

**COMPUTER ASSISTED
DECISION MAKING FOR IMAGE
UNDERSTANDING IN MEDICINE**

by
Paul Martin Taylor B.Sc. M.Sc.

Dept of Medical Physics and Bioengineering
University College London
Gower Street
London WC1E 6BT

Advanced Computation Laboratory
Imperial Cancer Research Fund
61 Lincoln's Inn Fields
London WC2A 3PX

Submitted for
the degree of Doctor of Philosophy
at the University of London

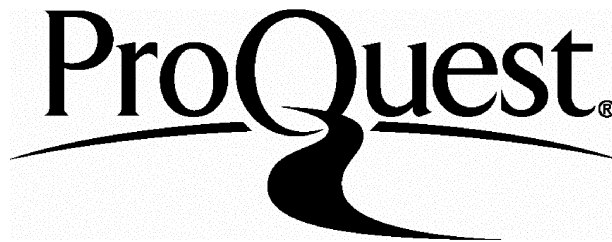
ProQuest Number: 10016126

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10016126

Published by ProQuest LLC(2016). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code.
Microform Edition © ProQuest LLC.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

*Leis na bha dhomh de bhreannachadh
gun d'rinn mi faileas stri`,
gun d'rinneadh gleachd le m' che'ill*

*(With all I had of apprehension
I put up a shadow of a fight;
my reason struggled)*

Sorley Maclean

ACKNOWLEDGEMENTS

Heartfelt thanks are due:

to my supervisors John Fox and Andrew Todd-Pokropek;

to David Ingram who gave me space to finish;

*to Claire Dicks-Mireaux, John Pritchard and Carol Young
of Great Ormond Street Hospital;*

*to Regina Pauli and the radiographers at the Jarvis Breast
Screening Centre;*

*to all the radiologists who contributed including Nigel
Barratt, Beatrice Barreau, Julie Cook, Jackie Davis,
Marie-Hélène Dilhuydy, Ruth English, Mike King,
Caroline Kissen, Mike Michel, Nick Perry, Basil
Shepstone, Kate Stoner and Kate Wharmsley;*

*to Lida Graupner, Saki Hajnal, Paul Krause, Simon
Parsons, Ali Ramenzadeh and Jeremy Wyatt, at the
Imperial Cancer Research Fund;*

to Chloe Hutton at University College London;

and to Jean McNicol.

Abstract

We are all familiar with the after-the-fact tone – weary, self-justificatory, aggrieved, apologetic – shared by ship captains appearing before boards of inquiry to explain how they came to run their ships aground, and by authors composing forewords.

John Lanchester

This thesis considers the different kinds of computer system which assist in the interpretation of medical images. These systems contain information represented in two very different ways: as images and as symbolic knowledge. If a decision aid is to provide access to all the information that might assist a radiologist, it must be able to employ information represented in images and as symbolic knowledge. Image processing is needed to provide the descriptions required by knowledge-based systems. Symbolic representations are needed to relate image data to the decisions radiologists take. This thesis sets out a design for decision aids which combine image data and symbolic representations.

The approach is based on a model of decision making, a symbolic decision procedure which constructs arguments, or lines of reasoning, about possible solutions to a problem. An extension to the decision procedure provides a model of three generic tasks in image interpretation: detection, classification and measurement. The extended decision procedure is implemented as a program which allows the symbolic decision procedure to draw on information obtained from processing images.

A generic architecture, based on the extended decision procedure, has been used in the implementation of two prototypes: one to assist in the interpretation of breast X-rays or mammograms and one to assist in the use of CT for the management of abdominal tumours.

The mammography system has been evaluated as an aid to the differential diagnosis of microcalcifications. An ROC analysis showed that radiographers who had been trained to interpret mammograms performed better on a differential diagnosis task when they had access to the information supplied by the system: the information was based in part on a symbolic representation of facts about microcalcifications and in part on the processing of digitised mammograms.

Contents

Chapter One: Radiology and Computers	12
1.1 Introduction	12
1.2 The Medical Context	15
1.2.1 Establishing the need for the examination	15
1.2.2 Performing the examination	18
1.2.3 Interpreting the image	26
1.2.4 Reporting the findings	30
1.3 Computer Support for Radiological Tasks	33
1.4 Summary	36
1.5 Chapter Outlines	37
 Chapter Two: Decision Aids for Radiology	 39
2.1 Introduction	39
2.2 Decision Support	40
2.3 Computers and Medical Image Interpretation	46
2.3.1 Image databases	46
2.3.2 Decision systems based on numerical methods	52
2.3.3 Expert systems	57
2.3.4 Image-processing systems	63
2.4 Conclusion	67
 Chapter Three: The Signal/Symbol Problem	 71
3.1 Introduction	71
3.2 Signal and Symbol Information	72
3.2.1 The signal representation of information	73
3.2.2 The symbolic representation of information	74
3.3 Systems Combining Signal and Symbol Information	76
3.3.1 Multimedia systems	77
3.3.2 Expert systems for image processing	81
3.3.3 Expert systems with image processing	85
3.3.4 Knowledge-based image segmentation systems	88

3.4 Discussion	99
3.4.1 Mapping between signals and symbols	99
3.4.2 The representation of signal data	106
3.5 Conclusion	109
Chapter Four: A Framework for Decision Support	111
4.1 Introduction	111
4.2 The Symbolic Decision Procedure	113
4.3 Decision Making in Radiology	121
4.4 The Extended Decision Procedure	123
4.4.1 Rules for proposing candidates and arguments	123
4.4.2 The interpretation rules	127
4.5 Supporting Image Interpretation Tasks	134
4.5.1 The detection of calcifications	134
4.5.2 The classification of calcifications	137
4.5.3 The measurement of tissue density	139
4.6 Discussion	142
Chapter Five: A Generic Architecture	147
5.1 Introduction	147
5.2 Principles for the Design of a Radiologist's Workstation	148
5.2.1 Integration of functions in a single environment	148
5.2.2 Combination of image processing and reasoning	149
5.2.3 A task-oriented interface	150
5.3 Components of the Architecture	150
5.3.1 User interface	152
5.3.2 Information sources	154
5.3.3 Information-processing components	159
5.4 Functional Components	161
5.5 Conclusion	164
Chapter Six: Decision Support for Mammography	167
6.1 Introduction	167
6.2 The Medical Context	167
6.3 Decision-Making in Mammography	170
6.3.1 The analysis of breast masses	171
6.3.2 The classification of microcalcifications	176
6.3.3 Assessment of tissue density	188
6.3.4 Conclusions	189
6.4 Image Processing for the Detection of Calcifications	192
6.5 Image Processing for the Classification of Calcifications	193
6.5.1 Magnin [Magnin1989]	194
6.5.2 Lefebvre [Lefebvre1991]	196
6.5.3 Patrick [Patrick1991]	198
6.5.4 Shen [Shen1993a, 1993b]	200
6.5.5 Wu [Wu1993]	201

6.5.6 Chitre [Chitre1994]	203
6.5.7 Nishikawa [Nishikawa1993, Nishikawa1994]	206
6.5.8 Lo [Lo1995]	208
6.5.9 Parker [Parker1995]	209
6.5.10 Discussion	212
6.6 Discussion	216
Chapter Seven: The Mammography Workstation	218
7.1 Introduction	218
7.2 Image Processing Operators	219
7.2.1 Detection of calcifications	219
7.2.2 The classification of calcifications	222
7.3 The Representation of Medical Knowledge	224
7.4 The Representation of the Protocol	230
7.5 Overview of the Prototype	232
7.6 Conclusions	236
Chapter Eight: Evaluation	237
8.1 Introduction	237
8.2 Accuracy of the Knowledge Base	237
8.3 Association between Knowledge and Image Processing	241
8.3.1 An informal evaluation of the image-processing	242
8.3.2 Agreement between radiologists and processing	246
8.4 Value of the Tool as an Aid to Decision Making	252
Chapter Nine: Decision Support for Abdominal CT	258
9.1 Introduction	258
9.2 The Medical Context	258
9.3 Decision Making in Investigations of Neuroblastoma	259
9.3.1 Diagnosis	259
9.3.2 Staging	260
9.3.3 Response to Treatment	262
9.3.4 Conclusions	263
9.4 Image Processing for the Management of Neuroblastoma	265
9.4.1 Detection of the Midline	265
9.4.2 Measurement of tumour volume	267
9.5 The Representation of Medical Knowledge	272
9.6 The Representation of the Protocol	275
9.7 Overview of the Prototype	277
9.8 Conclusion	278
Chapter Ten: Discussion	280
10.1 Introduction	280
10.2 The Potential for Decision Support in Radiology	280
10.3 The Signal-Symbol Question	283
10.4 The Extended Decision Procedure	285

10.5 The Generic Architecture	289
10.6 The Choice of Clinical Problem	292
10.7 The Mammography System	294
10.7.1 The construction of the knowledge base	295
10.7.2 Image processing	298
10.7.3 The evaluation of the mammography system	299
10.8 The Neuroblastoma System	300
10.8.1 The measurement of tumour volume	301
10.8.2 Decision support for staging and assessment	302
10.9 Conclusions	302
10.10 Future Work	304
Appendices	306
11.1 Image Processing Calculations	306
11.1.1 Compactness	306
11.1.2 Hu's Invariant Moments	306
11.1.3 Eccentricity	308
11.1.4 Elongation	308
11.2 Knowledge base used in the mammography system	310
11.2.1 Hierarchy of terms	310
11.3 Properties of calcifications	313
11.4 Positive signs for the possible diagnoses	314
11.5 The Mammography Protocol	317
11.5.1 Graphical Representation	317
11.5.2 Knowledge Base	320
References	326

List of Tables

Table 1: kinds of medical image referred to in this thesis	25
Table 2: the toplevel of the Symbolic Decision Procedure	126
Table 3: knowledge required for the example of a breast lump	127
Table 4: the three different image interpretation tasks	128
Table 5: a model to identify a region of interest	131
Table 6: rules describing the three image interpretation tasks	133
Table 7: rules and facts used to represent the detection task	137
Table 8: rules and facts used to represent the classification task	139
Table 9: rules and facts used to describe the measurement task	141
Table 10: a classification of circumscribed lesions [Tabar1983]	171
Table 11: spectrum of Breast Mass Appearances [Feig1992]	175
Table 12: characteristics of ductal and lobular calcifications	178
Table 13: microcalcifications seen in 1044 mammograms[Lanyi1987]	180
Table 14: classification of the American College of Radiology	182
Table 15: the predictive value of microcalcifications	185
Table 16: measures used in the classification of calcifications	214
Table 17: results from testing the four individual segmentors	221
Table 18: results from testing the three combined segmentors	222
Table 19 decision specifications for the mammography workstation	225
Table 20: the mean scores for the radiologists' ratings	241
Table 21: values for five calcifications	243
Table 22: weighted kappa values	248
Table 23: true and false positives, true and false negatives	256
Table 24: the INSS criteria for the staging of neuroblastoma	261
Table 25: the INSS criteria for the assessment of neuroblastoma	262
Table 26: results for the connectivity and threshold algorithms	271
Table 27: decision specifications for the neuroblastoma workstation	273
Table 28: stage 3 neuroblastoma	274

List of Figures

Figure 1: a planar X-ray of the compressed breast	19
Figure 2: an example of a breast X-ray or mammogram	20
Figure 3: a fan beam of X-rays is used to generate a 2-D image	22
Figure 4: an example of a X-ray CT slice	23
Figure 5: a model of static human vision, after [Pizer1990]	28
Figure 6: schematic outline of the examination of a visual display	29
Figure 7: Image Indexing by Content	50
Figure 8: one of the rules in Smets et al.'s system.	88
Figure 9: image data are combined with symbolic knowledge	100
Figure 10: AXON	101
Figure 11: I ² C	102
Figure 12: the Collins et al. system	104
Figure 13: AutoMEX	105
Figure 14: the decision procedure	117
Figure 15: the domino model	118
Figure 16: procedural interpretation of the rule describing detection.	130
Figure 17: the method used to measure the density of tissue	140
Figure 18: the operation of the rule describing the detection task	143
Figure 19: the components of the generic architecture	151
Figure 20: the image reporting function	161
Figure 21: the decision support function	162
Figure 22: the task-management function	163
Figure 23: the information retrieval function	164
Figure 24: the anatomy of the female breast	168
Figure 25: the five types of microcalcifications	177
Figure 26: a protocol [Dilhuydy1994]	184
Figure 27: a protocol [Monsees1995]	187
Figure 28: the knowledge model	226
Figure 29: the interface of the mammography workstation	233
Figure 30: the system used to display mammograms	234
Figure 31: the display of decision support	235
Figure 32: the segmentation of a cluster of calcifications	243
Figure 33: the segmentation of calcifications leads to errors	244

Figure 34: values of eccentricity and elongation	245
Figure 35: plots of mean radiologists' rating	250
Figure 36: plots of mean radiologists' rating	251
Figure 37: the decision support advice used in the evaluation	254
Figure 38: pooled ROC curves for four trained radiographers	255
Figure 39: an example of a CT scan for a patient with neuroblastoma	265
Figure 40: the output from the spinefinder program	266
Figure 41: a screendump of the neuroblastoma application	278

CHAPTER ONE

RADIOLOGY AND COMPUTERS

*Whilst my Physitians, by their love are growne
Cosmographers, and I their Mapp*

John Donne

1.1 Introduction

Consider the sequence of events through which a woman who has detected a lump in her breast is diagnosed as having cancer and given access to treatment. First she will see her GP, who will refer her to a breast clinic. There she will see first a surgeon and then a radiologist. A number of investigations will be performed. Almost certainly one of them will be a mammogram – a breast X-ray – performed by a trained radiographer. A sample of tumour cells will be taken and inspected by a pathologist. All of these experts – GP, surgeon, radiographer, radiologist, pathologist – will have undergone years of education and training and their expertise represents a scarce resource which must be used efficiently. Each encounter, therefore, is organised by support staff including receptionists, clerks, secretaries, nurses and managerial staff. Between these encounters, the opinions and findings of the experts are pooled, to allow

the most informed decision to be made at every stage. This requires that, at each stage, the experts have access not just to the information supplied by their colleagues but to other data about the patient and her family, as well as about the latest medical research and the locally available investigations and treatments.

The effective provision of medical care requires the careful management of different kinds of information: to ensure sound administration, permit clear communication and guarantee informed decision making. An obvious component of the required infrastructure is an appropriate computer system. Consider the role of the radiologist. Having met the patient, he or she must decide on the appropriate investigations. If these are to include a mammogram, the patient must be seen by the radiographer who will take the mammogram. The radiologist will then inspect it and provide a report for the surgeon. Advances in digital mammography mean that in the future the mammogram may exist only as an array of data in a computer. The trend towards electronic patient records and advanced hospital information systems suggests that a patient's history will exist only on computer. Medical databases will allow radiologists to obtain information about the available investigations and treatment options via computer. In such a scenario the computer will underpin much of the activity of the radiologist. What kind of computer system would provide him or her with the best possible tool?

A radiologist interpreting a mammogram would want a single computer system which combined the ability to display images with access to information about the patient and access to databases of medical knowledge and to recommended care pathways. He or she would want to use this computer to assist in interpreting the mammogram, processing the image in order to help detect abnormalities and classify certain features. This thesis is concerned with the design of a computer system that combines these features: an integrated radiologist's workstation.

Systems to help in the interpretation of medical images have been developed in a number of different areas of computer science, areas such as image databases, image processing, expert systems and numerical methods. One of the aims of the thesis is to evaluate this diversity of systems from a unifying perspective, one in which radiologists are viewed as decision-makers and computer systems are assessed for the contribution they can make to improving radiologists' decision making. The argument which will be made is that each of the different kinds of system has advantages and limitations, and there is therefore a strong case for exploring the extent to which different techniques can be used to complement each other. This, then, is the second aim of the thesis: to consider how the technologies developed within these different fields can be combined to provide a system within which different forms of decision support, based on different kinds of information, would be available for use as required. In particular, the question is asked: how can information represented in symbolic knowledge bases be combined with information represented in images?

Numerous papers have been published reporting optimistic results in tests of decision support systems, including 29 rigorous studies of the effects of such systems on clinical practice [Johnston1994]. Few have entered widespread clinical use. Recent papers have discussed the failure of research into decision support to change clinical practice and a number of explanations have been proposed. Researchers responding to the perceived failures of early decision support systems suggest that both effective performance and user acceptability require a co-operative model of decision making in which the abilities of user and machine are matched [Miller1990]. A third aim of this thesis is to develop a conceptual framework within which diverse forms of information can be made available to a radiologist as he or she requires. The final aim is to demonstrate the practicality of these ideas through the implementation of two prototype decision support systems aimed at quite different areas of medicine: the use of mammography in the investigation of breast cancer and the use of abdominal CT in the management of neuroblastoma.

The next section gives a brief introduction to the sequence of activities which are involved in carrying out a radiological investigation. This provides the background for a discussion of the role to be played by computers in assisting radiology.

1.2 The Medical Context

Much of the work that is done on the application of computers to the analysis of medical images is carried out in isolation from clinical practice. An attempt to provide practical assistance for radiologists making decisions about medical images ought to be based on a more thorough analysis of the tasks radiologists perform and how they perform them. The next four sub-sections consider four stages in the radiological process: establishing the need for an investigation, performing the investigation, interpreting the results and communicating the findings.

