION

The DAFODILL knowledge base was constructed from information that we elicited from three sources: expert radiologists, the literature, and case studies.

We conducted a series of knowledge-acquisition sessions with two radiologists who are experts in the field of abdominal MRI. These sessions involved in-depth discussions of the specific features of hepatic lesions on MRI. The experts provided subjective estimates for most of the probabilities in the belief network. To supplement and confirm the subjective estimates of our experts, we searched the literature for information about hepatic lesions and their appearance on MRI [1, 2, 3].

We further supplemented the knowledge base by reviewing clinical cases from the Stanford Hospital MRI center. Data from these cases provided a baseline estimate for many of the probabilities in the network; in fact, these cases were the only available source of data for some of the knowledge in our knowledge base. For example, the presence of a pseudocapsule in hepatocellular carcinoma has been estimated to be between 50 and 90 percent. Our case data yielded a value of 73 percent, which our experts deemed reasonable and which we used as the value in our network.

SYSTEM DESCRIPTION

DAFODILL produces a differential diagnosis of hepatic lesions found on MRI studies, based on specific information about the features that characterize the lesions, as well as on general clinical information about the patient. The general flow of program execution is shown in Figure 1.

The system comprises three modules: a user interface, a belief network, and an image processor. The user interface manages user input into the system, as well as display of images and system output. It also acts as the task controller by managing the activation of tasks by the other modules in the system. The image processor makes inferences about various features of the user-specified lesion. These inferences and the clinical information input by the user are used by the belief-network module to compute the differential diagnosis.

User Interface

For a given case, the user inputs general patient data into the patient-information window. The system retrieves a series of MRI images based on the patient's identification number, and displays them to the user. The user looks at each image individually, and marks the ROI (i.e., the lesion in question) using a mouse. The user also points to a region on the image that contains normal-appearing liver tissue.

Once the user has indicated an ROI on an image, the window coordinates defining the ROI and the region of normal liver are sent to the image processor. The image processor then makes inferences about the size, homogeneity, and intensity of the lesion, which are in turn displayed. This process is repeated for each individual image in the study, for up to six images. In addition to making

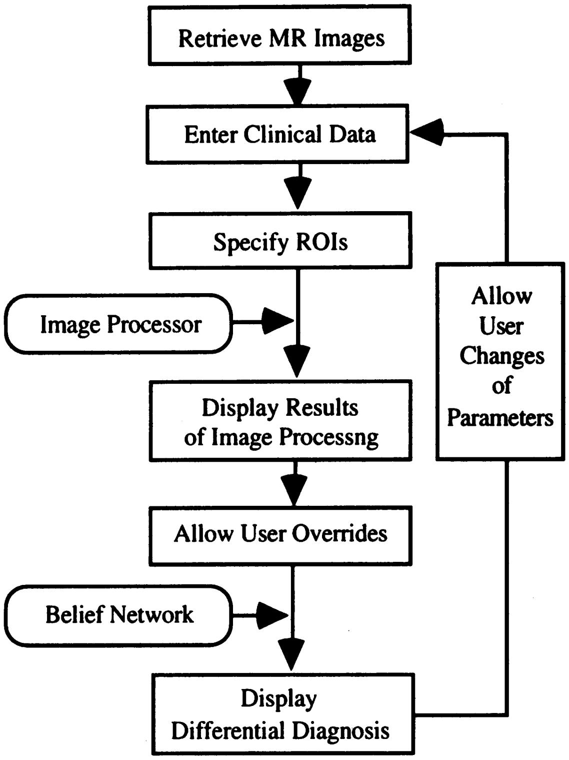

Figure 1. DAFODILL program execution. The rectangular boxes represent stages in the interaction with the user interface. The rounded boxes indicate the activation of support modules.

inferences about individual images, the image processor also makes inferences about the images as a group, such as the size of the lesion. These inferences are displayed in a separate summary window. The user has the ability to override any of the inferences by entering her changes into the various fields being displayed.

At this point, DAFODILL formalizes the feature inferences and other input by assigning an appropriate state to each corresponding node in the network. The assignment of states to nodes involves simple rules for several of the nodes in the network. For example, if the image processor infers that the feature *T2 relaxation time* has a value of *96*, then the state *greater than 90* is assigned to that node.

After the user interface assigns an appropriate state to each node, it sends the list of states to the belief network, whereupon the inference engine updates the nodes in the network and calculates the probabilities of all possible hepatic lesions. The probabilities are returned to the user interface, and are displayed to the user as a differential diagnosis.

Belief Network

DAFODILL uses ERGO [15], a commercial belief-network inference engine, to compute the likelihood of the 14 most commonly encountered hepatic lesions (Table 1). The belief network contains probabilities that define the conditional relationships among the 14 types of lesions and 13 specific features, 8 of which are image-related features, and 5 of which are patient-related clinical features (Figure 2).