1.2.1 Establishing the need for the examination

The process of radiology begins when additional information is needed to make a decision about a patient. The need may be identified by a clinician, by the patient or, in the case of population screening, by some other authority. It is frequently assumed in discussions of radiological practice that the decision involved is a diagnostic one, although radiological investigations are also used to inform decisions taken in the management of patients with known diseases. Establishing the need for the examination is a key part of the interface between radiology and clinical activity, it involves focusing on the clinical problem, identifying possible procedures, determining their appropriateness and appraising their impact on subsequent strategy. In the past the radiologist would often not have played any part in this process, his or her role was simply to interpret the examinations requested by the referring physician. There is now a greater variety of investigative techniques available and a stronger awareness of the

need to minimise unnecessary investigations and radiologists may help in the selection of an appropriate procedure.

A number of papers in medical decision making, e.g. [Kuhns1989] and [Chang1990] have considered how the need for radiological investigations is assessed and have argued for the use of mathematical models in appraising their value. The concern here is to minimise the number of unnecessary investigations by considering the impact on diagnosis of the information obtained from an investigation. Medical diagnosis is a ‘hypothetico-deductive’ process, in which the clinician uses some initial information to make a tentative assessment of the probability of disease and this assessment then guides decisions about what information to gather, which investigations to request, and the new information in turn allows the initial assessment to be updated. It is important to note that the clinician is dealing in probabilities: his or her deductions are, in most cases, not being made from principles that have the status of natural laws but from knowledge gained through collective experience and from imperfect clinical data.

One model of radiological decision making [Kuhns1989] requires clinicians first to calculate the degree of certainty they should have in a diagnosis before selecting a treatment – thus setting a threshold above which an investigation is superfluous because the already available evidence provides an adequate basis for recommending treatment – and second to determine the impact on that certainty which a positive test result would have – setting a second threshold below which an investigation is unnecessary since even a positive result would not provide sufficient evidence to warrant treatment. This approach assumes both that the decision can be clearly structured, with the set of possible diseases and their indications laid out, and that reliable data will be available on the prior probability of diseases and of the probabilities of the diseases given a particular test result. Often things will not be so straightforward, and when they are this will frequently have resulted in protocols or guidelines being laid down for what should be done in a particular situation. Such protocols are a particular area of

interest in this thesis. It is assumed that there will be an increasing move towards the establishment and use of consensus guidelines for clinical practice [Audet1990, Grimshaw1993] and that this will be supported by – and create a demand for – decision support tools [Renaud1994].

Mendelson [Mendelson1995] describes how the American College of Radiology is drawing up appropriateness guidelines for radiological investigations. The aim of these guidelines is to provide an orderly sequence of studies most likely to assist in the diagnosis and management of a clinical condition. Guidelines are produced only for clinical conditions meeting a set of criteria, to do with the importance of the condition, the variability of practice and outcomes in the absence of guidelines and the existence of evidence on which guidelines may be based. A panel of experts, drawn from all the relevant professional specialities, is initially asked to consider a table of evidence based on 15-20 studies, classified according to the type of study, the number of subjects, the purpose of the study and the strength of the finding. Successive questionnaires are then sent to each member of the panel, to elicit information about the appropriateness of possible imaging techniques. The process is terminated when 80% of the panel agree, the guideline is then written to incorporate abstracted recommendations and indicate areas of uncertainty. The guideline, which must meet standards of validity, reproducibility, clarity, flexibility and applicability is then disseminated in different forms (as a report, a reference guide and a patient guide) through government, healthcare and other agencies. Guidelines for the appropriateness of radiological investigations are being drawn up in other countries and by other organisations. Sometimes these will be national guidelines, others will be drawn up by a particular institution for internal use.

The research described in this thesis seeks to combine support for radiological guidelines with support for image interpretation and, where appropriate, to structure assistance for decisions around models of ‘best practice’ which embody the same kind of information as the guidelines described above. The work represents an attempt to

explore the value of combining support for protocol or guideline based care with support for image interpretation. The work has not included a rigorous attempt to identify or validate a particular guideline or a set of guidelines, but rather to develop an approach which can accommodate various kinds of guideline.

1.2.2 Performing the examination

Once the need for an examination has been established, a request has to be made to the imaging facility where the investigation is to be performed and the patient has to be invited to attend. The investigation itself will normally be performed by a technician or radiographer and the radiologist will only be present in emergencies.

The underlying principle in all radiological techniques is the measurement of radiation to which the body is semi-transparent. Different techniques are characterised in part by the radiation they use (X-rays, gamma radiation, ultrasound, perturbations of magnetic field), in part by the relative positions of camera, anatomy and radiation source (whether the detected radiation has passed through the body, been emitted from it or reflected by it) and in part by how the image is constructed from the detected radiation (whether a single image is taken, or images taken over a period of time and amalgamated, or an image is created from a set of different projections).

The oldest form of medical imaging is planar X-ray imaging, or radiography, in which emitted photons pass through a patient and create a two-dimensional image, which is a projection of the three-dimensional distribution of the X-ray attenuating properties of a section of anatomy. In traditional film-based radiography, the photons are detected by a screen-film combination in which the photons are absorbed by a phosphor coating which emits light on absorption, the light exposing the emulsion of the film. There are now a variety of techniques for acquiring a digital radiograph, for

example, using a photo-stimulable phosphor plate that stores a latent image which can be read out by a scanning laser beam.

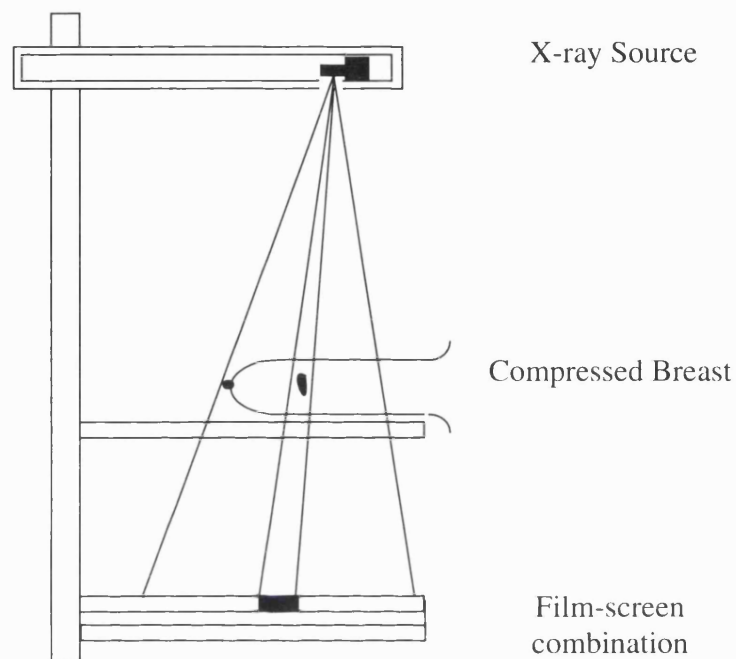


Figure 1: traditional mammography involves obtaining a planar X-ray of the compressed breast

One of the application areas considered in this thesis is that of planar X-rays of the compressed breast, known as mammograms. Breast cancer is the most common cancer in women in the UK and there are nearly 32,500 cases every year with about 14,500 deaths. Mammography is the investigation of preference throughout the detection, diagnosis and management of the disease.



Figure 2: an example of a breast X-ray or mammogram

It is a particularly demanding technique for a number of reasons. First, the difference between the X-ray attenuation of cancer and normal tissues is relatively small, which means that images must be created with photons emitted at relatively low energies, in order to optimise the absorption. Second, the range of transmitted exposures is relatively high, owing to the mix of tissue and radiolucent fat in the breast, which is a problem since higher exposure leads to higher film contrast only within a limited range of exposure values. Third, the objects which must be distinguished can be relatively small and spatial resolution must be traded against detector efficiency in determining the thickness of the film screen combination. A proportion of the photons arriving at the detector will not have been transmitted undisturbed through the breast but will have been scattered, the noise generated by this scattered radiation can, in part, be removed by a collimating grid, but this has the effect of increasing the dose required to expose the film adequately.

The difficulty of producing the best possible film mammogram, given these constraints, is in part due to the fact that the film is at once the medium of acquisition, of storage and of display. In digital mammography these processes can be separated and each one optimised. Digital mammography is the subject of much research, although at the time of writing little used in practice [Feig1995]. Much research is also being carried out into computer aids for mammography, since digital mammograms can conveniently be processed by computers, although the scarcity of digital mammography machines means that most of the research in this topic uses traditionally captured images which have then been digitised using a light-box and a CCD ('charge-coupled device') camera or by scanning a laser across the film.

The other class of images considered in this thesis is that of X-ray computed tomography (CT), in which anatomy is viewed slice by slice, forming a stack of two-dimensional images which can be used to create a three-dimensional representation. Each two-dimensional image is constructed from a number of projection images taken

through the body. In modern CT scanners a fan beam is rotated around the body to generate projection data from which a two-dimensional image of the slice is computed.

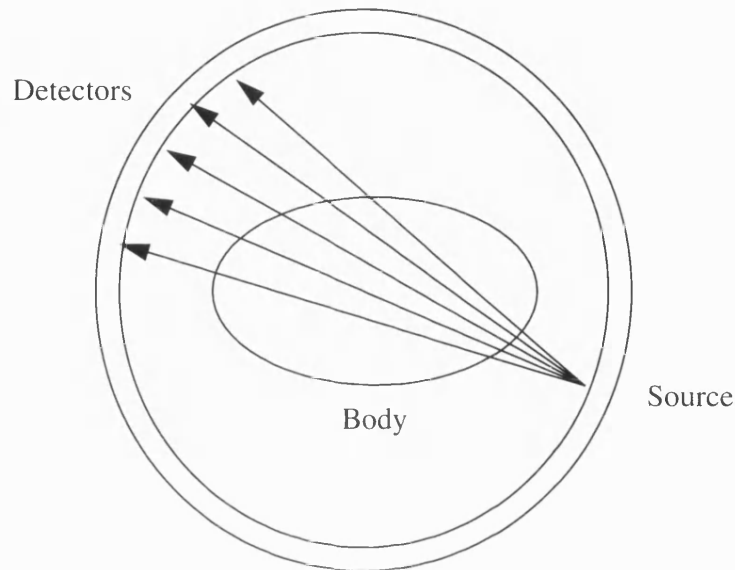


Figure 3: in X-ray CT a fan beam of X-rays is used to generate a 2-D image of a slice through the patient

If the slice is viewed as a grid of cuboids (known as voxels) then the value of detected radiation at any point is determined by the absorption properties of the tissue in every voxel through which it passes. If sufficient number of measurements are taken for different lines through the grid, then a value can be calculated for the absorption properties of the voxel. This value is called the Hounsfield number. A CT data set consists of a set of arrays of Hounsfield numbers. This is how CT data are stored, but generally not how they are displayed. The display of a CT scan usually involves altering the function used to map Hounsfield numbers into screen values so that the available contrast on the display medium is concentrated in the range of interest for the application. The application discussed in this thesis is that of abdominal tumours of

childhood; the interesting contrast is therefore between the tumour and the tissue making up the surrounding organs.

Mammography and X-ray CT are the two image modalities discussed in detail in later chapters. A number of other techniques are discussed in considering previous research in decision support for medical imaging. These are briefly introduced in the remainder of this section.



Figure 4: an example of a X-ray CT slice

X-rays are generated externally to the body and passed through it to create an image of anatomy. In another class of medical images, radiation is generated inside the body and detected externally to create an image of biological function. The distribution around the body of an injected radiopharmaceutical will depend on blood flow, blood volume and a variety of metabolic processes, so radiation detected using a gamma or scintigraphic camera, can provide information about these processes. In these images,

unlike X-ray images, the point of emission is not known, so some form of collimation is required to provide information about direction.

This technique, known as Nuclear Medicine or Scintigraphy is used to create both planar and tomographic images. Planar scintigraphic images may also be considered in two categories: static images, 2D single view images of the distribution of activity at a moment in time, and dynamic images, in which multiple images are taken over a period of time that may be milliseconds or a few hours. Important classes of dynamic planar scintigraphic images include those used to measure cardiac and renal function. Emission computed tomography is usually considered as two separate modalities, SPECT (single photon emission computed tomography), using radioisotopes where a single gamma ray is emitted per disintegration, and PET (positron emitted tomography) where two gamma rays are emitted simultaneously when a positron from a nuclear disintegration annihilates in tissue. Owing to their high cost, PET scanners are found only in major research institutions and have found few applications as yet in routine clinical practice.

A third kind of radiation is ultrasound. Like X-rays ultrasound is generated outside the body and attenuated differently by different tissue types. Unlike X-rays, it is usually the reflected rather than the transmitted signal which is measured. Since the speed of sound is slow enough to allow the time between pulse and echo to be measured for distances travelled in the body, the doppler shift can also be measured and ultrasound used as a functional imaging modality in the measurement of blood flow. Equally the speed allows all the data for an image to be collected in time to present, for example, a real time image of the moving heart. Ultrasound has the further advantages of being relatively cheap and low-risk. The principal uses of ultrasound are in obstetrics, gynaecology, and in abdominal, cardiovascular and superficial imaging

Name	Radiation	Configuration	Construction
X-ray	X-ray	Transmitted	Planar
Angiogram	X-ray	Transmitted	Difference Image
X-ray CT	X-ray	Transmitted	Computed Tomography
Ultrasound	Ultrasound	Reflected	Planar
Doppler Ultrasound	Ultrasound	Reflected	Time of flight calculation
Nuclear Medicine	Gamma Rays	Emitted	Planar or Difference Image
SPECT /PET	Gamma Rays	Emitted	Computed Tomography
MRI	Radio Waves	Emitted	Computed Tomography

Table 1: kinds of medical image referred to in this thesis

The other major imaging modality which should be discussed here is magnetic resonance imaging or MRI. Like X-ray CT, MRI is used to create sets of images providing a 3D representation. However, while X-ray images provide a map of electron density, which relates to physical density, MRI can be used to obtain information about proton density (or other relevant nuclei), and about tissue characteristics, such as the freedom of hydrogen containing molecules and the proportion of water in different regions of the body. MRI can also be used to obtain images of function. The basis for the radiation used here is the magnetic field created by orbiting protons in the molecules of the body. The net magnetic moment of a sample of nuclei is shifted when placed in a rotating magnetic field. This shift can be detected with a suitable coil. In fact different pulse sequences can be generated and measured to give information of different kinds. Saturation recovery, inversion recovery, spin/echo, longitudinal relaxation (T1) and transverse relaxation (T2) sequences are all used in constructing MRI

images. Spin/echo, T1 and T2 images are particularly important; the latter two provide information about vibrational motion in the lattice of molecules.

There are other methods of imaging the human body, such as the use of infra-red light to trans-illuminate, a technique which is used in the investigation of skin melanoma, and techniques which use visible light detected by miniature cameras inserted into the body, but the above represent the principal methods for which computer decision aids have been developed.

These different classes of image have quite different characteristics in terms of spatial resolution, dynamic range and signal-to-noise ratio and are used for very different purposes. In each, however, some form of radiation is used to create a signal which carries information about the body and stores it in a form which the human eye can read. This 'reading off' of information stored in images is so effortless a task for human beings that some imagination is required to understand what it must involve. This process is discussed in the next sub-section.

1.2.3 Interpreting the image

Designing appropriate tools to assist in the interpretation of images requires some understanding of how images are interpreted by radiologists. Perception, at the most basic level, involves a response to luminance values. At a more abstract level it involves the detection of significant form in these responses. The process of image interpretation is also an active one, and there are important questions concerned with how the image is searched for potential abnormalities.

Pizer [Pizer1990] describes a model of image perception which, although not universally accepted by the vision community, offers an adequate guide for this introduction. The model in Figure 5 is split into three stages: front-end, intermediate and

high-level vision. Perception starts with the spatial distribution of luminance values detected on the retina with a logarithmic sensitivity. The next column consists of samplings over fields within the retina. Most of these receptive fields signal the result of a comparison of the amount of light on a field of the retina with the average amount falling on the surround. The information transmitted from the receptive fields is therefore about intensity changes in the image and not about absolute intensity values. There are many different overlapping receptive fields and receptive fields of many different sizes. The result is that the early vision system provides information about changes at different spatial scales. The system also responds to higher order changes and so transmits information about orientation and curvature. The local measurements of features such as edges, bars and corners are combined in a way which allows edges corresponding to sharp luminance changes to be linked by 'subjective' edges which do not correspond to intensity changes in the image. The edges and corners perceived at different spatial scales are thought to combine in some way with a hypothesis, which may be generated by a tentative decision or by expectation.

A radiological investigation is performed with the aim of answering a particular question. The nature of this question determines the task which the radiologist performs in interpreting the image. He or she may be attempting to detect any of a class of known possible abnormalities, assessing the likelihood that a particular feature has a malignant cause or measuring the change in size of a tumour following treatment. These different tasks define sets of hypotheses which guide the radiologist in his or her reading of the image.

The human visual system is able to adapt to an extraordinary range of luminance values. However, for any given level of adaption it can make distinctions only within a relatively narrow range of luminance values. The properties which give the system its range of sensitivity to objects of different scales and luminance, also make it an unreliable indicator of absolute luminance. Judgements about the relative luminance of two separate regions within an X-ray, or other image, are not therefore

reliable. The role of hypotheses means that what is detected depends, to an extent, on what is looked for in the image. A faint or subtle finding to which a radiologist has been directed in some way is therefore much more likely to be perceived.