Table 1. Hepatic lesions included in DAFODILL's belief network.

Hepatic lesion	A priori probability (%)*
hemangioma	21
metastasis, hypovascular type	19
metastasis, hypervascular type	16
metastasis, sarcomatous type	6
hepatocellular carcinoma	13
cholangiocarcinoma	2
lymphoma	1
fibrous nodular hyperplasia	1
regenerating nodule	3
hepatic adenoma	1
focal fatty infiltration	3
abscess	2
simple cyst	6
other	6

*A priori probabilities estimated from patients studied at Stanford Hospital from 1990 to 1992.

Image Processing

Patient images are displayed as 8-bit images, which are scaled versions of the 12-bit originals; the images are downloaded from the MRI scanner. DAFODILL's image processor exploits the complete image information by processing the original 12-bit image. The image processor uses several image-processing functions to infer characteristics about the ROIs that are specified by the user on the displayed MRI.

CURRENT STATUS AND EVALUATION

In the current implementation of DAFODILL, the features outlined in the previous section are integrated and functional.

Knowledge Base

The current DAFODILL knowledge base, which is maintained in the belief network, is constructed

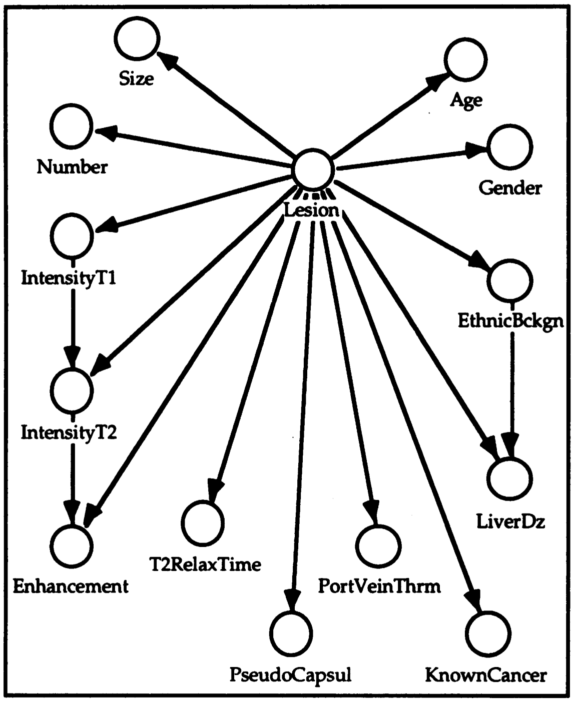

Figure 2. Belief network that constitutes the DAFODILL knowledge base. *Size:* size of lesion; *Number:* number of lesions; *IntensityT1:* intensity on T1-weighted image; *IntensityT2:* intensity on T2-weighted image; *T2RelaxTime:* T2 relaxation time; *Enhancement:* enhancement pattern with IV contrast; *PseudoCapsul:* presence of pseudocapsule; *PortVeinThrm:* presence of portal vein thrombosis; *KnownCancer:* presence of preexisting cancer; *LiverDz:* presence of chronic liver disease; *EthnicBckgn*: ethnic background of the patient; *Gender*: patient's gender; *Age*: patient's age.

from information that we elicited from expert radiologists during knowledge-acquisition sessions, as well as from data from case studies and the literature, as described earlier. Our experts were satisfied with the descriptions of most of the nodes in the current network, with the exception of *contrast enhancement,* which in its final form had ten values to allow for the possible variations in appearance of the lesions.

Image-Processing Routines

The following image-processing routines were implemented: identification of user-specified ROI for abnormal lesions and normal liver; determination of relative signal intensity and lesion homogeneity; calculation of T2 relaxation time; and identification of pseudocapsule around lesions.

Defining the threshold values for some of the image-processing functions was a tedious task. Additional fine tuning of the thresholds will be necessary to improve performance.

System Evaluation

We performed a preliminary evaluation of the program using three test cases of hepatic lesions. Certain of the image-processing routines were not integrated at the time of testing, so we evaluated the cases by having the user identify some of the image features without the help of inferences from the image processor.

In this initial uncontrolled evaluation, DAFODILL correctly diagnosed the hepatic lesions in the three test cases. In each of the cases, the correct diagnosis was made with a probability between 85 and 95 percent, with the probability of the second leading diagnosis between 5 and 15 percent.

FUTURE DIRECTIONS

Future work on DAFODILL should include three specific goals: (1) to refine the knowledge base as represented in the belief network, (2) to evaluate the accuracy of the image processor as well as implement additional image-processing routines, and (3) to evaluate formally the program overall to test its accuracy and utility.

Refinement of the Knowledge Base

The knowledge-acquisition task was difficult. Because our experts were unsure or inconsistent about certain probability estimates, we expected some error in the probabilities. The experts were confident about the choice of most of the features and values used by the system. There was some disagreement among experts, and it might have been useful to consult a panel of experts to achieve more accurate probability values.

More knowledge-acquisition sessions with expert radiologists will be necessary if we are to address those parts of the knowledge base that are incomplete or inaccurate. In particular, the features of contrast enhancement should be more accurately defined, as enhancement is a key aspect of interpretation of hepatic lesions on MRI.

The knowledge base could be expanded to include hepatic lesions that were not a part of the original problem domain. Specifically, lesions that are found in pediatric patients, such as *hepatoblastoma,* a type of tumor seen in young children, should be included.

Evaluation of the Image Processor

Although DAFODILL's image processor is simple, it is an important part of the overall design. After the routines are integrated, the accuracy of the image processor must be evaluated by recording the frequency with which users override the inferences that the algorithms provide.

Proposed Formal System Evaluation

Evaluation of DAFODILL's clinical and technical performance is a critical part of the system's future development. A formal evaluation of DAFODILL is necessary after the next phase of knowledge-base refinement and image-processing implementation is completed. The gold standard of such a study is the diagnosis of expert radiologists. It would be useful to test the effect of DAFODILL on the diagnostic accuracy of radiology residents and fellows in MRI of hepatic lesions and how they compare to the gold standard.

CONCLUSIONS

We conclude that belief networks provide a useful method for representing expert knowledge in the domain of medical image interpretation. We believe that DAFODILL, if fully implemented, could provide decision aid for nonexpert radiologists in the diagnosis of hepatic lesions on MRI. A formal evaluation is necessary to demonstrate the system's utility.

ACKNOWLEDGMENTS

We thank Winston Whitney, MD and Robert Herfkens, MD for their tremendous help in knowledge acquisition. This work was conducted with the support of the National Library of Medicine under grants LM-05305 and LM-07033.

References

[1] G.K. Kanzer., J.C. Weinreb. Magnetic resonance imaging of diseases of the liver and biliary system. *Radiologic Clinics of North America,* 1991 Nov, 29(6): 1259-84.

[2] D.D. Stark. MR imaging of focal liver masses. *Radiology*, 1988 Aug, 168(2): 323-6.

[3] R. Weissleder., D.D. Stark. Magnetic resonance imaging of the liver. *Magnetic Resonance Quarterly*, 1989 Apr, 5(2): 97-121.

[4] B.N. Nathwani., D.E. Heckerman., E.J. Horvitz., T.L. Lincoln. Integrated expert systems and videodisc in surgical pathology: An overview. *Human Pathology*, 1990, Vol. 21, pp. 11-27.

[5] D.W. Piraino *et. al.* Application of an artificial neural network in Radiographic Diagnosis, *Journal of Digital Imaging*, Vol. 4, no. 4, 1991, pp.226-232.

[6] J.A. Newell. Medical images and automated interpretation. *Journal of Biomedical Engineering*, 1988 Nov, 10(6): 555-61.

[7] H.C. Davidson., C.L. Andrews., A.G. Osborne., H.R. Warner. Expert system computer-guided diagnosis of sellar lesions. In: Association of University Radiologists Thirty-seventh Annual Meeting, Seattle, WA, 1989: 68.

[8] T.J. O'Leary., U.V. Mickel., R.L. Becker. Computer-assisted image interpretation: use of neural network to differentital tubular carcinoma from sclerosing adenosis, *Modern Pathology*, Vol 5, no. 4, 1992, pp. 402-405.

[9] M.G. Walker. Automated recognition of brain anatomy and pathology in MR and CT images. Technical Report KSL-90-38, Knowlege Systems Laboratory, Stanford University, 1990.

[10] E. Charniak. Bayesian networks without tears. *AI Magazine* 1991 Winter, 12(4): 50-63.

[11] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, Inc., San Mateo, CA, 1988.

[12] J.M. Agosta. Probabilistic recognition networks. PhD Dissertation, Stanford University College of Engineering, 1991.

[13] D.E. Heckerman., B.N. Nathwani. An evaluation of the diagnostic accuracy of Pathfinder. *Computers and Biomedical Research*, 1992 Feb, 25(1): 56-74.

[14] *The SuperCard Language Guide Version 1.5*, Silicon Beach Software, Inc., San Diego, CA, 1990.

[15] *ERGO User Manual*, Noetic Systems, Inc., Redwood City, CA, 1991.

[16] E.H. Smith. Complications of percutaneous abdominal fine-needle biopsy. *Radiology*, 1991 Jan, 178(1): 253-8.