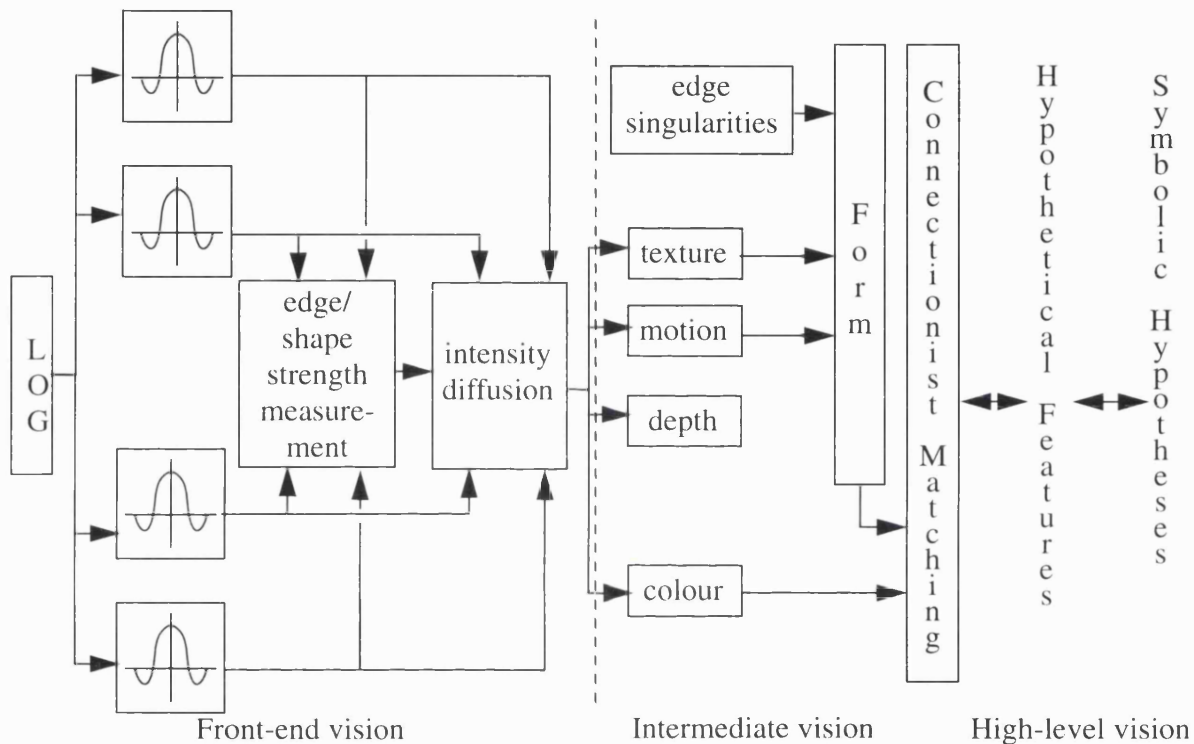


Figure 5: a model of static human vision, after [Pizer1990]

Greenes [Greenes1989] distinguishes two tasks in the interpretation of images: the detection and classification of features. This distinction is accepted by a number of authors [Gale1993a, Swett1993]. Gale presents a detailed account of how images are searched by radiologists, distinguishing between the 'somewhat parallel' initial processing of the image within the first glimpse and a second phase in which the display is serially examined using eye-movements. The model also provides a role for experience in furnishing radiologists with hypotheses. Kundel [Kundel1972] showed clear differences in how trained and untrained observers scan chest X-rays, variations

on the appropriate section of the image and in approximately 30% of cases the nodule received prolonged visual attention. Thus roughly one third of errors were errors of search, one third of detection and one third of classification. These findings have been used by researchers in computer aided radiology to justify work in enhancing images, providing prompts and in providing decision aids for classification, e.g. [Swett1993]. The roughly equal distribution of the errors across the three classes of error is a strong argument for providing a decision support tool which is capable of assisting in search, detection and classification. The role played by hypotheses suggests a potential role for a system which can provide information about possible diseases and diagnoses.

1.2.4 Reporting the findings

The next task in the radiological process concerns the reporting of radiological findings. Many authors identify this as a common source of problems. Robertson [Robertson1989] describes a study of communication problems in a mammography screening programme. In the course of the 11-week study 1,404 screening mammograms were taken, of which 63 required additional evaluation or biopsy. Written reports were sent to the referring physician and, in addition, the physician's office was notified by telephone of the need for additional investigation or biopsy. An average of 2.5 months after the initial examination was performed, the 63 cases were followed up: computers in radiology and pathology departments were searched for information about additional mammography studies or pathology results. Where no results were found, the relevant physicians' offices were contacted by telephone. These steps were repeated at fortnightly intervals. In some cases the patient was contacted directly. When the cases were first followed up, at 2.5 months, no action had been taken in 40 out of the 63 recommendations. After 3.5 months no action had been taken in 10 out of 63 studies and at 4.5 months four had still not undergone the recommended additional investigations.

Robertson lists the various explanations offered for the breakdown in communication but gives no indication as to their relative frequency. It seems that problems can occur at every stage in the process. Some letters were sent to the wrong address. In some cases the referring physician had moved on. In some cases the letter was misfiled. In other cases it was misunderstood by the referring physician and in others the difficulty was in getting in touch with the patient or with persuading them to comply with the recommendation. Robertson notes that the particular problem with screening is that 'no news is good news' and therefore when a letter fails to reach its intended recipient the failure is not noticed unless the sender follows up the communication.

D'Orsi [D'Orsi1995] describes in detail the communication issues associated with mammography. Reporting occurs at two levels: the communication of results and recommendations in lay language to the patient and the technical medical report to the patient's healthcare provider. Robertson cites a number of studies which have shown the importance of providing a report directly to the patient, and have shown that this results in improved compliance with recommendations, fewer delays in diagnosis and reduced confusion over later treatment options. Problems surrounding the sharing and communication of information with patients have been identified as one of the most influential factors in patient decisions to sue for malpractice.

The technical report to the patient's healthcare provider has also been the subject of study. D'Orsi describes how the American College of Radiology has supported the development of a standard approach to reporting which includes a statement of which elements should be included in a report and a lexicon of appropriate terms to be used in describing mammographic features. The aim is to encourage concise, understandable reports. The report should begin with a brief initial statement of the reason for the examination, the breast composition, significant findings, comparison with previous images (if applicable) and an overall assessment and recommendation. It should then document any relevant clinical concern and include a

statement of the communication that has taken place with the patient and the healthcare provider.

D'Orsi lists the statements that can be used to describe composition, the different abnormalities which can be described, the properties which may be attributed to them and the manner in which these properties should be expressed. Some authors e.g. [Heilbrun1994] have taken exception to this lexicon. Heilbrun states that the job of a mammographer is to recognise three kinds of lesion and classify them as benign, probably benign, suspicious and highly suspicious. Any findings which cannot be so classified require further analysis. Experienced mammographers have no difficulty in doing this, and in reporting succinctly that they have done so, inexperienced mammographers attempt to cover up their lack of confidence with wordy equivocations. It is argued that terminology is not therefore the root problem. Another difficulty is that the terms in the terminology concentrate on morphology and this inhibits the development of thinking about the underlying processes which cause abnormalities.

Coding schemes, such as that proposed by the American College of Radiology are becoming more widely used throughout the medical community. This is an important issue for the many computer based decision aids which require that relevant information about a patient be encoded in a form that is intelligible to the computer. The question whether the contents of radiological reports can be represented in this form is of crucial importance to this thesis. Coiera [Coiera1995] argues that coding systems can provide a practical basis for managing the language of medicine so long as it is understood that they define a limited and consensual language. The approach taken in this thesis has been to assume that codified medical terminologies can be developed for restricted tasks where the intended users and the context of the intended use are understood. A standardised reporting language for particular classes of medical investigation is therefore assumed not to be an impractical prerequisite for a radiological decision aid.

1.3 Computer Support for Radiological Tasks

The different tasks involved in diagnostic radiology have different information requirements and are amenable to different forms of computer support. A radiology department will normally exist within a hospital which will have its own information system providing for the storage, capture and transmission of patient-related and other information (a hospital information system or HIS). There will often be a separate computer system providing these functions within the radiology department (a radiology information system or RIS). There may be, in addition to these systems, some form of computerised patient record or electronic healthcare record (EHCR).

A number of the imaging modalities will involve digital capture, storage and display of images. There may be another system for the archiving and communication of images independent of the individual imaging systems (a picture archiving and communication or PACS system). There are clearly considerable benefits to be obtained from the integration of these different systems. Such integration or 'interoperability' is now a major topic of research in medical informatics and many groups are working on the development of standards and formal models that will facilitate the development of systems which can share information. Two directly relevant projects are the development of the DICOM standard and the MIMOSA approach to modelling PACS systems.

The American College of Radiology and the National Electrical Manufacturers Association jointly developed a standard for Digital Imaging and Communication (DICOM) which was published in 1993 [DICOM1993]. Continuing work on this standard has involved working groups from European and Japanese standardisation bodies and has been adopted by vendors of imaging devices and clinical workstations. The standard covers such topics as the encoding of image data and file transfer protocols as well as notions like the definition of Information Objects which allow the

representation of concepts such as 'patient' or 'visit' and the specification of operations performed on Information Objects as well as the expected behaviour of the user and provider of the operation.

The MIMOSA¹ project [Garfagni1994] developed an approach to modelling medical image management systems which tackles the issue of inter-operability between PACS and HIS/RIS. The project, which was supported by the EU as a research rather than a standardisation initiative, attempts to reconcile two different views of medical imaging: the HIS view which considers medical images as a subset of medical information and the PACS view which is concerned with the acquisition, processing and storage of images without being concerned with why they are produced or how the information they contain is used. The MIMOSA model consists of three loosely coupled models: the data model describes the structure and relationships of the data represented in the system, the functional model describes what the system does, the dynamic model describes the behaviour of the system over time.

It is worth reflecting on the ultimate goal of this work. It is theoretically possible to have a filmless, paperless hospital in which all information, image-based and otherwise, exists only on computer. This is not necessarily desirable. Indeed, CT images which can only be created by computer and which must be viewed on a computer screen in order to establish the appropriate display parameters, are routinely printed on film to be seen by the radiologist. Film offers many advantages to the radiologist: they can pick it up, view it next to other films, scan large numbers of images and carry them from office to office.

The development of networks of communicating computer systems supporting the different tasks performed in radiology will require balancing the advantages to be

1. Medical Image Management in an Open Systems Architecture

gained by standardisation and computerisation with the flexibility and familiarity of traditional ways of working. Consider the four tasks described in the previous section.

- Establishing the need for a radiological examination requires communication between a referring clinician and a radiologist. This might be supported by a computer, if both clinicians regularly use email or video-conferencing, but it may be more conveniently done some other way. It may involve reference to guidelines, it could involve reference to research material. Again, this kind of information can be conveniently accessed by computer, but it doesn't have to be. It could involve calculations or inferences made by a computer on the basis of some form of stored knowledge.
- Once the need for an examination is established the examination must be performed. The efficient use of imaging devices and the attendant personnel requires the use of computerised scheduling systems. Such systems are generally provided with radiology information systems and the basic concept, that of the worklist, is being defined in standards to allow interoperability of PACS and RIS systems. Many image generation techniques are now digital and require computers. Others may become computerised.
- The interpretation of images, if at some point they exist in digital form, may be facilitated by computer systems which either enhance, analyse or interpret the image or which provide information that the radiologist can use in making a decision.
- The generation of radiological reports can be assisted by systems of menus which provide access to standard lexicons, or by systems which provide access to reporting standards. Voice recognition and word-processing software can ease the creation of a written report and information systems can be used to manage the sending out of reports and appropriately timed follow up letters.

The research described in this thesis focuses on the kind of computer system which would be used by a radiologist to help make a decision about an image. The work, however, is also guided by a desire to consider this as part of a process of care and to bear in mind that a computer system designed to support part of a process must sit within a network of inter-operable systems supporting the whole process.

1.4 Summary

The opening section of this chapter set out four aims for this thesis:

- to consider the range of systems which could potentially improve medical decision making from a unifying perspective
- to consider how different kinds of system could be integrated into a single decision aid
- to propose a generic model for decision support tools which allows users to access relevant information
- to develop prototype systems which demonstrate the practicality of these ideas

The following section described four different activities involved in carrying out a radiological investigation: establishing the need for the investigation, performing the examination, interpreting the results and providing a report. Each of these tasks can be supported by the different computer systems which are currently in place in most radiology departments. A number of projects are now working on the integration and inter-operability of these different kinds of system. That work provides the context within which the research described in this thesis should be considered.

1.5 Chapter Outlines

Chapter Two reviews existing work on computer aids for decision making in diagnostic radiology. Systems developed in a number of different areas of computer science (image databases, numerical decisions aids, expert systems, image processing systems) are identified and reviewed. The value of the different approaches is assessed and it is argued that a system capable of drawing on different kinds of information source is desirable.

Chapter Three considers the problems involved in designing a decision aid to handle different kinds of information. Examples of existing systems which attempt to combine image processing and symbolic reasoning are considered.

Chapter Four sets out an approach intended to combine image processing and symbolic reasoning. The approach is based on a logical model of the processes involved in making decisions about images. In the model, decision making is viewed as involving the proposal of candidate solutions and then the consideration of arguments for and against these solutions. In the case of decisions made on the basis of the interpretation of images, some of the arguments may rely on information about what has been detected on the image, about how it can be classified or on measurements of the contents of the image. These different processes can all be described using logical rules, rules which can be implemented as a logic program and used as the basis for a decision support system.

Chapter Five gives details of a generic architecture for decision aids drawing on different information sources. The architecture sets out three components of a design: a set of displays which together make up the user interface, the set of information sources and a set of information processing modules – including the implementation of the

model described in Chapter Four – which draw on the information in the different sources to generate new information, which is then used to provide decision support.

Chapter Six describes the background to the prototype decision support system developed to assist in the differential diagnosis of calcifications on mammograms. The radiological literature on decisions made in the interpretation of mammograms is reviewed, concentrating on the problem of the differential diagnosis of microcalcifications. The application of image processing to this problem is considered and work in the area reviewed. The potential for a decision aid for the differential diagnosis of microcalcifications drawing on both image processing and symbolic reasoning, is assessed.

Chapter Seven gives a detailed account of the implementation of the prototype mammography system, concentrating on the development of the knowledge base and the selection of image processing measures. Chapter Eight describes the evaluation of the prototype mammography system. Chapter Nine describes a prototype system developed to assist in the staging of neuroblastoma, a tumour of childhood, on X-ray CT scans. Chapter Ten discusses the conclusions of the thesis, and highlights some directions for future work.

-

CHAPTER TWO

DECISION AIDS FOR RADIOLOGY

*The nurse conducts the exercise
in limited extinction;
the plate prepared, the source exposed
with no more than a glib nictation.*

*I am told to get dressed and go home.
Later, in my absence
the doctors make their guesses from
the holes left by my bones.*

Don Paterson

2.1 Introduction

This chapter reviews research into computer aids for radiological image interpretation. The aim is to bring together work from quite disparate areas of computer science and to consider the contribution it could make to improving the quality and efficiency of radiological decision making. Attention is therefore restricted to systems which would come into play when or after the image is displayed, and which would be used by a radiologist confronting a clinical dilemma. The various ways in which

computers can be used to create new kinds of image, the creation of tomographic images, the reconstruction of three-dimensional images, the enhancement of digital radiographs, the segmentation of MRI images and the registration of images of different modalities are not covered. Electronic information sources such as hypertext systems and medical databases are also excluded since, although they could assist in making decisions, they are not specifically designed to do so.

2.2 Decision Support

Shortliffe [Shortliffe1991] includes all systems designed to help health professionals make decisions within the definition of clinical decision support systems. He notes that “in a sense *any* computer system which deals with clinical data or medical knowledge is intended to provide decision support” and suggests that it is accordingly useful to consider three types of decision support function:

- tools for information management, which would include hospital information systems and bibliographic retrieval systems which make information and knowledge available to clinicians but do not help them apply it to particular cases
- tools for focusing attention, which would include systems which provide reminders or detect abnormal values, hence which provide advice on the basis of general rules covering different situations
- tools for a patient-specific consultation, which includes only those systems which provide tailored advice on the basis of patient-specific information

In practice, systems in the first of these categories are rarely considered in accounts of decision support systems and where a distinction is drawn between systems in the latter two categories, it is only systems of the third category that are considered as decision support systems. For example, Wyatt and Spiegelhalter [Wyatt1991] define

medical decision support systems as “active knowledge systems which use two or more items of patient data to generate case-specific advice”.

The term decision support system indicates a development in the conception of the role of knowledge-based systems, so that they are viewed as tools which can offer advice, but are designed to do so in a supporting role. An early definition [Keen1978] of decision support systems states that:

Decision Support Systems (DSS) represent a point of view on the role of the computer in the management decision-making process. Decision support implies the use of computers to:

- 1. Help managers in their decision processes in semi-structured tasks.*
- 2. Support, rather than replace, managerial judgement.*
- 3. Improve the effectiveness of decision making rather than its efficiency.*

The notion of ‘semi-structured’ tasks is crucial to this definition. These systems are not designed for tasks where a clear and rigid structure makes a completely computerised system a practicality, nor for those where there is no structure on which to base the design of such a tool, but rather for tasks between these two extremes. Johnston et al. [Johnston1994] found that decision support systems in medicine were more frequently shown to be successful when applied to tasks such as prescribing than when applied to diagnosis. It could be argued that prescribing is a better exemplar of the notion of semi-structured task than diagnosis, since – at least for a GP – diagnosis is a confrontation with all the variety, complexity and ambiguity of human life whereas prescribing is only attempted once the available data has been rendered in medical terms.

A successful advice-giving system depends on an understanding not just of the domain of expertise to be tackled but also of the decision problems in that domain for which a computerised aid may be of assistance. Decision support systems are defined as systems which attempt to support, rather than to replace, human judgement. If the developer of a decision support system is to respond effectively to this requirement he or she must have some conception of how human decision making works, and this conception, however simple, must guide the design of the system.

Shortliffe proposes the following list which reveals the differences between decision support systems.

- *System function*: what is the question which the advice provided is intended to help answer? Is it of the form “what is true?” or “what is to be done?”
- *Mode of giving advice*: some decision support systems provide advice on request, others have an ‘active’ mode of operation so that users are automatically sent reminders or warnings.
- *Consultation style*: some systems operate as consultants, so that the user provides information about a particular problem, on the basis of which the system is able to proffer advice, other systems expect the user to enter a judgement about a particular problem and the advice consists of a critique of the user’s reasoning or conclusions.
- *Methodology*: decision support systems can be distinguished on the basis of the techniques used to generate advice. Three different methodologies are identified by Shortliffe. The first is Bayesian probability, in which the most likely explanation of a patient’s state is determined by combining information about the patient with probabilistic information about how these parameters relate to the possible explanations. The second is decision analysis, in which information about probabilities is augmented by information about the relative value of different outcomes, and a formal structure, or decision tree, of outcomes and decisions is imposed on the problem. The

final methodology identified by Shortliffe is the use of AI techniques, such as those described in the section on knowledge-based systems.

- *Human-computer interface*: is the system integrated into a hospital network so that the data required can be obtained automatically, or must the user enter all the pertinent facts about a patient in the course of the consultation?

Shortliffe also notes eight reasons why clinicians resist the idea of decision support systems: fear that they will interfere with rapport between doctor and patient, fear of a loss of control, inertia, doubts about the capabilities of the machine, doubts about the value of artificial intelligence, legal concerns, the difficulty of data entry and the difficulty of learning how to use new technology.

Two issues do much to determine the acceptability of decision support systems. The first is to do with what Shortliffe terms 'transparency of function', by which is meant the extent to which the user can be confident of the basis for the advice being offered. A great deal of value is placed on systems which provide an explanation of the reasoning behind the advice they offer. This is a distinct advantage of systems such as those based on the techniques of symbolic reasoning associated by Shortliffe with artificial intelligence. This is related to the wider question of the extent to which the design of a system is based on a theory of how the supplied information will help the user make decisions. The second issue concerns the extent to which the decision support system is integrated into a wider computing environment which supports routine data management. This is widely thought to be a key determinant of the acceptability of decision support systems, since integration implies that the human-computer interface problems in data entry have been resolved, together with the difficulties of learning to use new systems.

As previously mentioned, Wyatt and Spiegelhalter [Wyatt1991] define medical decision aids as "active knowledge systems which use two or more items of patient

data to generate case-specific advice". Adopting this definition, it follows that all such systems must have both a computerised information source and a mechanism by which a user may quickly obtain relevant information from this source. Four quite different kinds of information source have been employed in decision aids for radiological image interpretation: images, algorithms for manipulating images, medical data and symbolic representations of medical facts. Different computations and methods of interaction are appropriate to each of these forms of information and there are therefore four distinct classes of decision support system, each based on one kind of information source. These are discussed in turn in the next four sections of this chapter. Systems which attempt to combine more than one kind of information source will be discussed in the next chapter. First, it is necessary to consider criteria to be used in assessing the research in the different areas.

If a decision aid comprises a computerised information source and a mechanism for making available relevant information from that source, then the first set of issues to be considered concerns the construction of the information source. The information it contains must be accurate, it must be complete (within limits which are understood by the user) and it must allow the retrieval of information.

The mechanism by which information is made available from the information source must also be well designed. There are two distinct issues here. The first concerns the interaction with the user. Since the decision support system must select *relevant* information from the source, it must first obtain information about the problem which allows an assessment of relevance: there must be an interaction with the user. The setting in which the system is to be used places clear constraints on what kind of interaction is practical: if a large number of images are being interpreted extremely quickly, as in screening, a useful system will require only minimal input from the user.

The other issue concerns the value of the information returned. One of the key factors here is the extent to which the designer has understood how the user will make use of the information in making a decision. Ideally decision support systems are designed as collaborative systems which complement the skills of the decision maker. Since the research reviewed is often work in progress, it is not always possible to say with certainty that a particular system will be able to improve decision making. It is however possible to assess the extent to which a design is based on a consideration of how information is used in decision making.

The above discussion can be summarised by listing three general criteria necessary for a successful decision support system:

- **VERACITY:** the information on which the decision support is based should be accurate and complete within understood limits.
- **PRACTICALITY:** the constraints imposed on a system by the medical domain and the clinical setting should be understood and the approach taken should work within these limits.
- **RELEVANCE:** the system should provide the user with information which is effective in improving his or her decision making.

Much of the work reviewed was published in technical journals without a detailed account of the intended clinical application. It would therefore not be appropriate to try and assess the extent to which systems meet a genuine clinical need.

In the next section four classes of decision support system are considered in turn and the extent to which research in these areas meets the criteria is assessed.

2.3 Computers and Medical Image Interpretation

In recent years an increasing number of articles, with origins in various strands of research in computer science, have appeared on computer aids for radiologists. The following sub-sections give a brief introduction to each of these topics and the research in them relevant to diagnostic radiology.

2.3.1 Image databases

Research into image databases began in the late 1970s when work in image interpretation created a need for systems able to store and retrieve large numbers of images. More recently researchers have been looking at image databases for use in multimedia applications and in what are called PACS (picture archiving and communication systems) for storing large numbers of clinical images. Grosky and Mehrota [Grosky1992] identify three generations of image databases. The first is implemented using a relational model with little or no image interpretation being performed by the system. The current generation is implemented using the object-oriented data model and includes some routines for detecting feature on images. It also addresses some issues associated with conventional databases, issues such as integrity constraints to prevent the entry of erroneous or inappropriate information. The third generation of systems is thought likely to include more sophisticated image interpretation facilities. Grosky and Mehrota highlight the importance of user-defined feature detectors.

The development of this technology has led to the possibility of using image databases to provide decision support: if unsure of the significance of an image feature, a radiologist could inspect database images with similar features but known pathology. This requires the storage of large numbers of images and the capacity to retrieve them quickly on the basis of their visual content.

Much of the research in the field of image databases generally involves attempts to develop “visual query languages” for this purpose, such as that of Chang [Chang1988] which allows “queries” to be built up from a reasonably small set of icons, a difficult technique to apply in clinical applications since it is likely that the user requiring assistance will have difficulty classifying the image feature. One group [Wiederhold1989] has developed a system for a database of 200 MRI images, in which the user indicates a section, or sections, of interest in the viewed image and this is used as the query in the retrieval of similar images. The designer of such a system faces two problems: first the process of matching the query with the images must be flexible enough to allow for normal variation between images of similar anatomy, secondly the process must be fast enough to be useful. In the Wiederholder et al. system an index consisting of binary images showing the outline of anatomical features is created from the database images, by noise reduction and simple thresholding. A set of graphical tools is provided to allow the user to specify image details, and relationships between details, to serve as a query. These details can be processed in a variety of ways to remove irrelevant data. The query is then matched with the binary images. The matching process is based on fuzzy mathematical morphology, using the query as a structuring element. In tests the system was able to retrieve similar images (those containing the same view) but the process took 40 seconds per database image. This is clearly too slow to be used interactively.

- **VERACITY:** the prototype contains 200 cases.
- **PRACTICALITY:** a great deal of thought has clearly been given to the design of the user interface and facilities for specifying a query. However, the matching process is very slow, and even given rapid advances in hardware remains a barrier to practical use.
- **RELEVANCE:** the study suggests that the retrieval technique is successful in identifying appropriate images, thus the system could provide the user with useful information.

Cohn et al. [Cohn1990] describe AXON, a knowledge-based system intended to assist in retrieving images from a database. In AXON each database image (chest radiograph) is stored with a set of keywords labelling the lesions and the disease. Associated with the database are frame hierarchies representing taxonomies of the keywords. For example, the disease hierarchy includes a frame for *tuberculosis* which inherits properties from *mycobacterial infection*, which in turn inherits properties from *infectious diseases*. The lowest level in the diseases hierarchy represents the cases known to the system. At this level there are links between the frames of different hierarchies, which represent the occurrences of lesions and connect the images with the appropriate cases. The simplest way to use the system is to enter a keyword and a condition. The system will match the keyword with its frame and then search down the hierarchy retrieving images that match the condition. For example, all images of neoplastic disease showing pulmonary lymphadenopathy could be retrieved by indicating that the system should search for cases in frames underneath the *neoplastic disease* frame and check for connections to the *pulmonary lymphadenopathy* frame before retrieving the images for the case.

This procedure has been made more sophisticated by using four “axes of clinical relevance”: radiographic findings, underlying aetiology, clinical findings and imaging modality. The stored images are grouped according to these four axes.

In addition, domain knowledge has been captured in search heuristics which have an application condition and a distance measure (an indication of similarity along an axis). One heuristic states that if fewer than five cases are found when searching along the aetiology axis, the immediate generalisation of the current aetiology should be used. When a user enters a query, the system first carries out a keyword search and gathers together all the heuristics whose application condition has been met. The heuristics are then applied in order of distance measure and the retrieved images are presented in order of merit.

AXON meets some of the criteria outlined in the introduction:

- **VERACITY:** the prototype contains only 60 cases and no mention is made of how an adequate coverage of image features and disease processes could be ensured.
- **PRACTICALITY:** the system is provided as part of an image display system in which both AXON and other kinds of information source are available on request.
- **RELEVANCE:** if the retrieval technique is successful the system would provide the user with information he or she could use in decision making.

The method for assisting in content based retrieval seems appropriate for a decision support system, in that it is based on an understanding of clinicians' ideas of relevance – although the paper does not explain where this understanding comes from. It is not clear, however, that the simple four-dimensional classification would be adequate to partition a search space much larger than 60 images, or that understanding what clinicians consider to be relevant could inform a sufficiently rich classification. The value of the heuristics is unclear. They are described as embodying knowledge about the domain, the domain in this case is as much “how to search a large database of medical images”, as it is “medical images” and the heuristics given don't suggest that there is much to be said about this domain.

The most complete account of an image database designed to provide decision support is given by Orphanoudakis [Orphanoudakis1994] who implemented a system in which automatic and user-guided image processing are used to derive representations of the content of images in the database. The automated image analysis produces “chains” of edge segments and the manual phase is used to edit the automated analysis and to define other regions of interest. Closed and open chains, (i.e. regions and

contours), together with characteristic features (e.g. area, position, perimeter) and the inclusion relationship are used as image descriptors. These image descriptors are created when images are added to the database and it is the descriptors rather than the images themselves which are the basis for retrieval. Rules embodying background knowledge are used to check the consistency of this representation. Retrieval strategies are also represented in the form of rules and are used, first, to reduce the complexity of a query description and, second, to control the invocation of query evaluation methods.

The retrieval is based on a two-phase process: inappropriate candidates are filtered out on the basis of gross characteristics of image segments and detailed matches are then sought for the midpoints and other characteristic features of polygonal approximations to the image segments making up the query. The first phase, medium-level matching, involves reducing the query to those elements which serve to discriminate between members of the class being searched, and then computing a similarity measure, a weighted sum of differences between specific attributes. The second phase, low-level matching, counts the number of “similar” linear segments in the database.

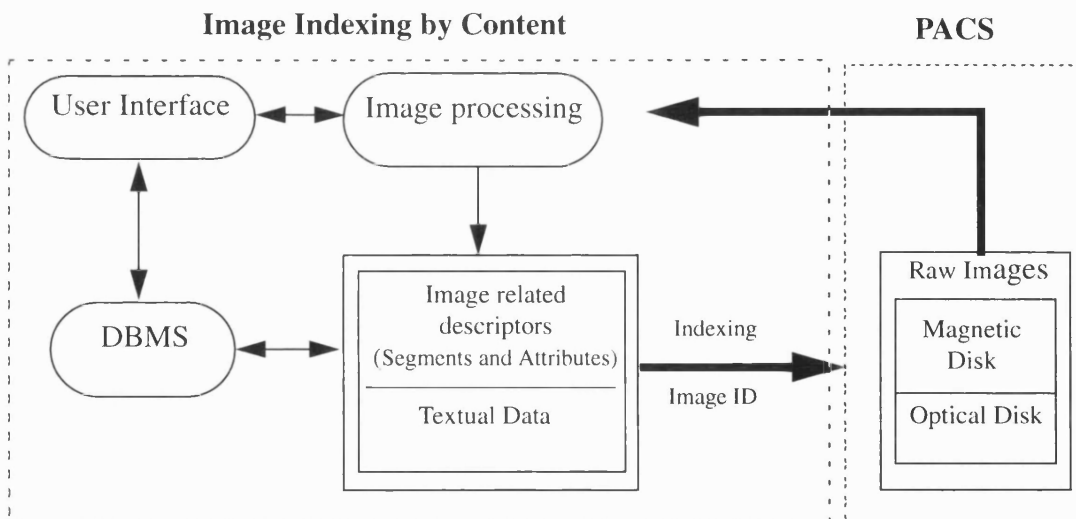


Figure 7: Image Indexing by Content, a block diagram of the Image Database System, adapted from [Kofakis1992]

- **VERACITY:** in common with the other systems described here, no mention is made of how the images included in the system are selected. There are, however, 400 images in the prototype system, rather more than in many of the others.
- **PRACTICALITY:** the paper gives a description of a user interface which could support different classes of user and different kinds of query. Users are able to narrow the search by specifying classes to be searched as well as by selecting image segments which are representative of image content.
- **RELEVANCE:** the system is presented as a tool allowing images to be retrieved on the basis of their content. It is argued that this could be useful for a variety of goals, including decision support. The authors, quite justifiably, are more concerned with explaining how retrieval by content could work than arguing for its value as a decision support tool. They record that experiments have demonstrated the efficacy of the retrieval mechanism.

The focus in all the papers reviewed in this section has been on the technical problems which must be solved before image databases can be used as decision support tools. Much of the work has been concerned with the retrieval of relevant images and other problems have received much less attention. One problem is that of ensuring that the image database itself is an adequate information source. There are conventional textbooks which provide atlases of radiological images e.g. Tabar and Dean [Tabar1983]. The analogy between such textbooks and image databases suggests that image databases are an appropriate way of providing decision support, but it isn't known how often clinicians consult such resources, nor in what circumstances, nor to what effect. Research into these questions could help inform the design of future image databases.

There are also textbooks which provide lists of all the possible diagnoses associated with every radiological finding for every class of image [Reeder1992], but

no attempt has been made to wed this to an image database. It seems clear that the proportion of images stored in digital form will continue to increase, and that the technology permitting their rapid transfer will become more efficient and more widely available. These developments will make image databases increasingly practical and accessible proposition, and will make the issue of their potential uses increasingly important. Much of this interest currently focuses on the use of image databases for teaching purposes [Sparacia1996], but the same technological developments – increases in speed, bandwidth and the availability of computer networks – have led to the availability of resources available which could be used in decision support tools.

2.3.2 Decision support systems based on numerical methods

Ledley and Lusted, writing about the reasoning that underlies medical decisions, characterised diagnosis as a logical activity of reasoning from symptoms to causes on the basis of “probabilistic” or “uncertain” information. They advocated the use of Bayes’ Rule to improve decision making [Ledley1959]. Their initiative led, eventually, to the development of computerised decision aids which use Bayes’ Rule to combine, for example, information about the frequency of a disease and the frequency with which certain signs are associated with that disease, to compute the probability that a patient showing these signs has the disease. A review of systems developed between 1971 and 1979 as aids to image interpretation, concluded that where the scope and range of the problem is narrowly defined, computer aids can perform better than humans [Lodwick1986]. The optimistic tone of the review is at odds with the impression that this was not a flourishing area of research, only eight systems were identified by the reviewer. Blois argues that such systems suit a collaborative model of diagnostic problem-solving since clinicians are much better than computers at identifying key symptoms and reducing the number of diagnoses to a manageable subset, while computers are better at performing the numerical calculations which identify the most likely option in that subset [Blois1980].

Good [Good1992] describes an experiment in which half the participating radiologists ignored the output of a numerical decision support system when rating mammograms for malignancy. This system was based on the descriptions that six radiologists gave of 429 sets of mammograms for which the clinical outcome was known. These descriptions were used as input to a classification program¹. An interactive program developed from the resulting classification system was then used by a different six radiologists to enter image descriptions and judgments of malignancy for 255 images. They were then given feedback based on the frequency of malignancy in similar images from the earlier set of images. Only three of the radiologists altered their judgments on the basis of the feedback, and their revisions were often in the wrong direction and in only a tiny fraction of cases would have led to differences in how the patients were managed.

- **VERACITY:** the information source in the system is the set of descriptions of images for which outcomes were known. Given that a panel of expert radiologists was used and that several hundred images were considered one might expect that this would be a full and accurate description of the domain. It is nevertheless possible that the resulting classification rule failed to capture information which could be of value to decision-makers.
- **PRACTICALITY:** the system requires fairly lengthy interactions to enter all the necessary data.
- **RELEVANCE:** Good notes that the lower the correlation between the users' initial certainty and the feedback, the less likely they were to make use of the feedback. This suggests that users were prepared to moderate their opinions given information which roughly agreed with their own perceptions but felt obliged to ignore information which differed wildly.

1. The author gives the name of the program used to derive the classification – 'ADALINE' – but no details of how it operated.

One of the best known attempts to use numerical methods in a decision aid for radiology is that of D'Orsi et al. [D'Orsi1992] which uses data derived from radiologists' interpretations of mammograms. The authors interviewed five expert readers to establish a set of 50 features indicating malignancy. A second set of features was identified experimentally. The experts were then asked to rate 24 images for similarity in an exhaustive series of pairwise comparisons, and a statistical technique, multi-dimensional scaling analysis, was used to derive a set of abstract dimensions that would explain the rating data. The two sets of features were then combined by the radiologists at a meeting where they identified the 23 most important features. 150 images were then rated for the presence of each feature by the same experts. This information was used to identify the 12 most significant features and their predictive values. These predictive values formed the basis for the decision aid.

The decision aid consisted of two components: a set of 12 scales on which radiologists could indicate their degree of certainty about each feature's presence and a computer program which combined the radiologists' ratings with the established predictive values to produce a probabilistic estimate of malignancy. The system was tested by six other radiologists, all of whom were experienced but were not specialists in mammography. They were asked to rate the probability of each mammogram being malignant and the pooled data was used to construct ROC curves¹, which showed that the decision aid increased accuracy. At a chosen threshold the decline in false negatives was found to be significant. ROC curves drawn up from data using only the 51 most difficult cases (difficulty being measured by the diversity of readers' judgments) showed a clearer improvement, both in false positive and false negative rates.

The same authors [Seltzer1992] used this methodology to develop a system for helping radiologists to combine the evidence from different imaging modalities. Feature lists and computerised decision aids were developed for breast X-rays and

1. See p254 for an explanation of ROC (Receiver Operating Characteristic) curves.

breast diaphanography. Tests of the decision aid showed no great improvement in the case of mammography, but considerable improvement in the case of diaphanography and significant improvement when used in decisions based on inspection of the two images. The authors claim that their tool allows users to merge information from the two tests at the “feature level” whereas radiologists inspecting images from different modalities would usually make separate assessments and only combine these gross judgments. Although the authors demonstrate the aid improved decision making based on the combined assessments, they do not show that these assessments were better than those based on the more effective of the two tests.

- **VERACITY:** the interviews and rating procedures impose a degree of rigour on the process of constructing the knowledge base, and should ensure that the most pertinent information is included.
- **PRACTICALITY:** nothing is said about the setting in which these aids might be used or how an interface to them would function. It seems likely that rating each image for 12 properties would be considered a heavy additional burden by many radiologists.
- **RELEVANCE:** the ROC tests, although on a small sample, demonstrate that some improvement is obtained through using these systems.

The methodology employed by D’Orsi et al. in the creation of their information source seems rigorous, but it is still based on radiologists’ judgments and not, for example, statistical data about the results of biopsies associated with mammograms. The probabilities are ‘subjective’ rather than ‘objective’. A criticism often levelled at this research is that the data on which it ought to be based is often not available, and if it is available, it may not apply in different geographical locations or at a later date. This is especially true in radiology where the incidence and value of radiological findings may be substantially altered by changes in imaging technology.

Kuhns et al. [Kuhns1989] have compiled tables of incidence and mortality rates for radiologically diagnosed diseases, together with information about the reliability of diagnostic tests and the risks and survival rates associated with treatments. Their aim is to allow a rough calculation to be made, of how certain one should be of a diagnosis before ordering treatment and of the impact a positive test result would have on that certainty. This provides an answer to the question: "Should I order this investigation?" The aim is to cut the number of investigations by helping radiologists identify cases in which the diagnostic certainty is already strong enough to warrant treatment and cases in which even a positive test result would not provide sufficient certainty.

Formal models of this kind have gained greater acceptance in relation to this category of decision than in application to decisions such as diagnosis. [Chang1990] uses this kind of formal 'decision analysis' to determine whether any of the currently available investigations could be applied to screening for prostate cancer. The screening application is an interesting one in this context since the challenge for the radiologist is as much setting an appropriate decision threshold as it is detecting and interpreting radiological findings [Dilhuydy1994].

It should be noted that numerical aids have been around for a long time in areas of image processing other than image interpretation, but that they have not been adopted with enthusiasm by the medical community. There are many possible reasons for this (see [Heathfield1993] for a discussion), some of them to do with social and political problems, others with the general difficulties of technology transfer. De Dombal et al. [De Dombal1992] discuss Bayesian decision systems and argue that one of the principal problems is ensuring that the physician is able to assess accurately the significance of the computer's recommendation. They advocate presenting overview data describing what happened to patients who received similar assessments, rather than making diagnostic predictions. There are advantages in using computers in decision making: computers vary less than people and can deal with a much greater volume of information. However, where the decision-making process has to remain

under human control, computers must provide appropriate input. It would have been interesting to know to what extent the improved performance found by D'Orsi et al. was due to the checklist of significant features, and to what extent it was the calculation of probability that improved performance.

A major weakness of the application of this technology to problems of diagnostic radiology is that all that such systems provide is an assessment of the significance of the user's interpretation. This kind of system is therefore dependent on the user being able to enter a sufficiently rich description of the finding couched in terms which correspond to the features identified in the information source. This leads to practical difficulties and to problems to do with the accuracy of information.

2.3.3 Expert systems

Expert systems are similar to numerical decision aids, the distinction being that in the numerical decision support systems described above the information source is made up of statistical data derived from past cases or radiologists' judgments, while in expert systems it is a symbolic representation of, in this context, clinical expertise. Many systems have been developed with the aim of providing diagnostic advice on the basis of information entered by the radiologist about what he or she sees on the image. Published accounts of such systems fall into two categories:

- small-scale systems
- novel developments of expert system ideas

Small scale systems are generally developed using a commercial expert system shell which provides the interface and inference mechanism. The system developer then works with a radiologist to construct a set of rules which provides the knowledge base or information source. Cook [Cook1989] developed such a system to distinguish between benign and malignant mammographic anomalies. The system contained a

number of rules (the number is not given) distinguishing 16 common mammographic manifestations. The radiologist assigned 'certainty values' to different conclusions. A certainty value of 0 eliminates a candidate, a value of 10 establishes it as a definite conclusion. Where an abnormality was proposed in the conclusion of more than one rule with differing certainty values, the average value was taken.

- **VERACITY:** the knowledge base was drawn up in part from information supplied in textbooks and in part supplied by an expert radiologist. The authors state that the most difficult aspect of this task was estimating the degree of certainty to be associated with the rules and that was left largely to the judgment of the radiologist.
- **PRACTICALITY:** no information is given about the user interface. However, the system was developed using a commercial package and took the form of a stand-alone system. A busy clinician who does not make continual use of a computer is unlikely to consult such a system other than in exceptional circumstances.
- **RELEVANCE:** in tests of the system it was found that the diagnostic accuracy of radiologists was not significantly improved, but that of residents improved dramatically so that it approached that of the radiologists. Even students of biomedical engineering, unable to interpret images on their own, performed with the same kind of accuracy as the radiologists when using the tool. The problem with such systems is that they are of little value to a specialist and yet too specialised to merit the attention of a generalist. They may be of some value to students of a speciality but suffer through being designed as decision rather than teaching aids.

Piraino [Piraino1989] developed a system for differential diagnosis using radiographic images of focal bony lesions. Users inspect prints of radiological images and associate them with appropriate radiological findings. The knowledge base links these findings to diagnoses, the link is labelled with a predictive value and a relative

frequency. These values are derived from elicited knowledge and textbooks. The input findings are first used to generate candidate diagnoses and the system then checks to see if any candidates are associated with findings that weren't input or that are not associated with candidates. The candidates are then ranked on the basis of calculated relative likelihoods. The system presented a list of candidates whose likelihood was greater than 50% of the most likely candidate.

- **VERACITY:** the knowledge base relates radiographic findings to diagnoses and gives predictive values and relative frequencies. The information was obtained from expert radiologists after reviewing standard textbooks.
- **PRACTICALITY:** the system suffers from the same constraints as that of Cook. The interaction is driven entirely by the user who is free to suggest findings. This fits the model of decision support as a means of assisting the user's decision making.
- **RELEVANCE:** when used by 268 people of varying experience the correct diagnosis was selected as the most likely candidate in only 12% of cases; in 78% the correct diagnosis did not feature at all. However, when the system was used by musculoskeletal radiologists, the correct diagnosis appeared on the generated list 71% of the time. Allowing passers-by to play with the system could perhaps be dismissed as an optimistic and unscientific test, but this is one of the few published studies to give a negative evaluation and stands as a useful corrective to many papers which report apparent success with tests carried out in highly controlled situations. One lesson which might be learnt is that the terminology used to build knowledge bases is often specific to a small community of potential users.

More interesting are the systems developed with the intention of extending conventional expert system technology in order to tackle the problems which beset the less ambitious systems. Visual Heuristics [Bonadonna1992] was developed to aid diagnosis based on descriptions of thorax radiographs. Visual Heuristics appears to be

the result of more thorough research than most of the systems reviewed so far in this section. The design is described as being based on experimental work aimed at understanding the processes involved in the perception of radiological images. In fact the design is simple and little is said about how the theoretical background served to shape it. The object-oriented knowledge base contains a representation of the thorax, including descriptions of its appearance in medical images and details of pathologies.

The knowledge base also contains structures for storing patient data, including visual schemata, scenarios and diagnostic cases as well as rules which relate cause and effect to explain the various pathologies and their characteristic presentations in images. The authors state that since anatomical objects in the system have a visual and a logical representation the system could be employed to generate a 'dynamic atlas' for use in teaching and consultation. How this might be achieved is not clear. Currently the system can be used in two ways: it can generate tests for the hypothesis that a pathology is present or can suggest a diagnosis on the basis of input patient data.

- **VERACITY:** the description of the thorax is detailed and well-structured. The description of the rules relating this description to diagnostic conclusions is, however, very sketchy.
- **PRACTICALITY:** the user interacts with the system to build up a description of the analysed scene, based on the model of the thorax.
- **RELEVANCE:** the operation of the system is an attempt to simulate the processes that take place when a radiologist interprets an image.

The systems reviewed so far in this section have not approached the problem of providing decision support as one of implementing a system intended to facilitate for co-operative problem-solving. One system designed with this in mind applies Miller's work on critiquing systems in a system for the differential diagnosis of lung disease from chest radiographs [Swett1987]. The knowledge base contains the usual production rules and frames and also a text generator. The user supplies clinical data,

findings and a proposed diagnosis. The production rules are then used to make further inferences from the supplied data and these, together with the supplied data, are processed using “expressive frames” to provide the content for the generated prose critique. The details of how this happens are not supplied in this paper. A later paper [Swett1993] describes a version of the system adapted for mammography and integrated with a version of the AXON system described in Section 2.3.1. The radiologist inputs the clinical and radiological findings and a proposed diagnosis. The system then uses these as a basis for providing context-sensitive advice. This paper is also rather sparing with technical detail.

- **VERACITY:** ICON is described as an experimental system designed to test the feasibility of the approach. The knowledge base contained 79 rules, not thought to be a complete representation of the domain.
- **PRACTICALITY:** the approach clearly has potential advantages in terms of usability, since the interaction is driven by the user’s construction of a traditional radiological report.
- **RELEVANCE:** no details are given of an evaluation.

A project described by Keravnou et al. [Keravnou1993] is developing an ambitious expert system to assist in the diagnosis of skeletal dysplasia on the basis of information entered by radiologists. The system is based on a model which sees diagnostic reasoning as comprising three activities: the triggering, differentiation and evaluation of hypotheses. At the beginning of a consultation the user inspects the image and enters some findings into the system. If a finding is known to be strongly suggestive of a dysplasia, it is triggered and becomes an active hypothesis. The differentiation module selects the most plausible active hypotheses, divides them into clusters and then investigates each cluster in turn, a process which involves requesting information that can distinguish between hypotheses, and which may in turn trigger new hypotheses. Less plausible hypotheses can be added to clusters which contain promising hypotheses.

To evaluate the hypotheses the system compares the findings that were entered and the findings which would be expected given the hypothesis, in order to determine both how many of the entered findings are explained and how many of the expected findings have been entered. In addition to knowledge about dysplasias, the system contains two types of “background knowledge”: foundational knowledge related to diagnostic findings (e.g. “if a part of a component is abnormal then a component is abnormal”) and a temporal model which supports reasoning with time-related facts about the case and about dysplasias. No evaluation of the described model has yet been performed although a limited evaluation of an earlier prototype showed a marked improvement over “manual diagnosis”.

- **VERACITY:** the bulk of the effort in the project has been devoted to developing a well-designed representation of the domain and also of some less formalised “background knowledge”.
- **PRACTICALITY:** published reports give no account of the interface or how it might meet the demands of the clinical setting.
- **RELEVANCE:** the question of how the information supplied by the system might be incorporated into the user’s decision-making process has not been explicitly addressed. The system is described as including a model of diagnostic reasoning, and it may be that this is assumed to guarantee the relevance of the information provided, although this would require that the appropriateness of the model is apparent in the presentation of information.

In recent years expert systems have been applied in a number of fields. They have tended to be most successful in clearly delimited domains where a large body of formalised knowledge exists and most useful where there is an identifiable class of users with a frequent need to consult this information. Expert systems for image interpretation have not yet been developed successfully because there is no real demand for the simpler systems and a number of research problems need to be solved before more

sophisticated systems will become useful. Users need to be able rapidly to enter unambiguous descriptions of images and receive succinct descriptions of the likely diagnoses. This requires sophisticated interfaces for user input, novel techniques for knowledge representation and inference mechanisms which are designed as collaborative problem solvers.

2.3.4 Image-processing systems

The term 'image processing' can be applied to any mathematical transformation of a pixel array. The term covers the numerical treatment of images for quantitative measurements, image enhancement, object recognition, segmentation, 3D reconstruction, tomography. Work in all of these areas could be considered relevant to decision support, however very little of it was explicitly intended to be. This section concentrates on the detection and classification of imaged objects, since image processing with these goals has been used to provide clinicians with decision aids. In contrast, tools for the segmentation of medical images are generally viewed as valuable for facilitating image measurements.

Perhaps the simplest object recognition technique is to set a threshold on the luminance values in the image, detecting objects which give rise to pixel values of a distinctive brightness. In practice this technique can be used only where some additional subtlety is employed in setting the thresholds, for example in the technique developed by Parker et al. [Parker1994] which detects microcalcifications in mammograms on the basis of significant peaks in the distribution of grey levels.

In thresholding systems objects are associated with a range of values on the single dimension of pixel value, objects can also be matched with regions of a multidimensional space, in which the dimensions correspond to photometry-based measures such as area or mean brightness. Object recognition is then performed by identifying

the region of interest on the image, computing the feature vector or the values of the various photometry-based measures for that region and classifying it accordingly. Systems of this type, let us call them 'feature vector classification systems', are often developed by feeding the photometry-based measures for regions into software which is able to combine them with varying weights until a combination is found which produces a reliable classification. The classification software could be similar to that used in Bayesian decision systems. Kegelmeyer [Kegelmeyer1994] for example, used 46 features of regions identified as microcalcifications as input to a 'Binary Decision Tree' in order to generate a classifier which would filter out false positives.

The simplest model-based systems use models which can be matched directly with the image. Matsumoto [Matsumoto1992] describes a system in which an idealised image of a projected 9mm lung nodule is used as the pattern in a system for the detection of lung nodules. Since similar objects can give rise to widely different image patterns some flexibility must be built into the process. In model-based object recognition, the flexibility is generally built into the function which attempts to match the model to the image pattern, although some systems build it into the model. The standard example of a flexible matching function is the Generalised Hough Transform. This is used to match templates to image data in cases where the templates may be described using a set of parameters. This method has been used to detect stellate lesions [Astley1993] in mammograms.

The kinds of flexible model used in object recognition include those implicit in neural networks. Tourassi and Floyd [Tourassi1993] used neural nets to detect cold lesions in simulated noisy SPECT data. A network with 256 input nodes, eight hidden nodes and one output node was trained on neighbourhoods of 16 by 16 pixels; 220 patterns were used in training the net: 80 lesion free neighbourhoods and ten each of seven different sizes of lesions were viewed at two different counts levels.

The network was tested on 64 by 64 pixel images, again of simulated SPECT data. The 16 by 16 pixel input region of the neural net was moved over the images, pixel by pixel, and the response at each point was recorded in an image of the neural net's output. The image was filtered using a noise reduction technique and then regions of the output image where ten or more adjacent positive responses coalesced were considered as representing detected lesions. Tested on 600 images, the network has 100% sensitivity on larger lesions while 80% sensitivity was found at a threshold of 50% specificity for 1cm lesions.

In addition to object recognition tasks like this, neural networks have been used to segment images on the basis of texture, e.g. to classify abnormalities on the basis of a user's input feature descriptions [Goldberg1992, Kippenham1992]. Shen et al. [Shen1993a] describe a neural network for classifying abnormalities on the basis of computer derived feature descriptions. Pattern recognition techniques are used to identify potential microcalcifications and then three shape parameters based on Fourier descriptors, moments and compactness are calculated for each microcalcification and used as inputs into a three-layer neural net. The net is trained with back-propagation to identify malignant calcifications. Tested on four images containing a total of 52 benign and 241 malignant calcifications, the system detected 85% of the abnormalities with a total of 29 false positives. Of the malignant microcalcifications, 87% were correctly classified.

Much of the work reported here attempts to improve techniques that may become useful in the future, rather than to demonstrate their value today. Even techniques that perform better than human observers on a particular detection task will not be useful if they are only slightly better than humans, if their operation is poorly understood or if they are attempting to answer only one of many questions involved in the interpretation of an image. Until recently few papers published on object recognition techniques in radiology described how the system could be used to support a clinical user. One paper which does is that of Astley et al. [Astley1993] in which the

authors propose that research be focused on areas in which human interpreters need assistance and stress that we need to understand how well these techniques work and how they can be used effectively.

Astley et al. have concentrated their work on the problems and challenges created by the UK's national breast screening programme. Analysis of the errors made by radiologists can be used to provide evidence about the areas in which assistance is required. Techniques for detecting the different classes of lesions are described, but the authors accept that these are inadequate either as a basis for an automated sort prior to screening by radiologists or to serve as a second reader. The authors believe that existing object recognition systems may be most valuable in providing prompts to guide the radiologists' search, and they report research which attempts to establish how they can be used in this way; studying the effects of prompting on radiologists' ability to detect microcalcifications [Hutt1994].

- **VERACITY:** the information source in this case is the image-processing algorithm, which was adequate as a detector of microcalcifications but ignored other abnormalities, hence the VERACITY criterion is only partially met.
- **PRACTICALITY:** the system is designed for a specific setting and focused on areas where interpreters are known to need assistance.
- **RELEVANCE:** detection was significantly better with prompts even with a moderate false-positive rate, hence the information provided was clearly of value.

Even when more effective techniques have been developed they will remain unused in clinical practice unless more thought goes into ways in which object recognition systems can be employed to improve the performance of human radiologists. The challenge is to design a framework based both on an acceptance of the limitations of the image-processing technology and an awareness of the needs of clinicians. Many

important questions remain unanswered. These questions are closely related to the criteria drawn up in the introduction to this review. Mammography is identified as an area where there is a need for support and the constraints which screening puts on how that is provided are understood, meeting our first two criteria. Meeting the VERACITY criterion would require detectors for a useful range of abnormalities. It is not clear if this has been met by the research reviewed here, since only some abnormalities can be reliably detected. Gale et al. [Gale1993b] argue that strong individual differences in error rates on certain types of abnormality mean that this is enough to be useful. However, it seems unlikely that software which attempts to detect only one kind of abnormality will be taken up. Nor is it clear that favourable results found when prompting for one abnormality will transfer to prompting for all abnormalities. Little is known about how prompting improves performance or how its capacity to do this is affected by parameters such as the false positive rate of the detection software.

2.4 Conclusion

Two challenges face the development of computer aids for image interpretation. The human visual system's capacity to detect structure is still, despite decades of research into computer vision, much more powerful, flexible and successful than any technological alternative. The design of a decision aid must therefore take into account the strength of human perception. Our capacity to design tools to augment, rather than to replace, human interpretative skills is, however, limited by our understanding of how the human perceptual system works and a decision aid must work within this constraint.

The research reviewed here meets these two challenges with varying degrees of success. The idea of image databases for decision support is an appealing one since the human visual system can effectively extract information from images. If a computer is to be used to provide decision support using image databases, a system must be devised

which allows the user to retrieve an image on the basis of its content. Progress is being made both on ways of analysing images in order to identify their components [Wiederhold1989] and methods of matching the segments with reference images. Such systems will, however, only be practical where a user's questions are answered by the presentation of a reference image. The most appropriate technique for the representation of the user's question isn't clear. In [Wiederhold1989] an element of the image is used as the query, in [Cohn1990] a symbol is matched with terms in a knowledge base, but it is likely that both forms of information will have to be used if a range of queries are to be supported effectively.

Numerical methods and expert systems can contain a broader range of information about the significance of findings and their relationship to medical conditions. As the computerised collection of medical data develops, the potential for systems that make numerical data available to help in decision making will increase, and so will the need for research into the most effective ways of presenting statistical information. The difficulty with applying expert systems to this domain seems to be that, although it is relatively easy to represent facts such as those found in textbooks about the relation of findings to diseases, this is only a small part of the knowledge that is acquired in radiological training. The challenge is to develop a system which is able to handle a domain complex enough for it to be useful. The development of more sophisticated systems in this domain is hampered by the difficulty of capturing, in the symbolic form which these systems require, the capacity to recognise and discriminate between features. One of the problems is that of constructing a knowledge base that allows users adequately to describe the features of visual images, another is developing a style of interaction which ensures that the expert system is complementing the user's expertise. This suggests that there is a need for image processing which can automatically extract information about images.

The use of image processing is problematic because the human visual system is still better than computers at distinguishing between normal and abnormal structure in

a previously unseen image. There is, however, increasing evidence suggesting that these techniques can be applied in certain cases and in these instances do improve radiologist's decision making. The trick here is to detect certain classes of abnormality on images which humans find difficult. Even then, current algorithms are sufficiently sensitive to be useful only when a high false positive rate is permitted. It has been shown that, even if there are two false positives per image, radiologists' detection rates do improve if they use the positive responses from an abnormality detector as prompts. However, if a battery of detectors is to be provided for images in which a range of abnormalities could be detected, the false positive rates must be kept much lower if the total number of prompts is to be held at an acceptable level. This suggests that image processing must be provided within a framework which ensures that the information being presented – whether as prompts or in some other form – is appropriate to the particular decision being taken. One possible hypothesis, explored in this thesis, is that a knowledge-based decision support tool could provide an appropriate framework for making available information from image processing.

This review has highlighted a number of problems in the development of decision aids for image interpretation. First, the choice of domain. There are many areas in which either the images are difficult to interpret, or the expertise required is in short supply, or the number of images generated stretches the available resources: areas, in other words in which computer aids could be useful. The problem is that these are not necessarily the areas for which it is easiest to develop a decision aid. In a number of cases it seems that research has been driven more by what is possible than by what is desirable. Second, there are the problems relating to how the system is used. A successful system must be designed for the setting in which it is to be used. It is less obvious what the standard of usability should be for a research prototype, especially when one considers the pace at which digital technology is changing the nature of radiology. Three points very quickly become clear from an observation of the working practices of clinicians, and are repeatedly made in discussions of clinical decision support systems [Shortliffe1991, Greenes1991]:

- a system which requires any lengthy interaction is only going to be used as a measure of last resort
- any system which is integrated into the normal routine of the clinician and which performs some clerical tasks is more likely to gain acceptance
- clinicians will be unwilling to learn how to use many different decision support systems

Few medical decision support systems, and none of the early systems, have been developed as collaborative problem-solvers. Their makers assumed that it would be possible to develop a system that performed demonstrably better than a human. Such a system could then be installed and would be consulted by the decision-maker when he or she required special assistance. Few of these systems entered widespread use. One reason for this failure is perhaps that the designers were more concerned with ensuring that their systems out-performed clinicians than they were with improving the performance of clinicians. This leads to weaknesses because it means that the systems haven't built on the users' own expertise. Perhaps more importantly, it leads to a failure of user acceptance, since it is not clear how the user should behave when he or she feels that the system is in error. The responsibility for the decision lies with the clinician and even if the machine is more reliable than the clinician, it is certainly not infallible.

Computer technology, of the kind reviewed here, may be useful both as a mechanism for independently identifying relevant image features and in making information available to help in decision making. This review has surveyed decision support systems based on different kinds of computerised information source. No one kind of knowledge is likely to prove pre-eminent and yet a proliferation of systems would mean redundancy and be confusing and daunting for users. There is therefore a strong argument for attempting to develop an integrated decision aid which is able to provide information of different kinds on request, or in response to different problems. The next chapter explores the possibility of developing such a decision aid.

CHAPTER THREE

THE SIGNAL/SYMBOL PROBLEM

If pictures represent, they do not represent in the same way as descriptions. Yet, as manifest as this difference is, it is far from obvious how to explain it.

Dominic Lopes

3.1 Introduction

Chapter Two reviewed research into decision support systems and suggested that there may be a need for systems which allow different sources of information to be used to their best advantage. An aim of this thesis is to identify the technical problems which must be solved for such systems to be built and to propose principled solutions to these problems. The four kinds of information employed in the decision support systems identified in Chapter Two fall into two categories: signal information, such as that contained in images and manipulated in image processing, and symbol information as employed by knowledge-based systems and Bayesian decision aids. In this chapter, the literature on systems combining symbolic knowledge bases and signal data is reviewed, and discussed. Three ways in which signals are mapped to symbols are

identified (pattern matching, knowledge-guided access to an image database and image interpretation). In each of these there is a stored association between elements of the symbolic representation and the results of image-processing operations. Different approaches to the representation of these results are identified and compared. A suitable approach is proposed for the applications to be considered in this thesis.

3.2 Signal and Symbol Information

The distinguishing characteristic of this work is a concern with systems in which decision support is based on information from more than one computerised source. Consider briefly the potential for such a system in mammography. A radiologist is inspecting a set of mammograms taken following the referral of a patient from a screening programme. A cluster of microcalcifications is clearly visible on the images and the radiologist must decide whether their form and distribution are characteristic of cancer or of one of the benign processes which result in calcification. A computerised decision aid could provide the radiologist with access to a knowledge base representing the best available expertise on the subject, could aid quantitative measurements of the image and enable comparisons with a database of images with known pathology. Any decision on the classification of microcalcifications will be followed by a decision on the subsequent course of action. The decision aid could include a representation of the hospital guidelines or protocol to be followed in such cases together with a record of the procedures followed in the handling of this particular patient.

To make a correct decision about whether or not a patient should be referred for a biopsy, the radiologist must weigh up a number of different factors: the patient's age, her personal and family history of disease, facts about the incidence of the disease and the range and consequences of the available investigations and treatments. This information can be represented on a computer in a symbolic knowledge base. The portions of the mammograms that reveal the microcalcifications, however, would be represented as

signal information in a digitised image. Any decision to be made about the management of the patient must involve a consideration of both forms of information; similarly, the design of a computer aid for this sort of decision making must also be based on both.

3.2.1 The signal representation of information

Wittgenstein wrote, in reference to philosophical problems in understanding human perception, that “we find certain things about seeing puzzling because we don’t find the whole business of seeing puzzling enough” [Wittgenstein1958] and it is true that one of the difficulties in understanding how information is represented in images is that information which seems to us simply to be manifestly present in the image is only available to us as a result of a complex sequence of operations which it is not yet possible to reproduce as a computer program.

A medical image is created from the pattern of radiation falling on a detector after it has passed through, or been emitted from, the human body. In the case of the mammogram mentioned above, a beam of X-rays passes through the breast to fall on a film/screen combination. The brightness at a point in the resulting image is a function of a number of parameters: the characteristics of the beam, the exposure time, the characteristics of the film/screen combination, the thickness of the breast when compressed in the mammography machine and the relative proportions of the different tissue types in the narrow column through the breast above that point in the image. From the viewpoint of the radiologist, it is only by virtue of the last parameter that the image contains interesting information and can assist in the detection, characterisation and localisation of anatomy. The perfect image would be one where each pixel was a measure only of the relative proportions of the different tissue types in the column through the breast above that pixel [Highnam1994].

A radiologist might be able to see from a mammogram that there is evidence of calcification in the breast. This information would be present in the image because calcium absorbs a much higher proportion of incident X-rays than normal tissue, resulting in decreased exposure and increased brightness at a corresponding point in the film. However, other regions of the image will have a similar brightness to that corresponding to the calcification, because of a chance arrangement of normal tissue or the scattering of radiation as it passes through the breast. The majority of these regions do not confuse or mislead radiologists because a radiologist's perception is not simply a response to the brightness in the image. The eye is responsive to discontinuities in brightness rather than to absolute illumination, and the perceptual system constructs an interpretation which is based not only on the patterns of stimulation on the retina, but also on expectations about what these patterns correspond to [Pizer1990]. The information cannot simply be "read off" the luminance values.

The representation of information in images has the following characteristics. First, the precise values of the data – in this case the brightness at different points in the image – are not a simple, direct function of the imaged anatomy. Secondly, salient information – about significant features, evidence of disease and so on – may only be a small proportion of the total information in the image. Thirdly, this information is implicit in the spatial arrangement of the data. This last point must be particularly stressed: the human visual system is a powerful instrument for analysing images and it is easy to forget that obtaining information from images requires complex computations and, even then, may only be possible on the basis of background knowledge.

3.2.2 The symbolic representation of information

Just as it is easy for us to ignore questions of how information is represented in images, it is easy to get confused about what it means for information to be represented in computer programs. We commonly talk about facts or information being stored in a computer and explain how programs work by making statements about what the

computer 'knows'. It is worth thinking a little more carefully about how a computer program can serve as a representation.

A decision support system must contain a representation of a particular domain. In a knowledge-based system this representation will consist of a set of symbols, a mapping between those symbols and the relevant properties of the chosen domain and a set of rules for manipulating the symbols in ways which produce accurate results. The symbols and the rules for manipulating them are realised in the computer system; the mapping between the symbols and the domain is not. The mapping is created by the programmer when he or she chooses this or that symbol to represent this or that aspect of the domain. This is worth remembering since the success of the system depends on this mapping being understood by the user. A computerised decision support tool is, fundamentally, a tool for communicating the intentions of the knowledge base author to the eventual user. Thus if I use the symbol 'microcalcifications' to represent microcalcifications when writing the knowledge base, I must be able to communicate to the user that I am using it in this way. The difficulty, in practice, is not in conveying the explicit association between the atom and the idea it represents, but in avoiding the usual connotations of the word, associations which have necessarily been neglected in building a restricted formal model of the domain [Winograd1986].

It should be stressed that the opposition of symbol to signal representation is not equivalent to that between qualitative and quantitative information, and we would expect some of the properties in the symbol representation to take numerical values.

The set of symbols used to denote the object classes and their interrelationships must be, to some degree, pre-defined. Systems which learn new rules or which can generate new abstractions do exist, but the basis for this learning is pre-determined. In a system to assist in the interpretation of images, there must be a representation of everything pertinent that might be present in the signal – or the building blocks from

which one could be constructed – before analysis of the signal starts. In practice this places clear limits on the kind of system which can be developed and the kinds of application which can be tackled.

3.3 Systems Combining Signal and Symbol Information

The key points of the above discussion are that signal data conveys information through the arrangement of values for some physical parameter (for example, through a spatial arrangement of values corresponding to the attenuation of X-rays in tissue) while a symbolic representation conveys information through the association of concepts and ideas with an abstract model composed of a set of tokens or symbols.

This section reviews examples of the different ways in which signal and symbolic information can be combined. Since much of the work is intended to contribute to the analysis of medical images but not to provide decision support for the interpretation of medical images, the review is complementary to that in the previous chapter, although some systems are mentioned in both. The systems described are grouped into four categories:

- multimedia systems
- expert systems for image processing
- expert systems with image processing
- knowledge-based image segmentation systems

For each system I briefly describe the signal information in the system – that is the image data and the data generated by image-processing operations – the symbolic knowledge, some details of the operation of the system and of how the signal data and symbolic information are combined.

3.3.1 Multimedia systems

The simplest way of combining signal and symbolic information is to make them both available within a single system which gives the user access to both but which makes no attempt to derive symbolic information from signal data. Examples of such system are image databases with knowledge-guided retrieval tools and decision support systems which also provide access to images and image processing.

3.3.1.1 AXON: a knowledge-based image retrieval system

This system [Cohn1990] is a knowledge-guided image retrieval system. The idea is to allow a database of images to be used as a decision support tool by facilitating the retrieval of relevant images. A knowledge base of information about the domain and a set of search heuristics is used to identify a subset of relevant images in response to a query from the user.

Images and image processing

The system works with digitised chest X-rays. There is no image processing.

Symbolic knowledge

A hierarchy of frames is used to represent the necessary information about the medical domain. The most specific level in the hierarchy, the leaf nodes, represent specific instances, that is to say the individual cases and actual medical images included in the database. A set of search heuristics completes the symbolic knowledge.

Operation

Processing is initiated when the user enters a query. Keyword matching is performed to identify a set of relevant images. Matches are sought by searching along

four “axes of clinical relevance” which map onto the frame hierarchies. Search heuristics are used to control the widening of the search if insufficient images are identified. There is no processing of images, the aim is simply to guide the retrieval of images.

Combining symbols and signal data

Terms in the symbolic knowledge base are linked to images in the image database. Thus terms in the knowledge base are understood in part through association with a set of images.

3.3.1.2 I²C: a database for retrieving images on the basis of visual content

Orphanoudakis et al. [Orphanoudakis1994] describe a system in which automatic and user-guided image processing are used to derive representations of image content, representations which can be used to retrieve images from the database.

Images and image processing

The system is designed to work with different classes of image and with different approaches to processing images. The published papers refer to a database of brain MRI images and use edge detection methods to derive a description of imaged objects in terms of contours, edge segments and points.

Symbolic knowledge

The representation of the symbolic knowledge is based on the notion of classes. Images are organised into classes on the basis of primary criteria – such as modality, the part of the anatomy, orientation, plane of cut – and secondary criteria which may be derived from the clinical interpretation of the image or assigned by a machine learning algorithm. The system maintains an encapsulated object for each class, which stores general information about each image type and about image description algorithms.

Each image in the database is also associated with the description obtained by applying the relevant image-processing operations. These stored descriptions are called ‘logical images’.

Operation

When an image is added to the database the default description type builds a logical image. In the implemented prototype the description is a hierarchical structure at each level of which attributes relating to the statistical and geometric properties of contours and segments are stored together with an estimate of their importance. Two forms of content-based query are permitted: image queries and sketch queries. In both cases inappropriate candidates are filtered out on the basis of gross characteristics of image segments and then detailed matches are sought for the midpoints and other characteristic features of polygonal approximations to the image segments in the query.

Combining symbols and signal data

The system contains symbolic information about the clinical interpretation of an image, information which is stored in the class object associated with an image. Thus, as with AXON, a link between symbolic information and signal data is made in the database. I²C, however, also uses image processing to generate a ‘logical image’. This logical image is a symbolic description which is derived from signal data and which can be matched with similar descriptions derived from other signal data.

3.3.1.3 VIA-RAD: an interactive problem-solver

The system [Rogers1995], intended to assist radiologists interpreting medical images, is based on a psychological model of image interpretation which is used to control diagnostic reasoning and visual interpretation and also the application of relevant image enhancement operators.

Images and image processing

The system was developed to assist in the interpretation of digitised chest X-rays. It includes a limited number of image-processing operators which perform image enhancement tasks, one such operator being histogram equalization.

Symbolic knowledge

The symbolic knowledge in the system consists of four basic components describing the domain: landmarks, findings, features and diagnoses. The landmarks frame provides a representation of all the anatomical objects which are visible or are expected to be visible in the image, of the possible aggregations of landmarks and of their identifiable components. The findings are represented at three different levels of inference from the perceptual data: facts about the image (a lucency), radiological findings (a mass) and diagnoses (a tumour). Features are the concepts which are used to describe findings. The diagnoses frame represents the associations between findings and diagnoses and the information required to calculate the appropriate degree of belief for the hypothesised diagnoses. In addition to these components, the system contains knowledge about diagnostic strategies and image enhancement operations. The five diagnostic strategies represented are pursue, rule-in-rule-out, not-enough-information, conflict and deliberate-landmark-search. The knowledge about image enhancement operations is associated with the representation of image features.

Operation

The system is based on a blackboard on which are posted visual hypotheses about what is present in an image and reasoning hypotheses about explanations for objects and collections of objects. The user is viewed as an additional information source, reading from and writing to the blackboard in an attempt to solve the problem. The processing is controlled by three modules, all identified in highly anthropomorphic terminology: the 'hypothesis manager', the 'attention director' and the 'strategy selector'. The hypothesis manager controls the hypothesis-related information sources and presumably functions as the inference engine although this is unclear in the

account. The attention director alters the image and posts information about what to look for. The strategy selector decides how attention should be focused, making judgments about the relative merits of the suggestions from the attention director.

Combining symbols and signal data

The representation of features includes information about image enhancement techniques, which may be posted to the blackboard. The attention direction module can invoke these operators to alter the appearance of the image. So symbolic terms are associated with image-processing operators but not with signal data.

3.3.2 Expert systems for image processing

The successful application of image processing requires an understanding of the image-processing operations, of the class of images to be processed and of how the imaging process works. Nazif and Levine [Nazif1984] were the first to propose that developers attempt to represent this knowledge explicitly in the software. There are now several kinds of system which may be considered as expert systems for image processing and analysis (ESIPAs). Matsuyama [Matsuyama1993], reviewing this work proposes a general architecture for ESIPAs comprising:

- a user interface
- a library of image-processing operators
- knowledge about image-processing techniques
- database of characteristics of image data
- image processing and analysis executor
- reasoning engine

He considers four kinds of ESIPA:

- a consultation system for image processing
- knowledge-based program composition system
- rule-based system for image segmentation
- goal-directed top-down image segmentation system

The last two categories are dealt with in later sections of this chapter. There is at least one example of a consultation system for medical image analysis [Rosis1992], which provides an intelligent interface to image-processing operators. A more developed example is of a knowledge-based program composition system, as described by [Kulikowski1995].

3.3.2.1 MICRORAY-1: a consultation system for medical image analysis

De Rosis et al. [Rosis1992] describe a system aimed at guiding a doctor using image analysis techniques in the interpretation of a medical image.

Images and image processing

The system is designed to be independent of any particular imaging modality or set of image-processing algorithms. The published paper includes examples of image-processing routines such as linear contrast enhancement used in the analysis of CT images of cancer in the eyelid.

Symbolic knowledge

The system contains application-independent knowledge about various goals of image analysis, the methods for achieving them and about the image analysis software functions supported by the system at any one time. This information is represented in a network of frames. Information acquired during the consultation about the particular problem the system is being used to address is represented in frames which describe

both the image and the anatomy it depicts. The system also contains a simple set of rules which implement an image analysis strategy, in essence that goals are achieved by applying appropriate operators. The strategy involves testing that the application conditions for an operator have been met and, once it has been applied, checking that the goal has been achieved.

Operation

The system is used by a clinician who is not expert in the use of image analysis tools but who wishes to use them in interpreting an image. The user views the image and then enters into the system an image-analysis goal, such as 'enhance contrast'. The system elicits information from the user about the image, applies an appropriate image-processing operator to the image then requests an assessment of its success in achieving the goal. The design provides explanations of the system's behaviour and stored chunks of text are included to describe the various supported operations.

Combining symbols and signal data

The connection between the signal data and the symbolic reasoning is mediated entirely by the user, who supplies the information about the appropriate image-processing goals and the effects of the image processing in achieving those goals.

3.3.2.2 VISIPLAN: a knowledge-based program composition system

The system [Kulikowski1995] is a knowledge-based system for automatically composing image analysis programs. The aim is allow the automated segmentation of MRI scans to be performed in response to queries couched in the terms typically used in patient records.

Images and image processing

This system is designed to work with MRI images. It employs four segmentation methods: local binary thresholding, global thresholding using a single image, global thresholding using multiple images and line detection using primary edge segments. Two schemes are used to identify segmented regions: a region classifier and a line-fitting operator.

Symbolic knowledge

The system uses knowledge about structural relations between objects and their components, spatial relations between objects and their components, morphological characteristics of objects and their components and knowledge about the task specification. The first three of these together represent a model of the expected objects and are represented in semantic networks. The fourth consists of knowledge about how decisions are made in the selection of reference objects, MRI modalities, foci, focusing operators, segmentation methods and recognition schemes. This knowledge is represented using production rules.

Operation

The user specifies the analysis in terms of a goal, available image modalities and a reference object. For a given task, an abstract plan is formulated in terms of a set of reference objects which must be found. For each reference object the system selects a focus, focusing operator, segmentation method and recognition scheme.

Combining symbols and signal data

The reference objects are defined in the knowledge base in terms of their spatial characteristics, their spatial relationships to other objects and their spatial separability from other objects and it is on the basis of these definitions that the segmentation operator is chosen. The choice of recognition scheme (region-based or line-based) is determined by the morphology of the object. So the symbolic information which

describes objects is used to identify a set of image features which image processing can be used to detect. Once these features are detected, no further use is made of the signal information in the region which is taken to correspond to the object.

3.3.3 Expert systems with image processing

Another class of systems consists of those which provide an expert system for an image interpretation task and incorporate some image-processing operators into the system, so that at least a proportion of the data required by the expert system can be obtained automatically from the image.

3.3.3.1 AUTOMEX: knowledge-based classification of breast lesions

Stewart [Stewart1994] presents a knowledge-based system for automatically classifying breast lesions.

Images and image processing

The system is for use with digitised mammograms and includes image-processing routines for measuring four characteristics of circumscribed lesions: density, contour, size and form. It is not clear from the published account how the abnormalities are detected.

Symbolic knowledge

Knowledge about the abnormalities found on mammograms and their diagnostic significance is represented as production rules.

Operation

The system works by backward chaining into a system of menus which elicit from the user the required characterisations of mammographic signs in order to prove

or disprove a hypothesis about the sign. The hypothesis will normally be a diagnosis. In addition, the system includes an image-processing module which measures four features of images of circumscribed lesions, namely density, contour, form and size. Thresholds were determined to split the measurement range into sub-ranges corresponding to radiologists' assessments.

Combining symbols and signal data

Circumscribed lesions, as defined in the symbolic knowledge base, have a number of properties and each of these has a number of values. For four properties there is a mapping between the symbol and the output of an image-processing operator which computes a value for a relevant property of the photometry. The values obtained are split into sub-ranges which correspond to the symbols for property values in the knowledge base. This correspondence was established experimentally.

3.3.3.2 Rule-based reconstruction of 3D angiograms

Smets et al. [Smets1990] describe an expert system for the labelling and 3D reconstruction of coronary arteries from coronary angiograms representing two projections.

Images and image processing

The images are stereo pairs of 2D angiograms. The image processing delineates the blood vessel segments using a ridgepoint detector which finds the maximum intensity point in all one-dimensional segments of an image. Since a blood vessel appears as a high intensity region bordered by two almost parallel lines, a region-growing algorithm is used to identify the parallel lines on either side of the centre found by the ridgepoint detector.

Symbolic knowledge

The system contains rules which capture constraints on the possible arrangement of blood vessels. Some of the constraints concern properties of individual segments: position, direction, grey-level and thickness. Some concern relations between segments: above, thicker-than, left-of, connected-to and same-direction.

Operation

The process has three stages. First the images are processed to delineate the blood vessel segments, second the coronary tree is identified using a rule-based model of normal and abnormal anatomical structures and finally the two projections are used to form a 3D reconstruction. All segments are first initialised to all possible interpretations and the rules used in the recognition of the coronary tree serve to eliminate impossible or unlikely interpretations. In general the output at this stage leaves one or two remaining solutions. The 3D reconstruction requires matching the two projections, this is trivial if the output from the previous phase is unambiguous, otherwise it involves calculating the permitted disparities, finding potential matches and calculating goodness-of-fit based on: diameter (which is the same in all projections), similar grey value and longest matching length.

Combining symbols and signal data

The signal-symbol mapping takes place in two stages. First a set of abstractions is created from the image data, that is, the segments. Next a labelling operation is performed, proceeding on the basis of rules which eliminate impossible labellings. Some of the rules include signal data, as shown in Figure 8.

```

(view^id RAO)
{(segment_lca ^id <id>^lad T
  ^begin <begin> ^end <end>) <segment>}
(it_is_so^that (segment_lies_in <begin> <end>
1 1 1 1 1 1 0 0
1 1 1 1 1 0 0 0
1 1 1 1 0 0 0 0
1 1 1 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0))
(it_is_so^that (vertical <begin> <end>))
-->
(modify <segment> ^lad F))

```

Figure 8: one of the rules in Smets et al.'s system. The rule consists of a condition and an action, to be performed when the condition is fulfilled. A complete explanation of the syntax would be inappropriate here, but this figure shows how image data is actually incorporated into the symbolic representations: the pattern of zeros and ones represents a portion of the image and 'segment_lies_in' is the name of procedure which identifies the position of the segment in the image. The condition requires that the segment lies in the top left corner of the image.

This suggests a mapping from spatial to symbolic information which makes very particular assumptions about where symbols will be on the input image. Such assumptions can be made in certain classes of medical image in certain applications, where the same projections are used and where images even from strongly abnormal patients share certain key features.

3.3.4 Knowledge-based image segmentation systems

A great deal of work has been done on the interpretation of X-ray CT and MRI images. The word 'interpretation' in this context is generally used to refer to the task of establishing a mapping between the signal data and a symbolic model of the expected

anatomy. The systems in this section, in contrast to those in the previous one, are not necessarily designed as advice-giving systems: the knowledge-based element here is a component piece of technology and does not define the application.

3.3.4.1 Knowledge-based segmentation of CT scans

Dhawan and Juvvadi [Dhawan1990] present a system for the knowledge-based analysis and understanding of CT images of the human chest. Their motivation is to provide a 3D anatomical image which can be correlated with an image of biochemical activity (e.g. a PET scan) and thereby assist in the interpretation of this image and the resulting differential diagnosis.

Images and image processing

The images are first preprocessed to remove noise and enhance features of interest. Copies of the image at successively lower resolutions are linked to form a multi-resolution pyramid on which the low-level segmentation algorithm operates. This algorithm identifies homogenous regions on the basis of proximity and similarity within and between levels in the pyramid.

Symbolic knowledge

The symbolic knowledge can be considered in three sections, a set of rules for revising the output of the initial data-driven segmentation, a set of rules controlling the process of matching image regions to the anatomical model and the anatomical model itself. The model of chest anatomy used comprises only five regions.

Operation

A set of global features is derived (e.g. *area*) which have symbolic values (e.g. *large*) for each region. These are used in the rule-based segmentation-revision process aimed at merging insignificant regions and producing a grouping of regions close to the model of chest anatomy. The process is controlled by 'strategy rules' which move the

focus of attention from one region to the next and activate merging and splitting rules as appropriate. Seven intermediate level features are then extracted. These geometrical facts about regions form the input to the high-level stage in which input is matched with object models. This is initially performed by forward chaining rules whose conditions match a region's feature vector and whose conclusions assert a fuzzy match value identifying an anatomical region. Identification is defined as fuzzy matching that exceeds a pre-defined threshold. Failure to find a complete match with a model can force the initial segmentation to be revised.

Combining symbols and signal data

The signal-symbol mapping is performed, in part, in the initial segmentation phase which assigns a symbolic identifier to a region. The image data associated with each region is represented as a vector of values for seven properties: area, average grey level, centroid, horizontal and vertical bounds, orientation, elongation, and neighbours. Implicit in this is a representation of the co-ordinates of the image, the representation, in other words, is not wholly symbolic. The later matching phase is performed entirely on this representation without referring back to the original image data.

3.3.4.2 ERNEST: the knowledge-based interpretation of scintigrams

The system [Niemann1990] is a generic signal analysis tool which has been applied in a system for analysing sequences of scintigraphic images of the heart [Niemann1985b] and more recently for MRI-based diagnosis of the knee [Weierich1993]. The system is intended be able to derive a symbolic description of the medical evidence contained in an image, without any human input whatsoever.

Images and image processing

In the scintigraphic system, the input consists of sequences of between 12 and 32 images of 64 by 64 pixels. Each one is smoothed and then a simple set of heuristics

is used to identify a point which, it is certain, will be within the left ventricle. The image is then reconstructed in polar co-ordinate space around this point and an edge detection program used to identify the ventricle boundary. Direction and expected ranges of occurrence are used to divide the edge into the four ventricular regions. Subjective evaluation suggests that these low-level processes successfully analysed 93% of 420 test images.

Symbolic knowledge

The symbolic knowledge representation has two components: a declarative component comprising an 'is_a' and a 'part_of' hierarchy and a procedural component containing rules for computing areas, lengths etc., for deriving descriptions of heart motion from area changes over the sequence, and for combining medical evidence.

Operation

The operation of the system combines data-driven, bottom-up image processing and query-driven top-down knowledge-based processing. The user asks if a particular concept represented in the knowledge base is true of the sequence under analysis and the system computes a likelihood. The query must match exactly a concept represented in the knowledge base. The answer is computed by breaking down the concept into its components in order to generate a set of primitive concepts, which correspond to the results of the image processing.

Combining symbols and signal data

Symbols and signal data are combined at the leaf-nodes of the symbolic representation. Control knowledge is used to select the most promising match between image features (the results of the image processing) and the concepts required to answer the query. This basically means using an optimised search strategy to find the matches which provide the best fit values at all levels of abstraction. The basis for the optimisation is a measure of 'goodness of fit', which is used to ensure the system explores first the interpretations based on the most promising partial matches.

3.3.4.3 Knowledge-based classification of MRI scans

Li [Li1993] describes a knowledge-based architecture for classifying brain tissue in MRI images.

Images and image processing

The starting point is the raw MRI data which consists of three separate images, the T1, T2 and PD images. A segmentation algorithm ('unsupervised fuzzy C means') is used to identify ten classes in the 3D image space. Further image-processing operations are then performed as required by the symbolic reasoning process. Some of these are rather ad hoc: for example, a deformity is identified in the white matter on the basis of lines drawn outward from the image centre 22.5 degrees apart. If a line doesn't cross white matter on its way to the edge of the image, then a gap is noted.

Symbolic knowledge

The symbolic knowledge is a set of rules representing knowledge about the application of image processing to MRI scans. For example: *white matter is clustered into one or two classes by the UFCM procedure, it is smaller in T2 than gray matter and sometimes smaller than one or two skull classes*. Facts like these are used to design procedures for identifying the tissue classes and determining if a tumour is present.

Operation

The segmentation procedure is applied to the images to separate them into ten regions. Thus the images are deliberately over-segmented. Then the rules are applied in a sequence of four steps which identify all the normal tissue in the brain, and label any remaining tissue as abnormal. First the skull tissues are separated from the tissues of interest in this application, and then, white matter is located. A test is next performed to see whether the white matter is deformed. If no deformation is found the cerebro-spinal fluid is identified and tested for deformity. Finally the grey matter is detected and a

further test performed to see whether small deformations of white matter should be considered significant. Each step involves the application of rules which represent a combination of knowledge about the results of applying image processing to a region associated with the tissue type and the expected location of the tissue.

Combining symbols and signal data

The first phase of the processing, the segmentation, results in a representation of what are termed “clusters”. The later rule-based processing involves the application of operators to signal information, which is stored with the clusters. The image-processing operators called from the rule-based processing map the signal data which correspond to the clusters to the symbols which identify the relevant properties of the brain regions. The relevance of these properties is established through experience in the application of image processing to MRI images, and is not “deep” knowledge about brain anatomy. For example, one rule states: *a deformity is identified if there is a gap of over 25 pixels, or more than one gap on the left or the right or there is not a gap at the front of one of the clusters.*

3.3.4.4 A blackboard system for interpreting CT scans

Li et al. [Li1994] describe a system for model-based interpretation of CT scans.

Images and image processing

The system is designed to work with brain CT scans. The image processing includes a preprocessing step based on grey-level thresholding. More sophisticated region-based segmentation operators are provided, termed the ‘edgementation’ operator and the ‘cavity detector’. A curvature detection algorithm is provided to detect points of maximum and minimum curvature.

Symbolic knowledge

The system contains five different ‘information sources’:

- brain CT-scan domain knowledge
- knowledge about image primitives (points of interest, edges, regions)
- image-processing domain knowledge
- knowledge about the mapping between the brain model and the image
- control knowledge

Operation

At the heart of the system is a blackboard with five panels, one for the image data, one for the image primitives, one for the low-level symbolic representation and one for the brain model. The interpretation of the scans – by which is meant the mapping of image regions to brain model objects – is triggered by a “hypothesis generation” module which creates hypotheses on the blackboard. These trigger “model directed heuristics”, actually instructions – e.g. find edge – to start processing in order to produce the image primitives. Data-directed heuristics are similarly unsophisticated (one suggests splitting a region and is triggered whenever the region detector 'discovers' a region which is too large to be plausible). In addition to the application of heuristics, processing is guided by the application of constraints on the correspondence between features of regions and features in the object hypotheses.

Combining symbols and signal data

The signal to symbol mapping takes place in part when image primitives such as regions are identified by the various image-processing modules. Image primitives are represented as symbols, but some of the spatial information remains implicit since the positions of image primitives are represented in co-ordinate form. These co-ordinates must be converted into the statements about the geometrical arrangement of brain objects that figure on the brain model panel as the mapping from image primitives to brain objects proceeds. Little information is given in the paper about the constraints used to represent the correspondence between features of regions and

features in the object hypotheses. This is a shame because these constraints must encapsulate the information used in mapping the image primitives to brain objects.

3.3.4.5 Model-based object recognition

Robinson et al. [Robinson1994] describe a system which uses a symbolic model of human anatomy in the recognition of brain objects.

Images and Image Processing

The system has been used to provide a labelling of structures near the midline of a three-slice brain MRI data set. The slices are segmented using an edge detector based on the Canny edge detector. Constraints on the size and position of the identified regions are used to select a set of regions as candidate anatomical structures. A shape description module computes a description of these regions in terms of their skeletons and boundaries.

Symbolic knowledge

The model of brain anatomy consists of a set of frames describing the properties of anatomical features. These are connected by three sets of links which form: a spatial adjacency graph, a part hierarchy graph and an inheritance network. Each frame also has associated with it a model of the shape of the anatomical feature it represents. These shape models are derived from a standard set of MRI images known as the Talairach atlas. The shape models are computed from points identified on the images by the investigator, which allow segments representing the skeleton and the boundary to be determined by the computer. The final description is in the form of a lattice containing skeleton and boundary information about components of the region.

Operation

The labelling of image parts is achieved in the following steps. First, image processing is used to identify a set of regions. Next shape models are computed for

these regions which are then passed to the shape description module. The resulting descriptions are matched against the shape descriptions associated with frames describing midline structures. A good match results in a provisional labelling of the structure. The system uses a technique which allows the modelling of hypothetical alternatives to construct a set of consistent labellings. This involves the notion of a viewpoint, or set of consistent labellings. Each provisional labelling sets up a new viewpoint. The application of rules which merge consistent sets of labellings and remove those containing contradictions converges on the interpretation with the highest number of image to model matches.

Combining symbols and signal data

The shape descriptions computed from the Talairach atlas represent an image description derived from signal data. This representation makes explicit some of the information about the spatial arrangement of features and can be viewed as an intermediate representation between the signal data and the symbolic model. The link between the image data and the symbolic knowledge is made in part when the descriptions are compiled, and in part in the process of matching the compiled descriptions with those stored in the frames.

3.3.4.6 Atlas-based segmentation of MRI data

Collins et al. [Collins1995] describe a technique for the segmentation of 3D MRI images of the human brain, based on a model of brain anatomy. The term ‘model’ is used to refer to a segmented and labelled MRI image, also referred to as an ‘atlas’.

Images and image processing

The images used are 3D MRI brain scans and the image processing is an iterative hierarchical registration procedure which matches the image to be segmented with the model, an image whose segmentation is known. The registration consists of two phases, a linear and a non-linear transformation of the data, in both of which image

intensity and gradient magnitude are matched. In the linear phase the best affine transformation matrix is found in order to maximise the correlation of features. This linear registration is performed at increasing spatial scales, starting with very blurred data. In the non-linear phase the same method is used in small neighbourhoods of the data set. The non-linear warp is composed of a set of local linear transformations which are performed iteratively, revising the estimate of the required deformation.

Symbolic knowledge

The only symbolic knowledge in the system is the labelling of the model.

Operation

The segmentation of the brain MRI scan is performed by registering the scan with the model, using the linear and non-linear transformations described above. The labels in the model are then mapped onto the image through the inverse of the transformation determined by the registration procedure.

Combining symbols and signal data

The link between a symbolic representation and one in signal data is achieved through the manual labelling of manually segmented signal data.

3.3.4.7 Knowledge-guided segmentation of CT scans

Kobashi and Shapiro [Kobashi1995] present a detailed account of a knowledge-based system for the accurate segmentation of CT scans.

Images and image processing

The following image-processing operations are performed on abdominal CT scans: thresholding, morphological opening and closing, set operations, connected components operations, measurements of shape, position and size.

Symbolic knowledge

The knowledge base includes a representation of the spine, aorta, kidneys, spleen and liver. The following properties are recorded for each: rank order of brightness, grey-level range, height of grey-level cliff, location relative to landmarks, adjacency with other organs, relative size, relationships with slices at other levels, positive and negative shape constraints. In addition to this “domain knowledge”, the system contains a representation of knowledge about the application of the image-processing operators for this problem. This knowledge was established experimentally and determines first the order in which the organs are identified; second, the order in which slices are processed and third, the sequence in which the image-processing operators are applied to each slice.

Operation

The system identifies each of the required organs in turn, in each of the slices in turn, starting with the slice that is two fifths of the organ’s expected height from its bottom-most expected slice. The segmentation can involve a sequence of up to eleven image-processing operations before the outline of the target organ meets the set criteria.

Combining symbols and signal data

The organ properties in the symbolic knowledge base are formulated so that they are directly computable. Symbols and signal data are combined by the set of image-processing operations which are applied to the image and return a value which is used in the symbolic processing.

3.4 Discussion

Chapter Two reviewed research into different kinds of system which could be used to support decision making in diagnostic radiology and evaluated these systems against a common set of criteria. The review in this chapter has a different goal, the aim is not to evaluate the various systems but to consider the methods used to combine signal and symbol information. The next section reviews the different kinds of signal-symbol mappings that these systems use. Each of the approaches employs some form of representation of signal data which can be associated with a symbolic representation. The different approaches are discussed, and conclusions drawn about the appropriate path to follow in this thesis, in Section 3.4.2.

3.4.1 Mapping between signals and symbols

In general considering mappings between signals and symbols means comparing the ways in which signal information can be mapped to symbolic representations, or symbolic representations can be extracted from signal data by image-processing operations. It is worth noting, however, that in a number of systems where both image data and symbolic information are used there is no mapping from one to the other, or rather any such mapping is done by the user. The systems described by de Rosi et al. [Rosi1992], Kulikowski [Kulikowski1995] and Rogers [Rogers1995] associate symbolic terms with image-processing operators and use symbolic reasoning to control the processing of medical images, but they do not associate symbolic terms with information derived from images. The remaining systems all provide some form of mapping between information represented in signal data and information represented symbolically. The diagram in Figure 9 shows the different kinds of mappings found in these systems.

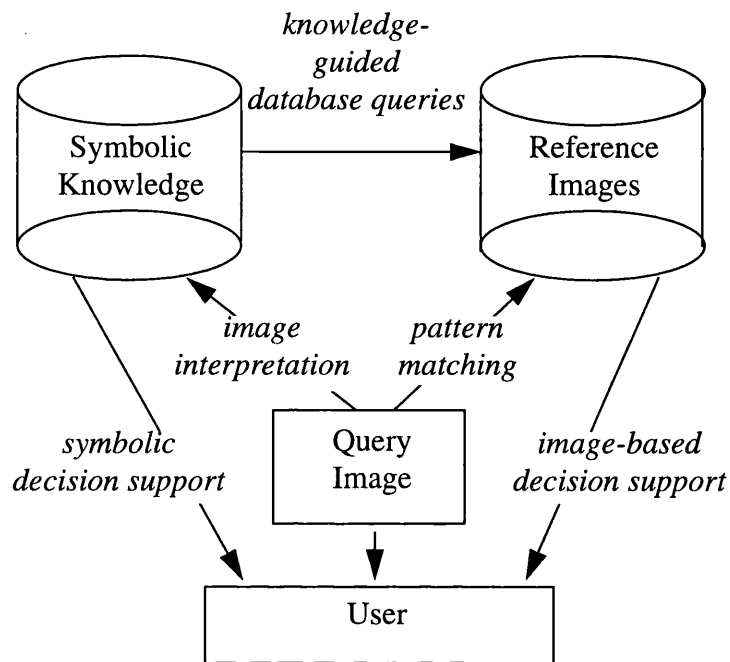


Figure 9: a diagram showing the different ways in which image data in the image being interpreted (the Query Image) and image data in previously interpreted images (Reference Images) are combined with symbolic knowledge to provide decision support.

A system which supports some kind of signal-symbol mapping could, in theory, support any mapping from symbols to signal data, or from signal data to symbols. In practice systems support mappings:

- from the symbolic knowledge to the reference images (knowledge guided access to an image database)
- from the query image to the reference images (pattern matching)
- from the query image to the symbolic knowledge (image interpretation)

The next three sub-sections consider these kinds of mapping in turn.

3.4.1.1 From the symbolic knowledge to the reference images

In AXON symbolic knowledge guides access to a database of previously interpreted images and images retrieved from this database form the basis of the decision support. In this system the mapping between the symbolic knowledge and the image is achieved by associating each image with a set of keywords which describe the image, presumably on the basis of an expert's interpretation. The keywords are organised into a set of taxonomies representing different views of the domain supported by the system and a set of search heuristics is used to ensure that an appropriate selection of images is retrieved for any given query.

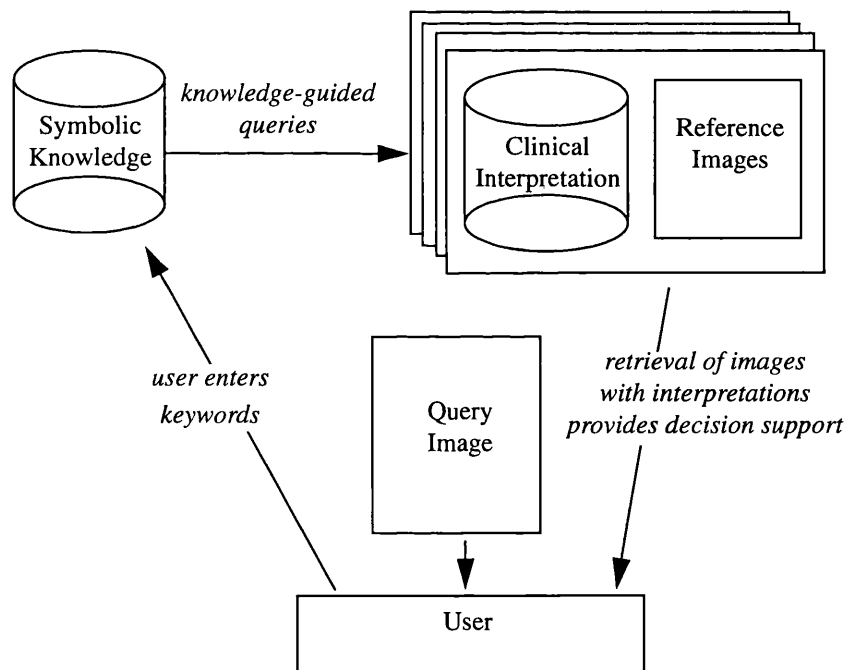


Figure 10: AXON, a system providing knowledge guided retrieval of information from an image database. Information about the clinical interpretation of each image is stored with the relevant image, as represented above by the stack of overlapping rectangles. The user submits keywords which are used by the knowledge-based search engine to search for appropriate images in the database. The retrieved images, and the associated clinical information form the basis of the decision support.

This approach can be developed further, for example by providing a richer symbolic description of the content of images than simply a list of keywords. Chang et al. [Chang1988] have proposed the use of 2D strings to index image databases on the basis of the relationships and attributes of objects in the image. Authors such as Karssemeijer [Karssemeijer1988] working in knowledge-guided image interpretation have developed semantic net representations of the anatomy imaged in different classes of medical image which could be used to describe image contents in a way that could guide image retrieval. It is not, however, clear that this kind of formalism captures the sort of information that users would use in queries any better than a set of keywords does. The most important use of an image database is, perhaps, to access the range of images representing the possible presentations of a known lesion and that there is only rarely a need to retrieve the set of images in which object A appears in relation B to object C having property D.

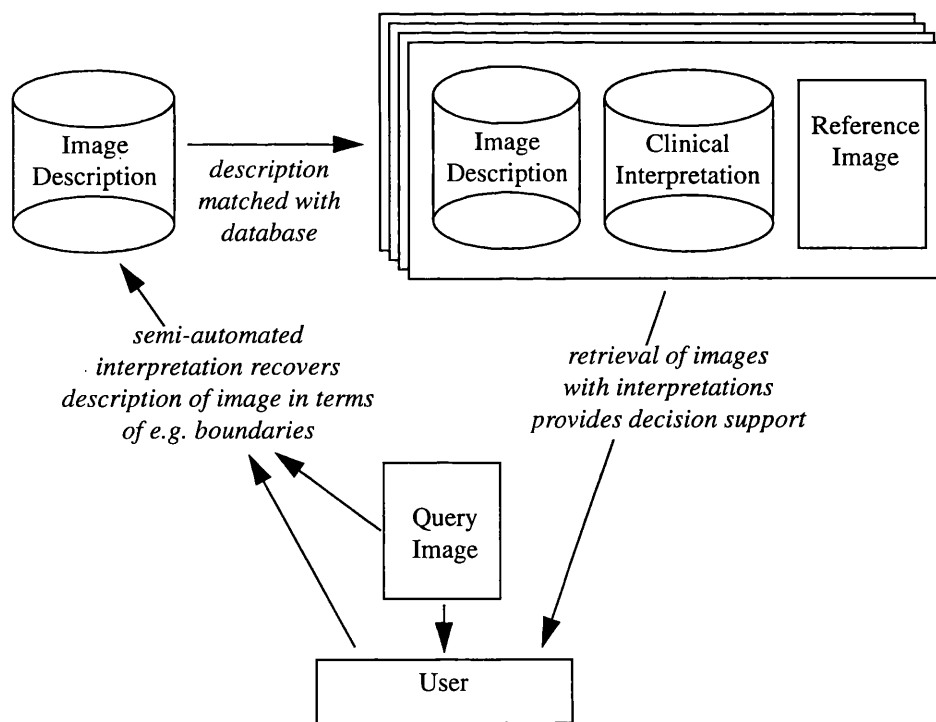


Figure 11: I^2C , a system for indexing an image database on the basis of image content. Each image is associated with stored information, which includes details of the clinical interpretation and an description of image features derived from image processing. A semi-automatic interpretation of the query image can be used to retrieve matching images from the database.

Orphanoudakis et al. [Orphanoudakis1994] have developed a system in which images are classified using criteria based on clinical interpretation but also on semi-automated analysis of image features. This analysis recovers a symbolic image description which is in terms of properties of the image data. The authors state that the system is designed to be independent of any particular approach to image description but their account includes a description of a technique in which a segmentation algorithm identifies chains of edge segments which define the boundaries of regions of interest. The properties of these symbolically represented descriptions can then be compared with descriptions derived from the query image. The mapping from the query image to the symbolic representation is semi-automated.

3.4.1.2 From the query image to the reference images

In the Collins et al. [Collins1995] system the query image is matched with an image from an atlas of reference images and the labelling associated with the reference image is applied to the query image. The mapping from the signal data in the query image to the symbolic labelling is achieved by associating each stored symbol with a description in terms of actual image data which can be matched with the query image. The same pattern-matching approach is used in systems where particular anatomical features or abnormalities are represented using templates or idealised images which can be matched with the query image. For example, in a system described by Matsumoto et al. [Matsumoto1992] the correspondence between the reference images – which are templates of idealised lesions – and the query image is the basis for the decision support (in this system there is no symbolic knowledge). Where pattern-matching is used to support the segmentation of a query image, as in the Collins system, the final solution may represent some kind of best fit between the query and the set of reference images. Where it is used, as in the Matsumoto system, to assist in the detection of abnormalities, a threshold must be set on the function which matches image data and template to determine when an adequate match has been found.

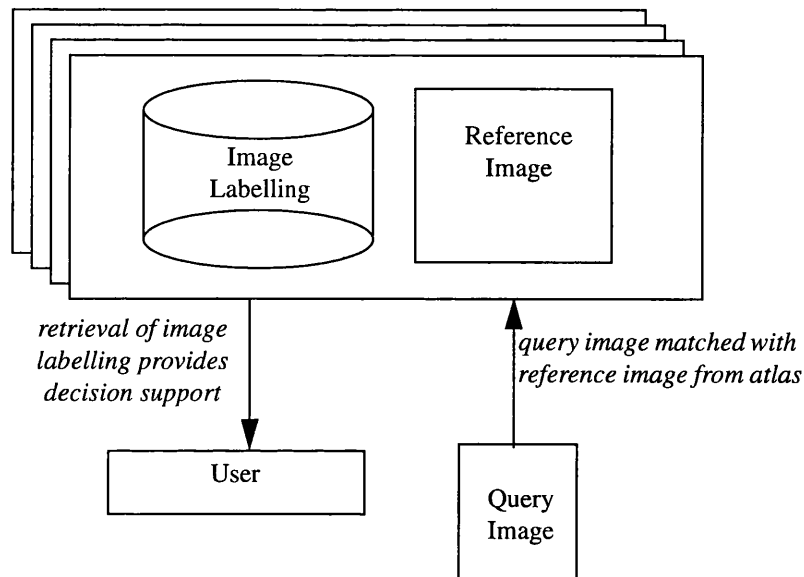


Figure 12: in the Collins et al. system the association between the symbolic labelling of reference images and the data in the images is used to assist the user by providing anatomical labels for a segmentation of the query image.

3.4.1.3 From the query image to the symbolic knowledge

The phrase ‘image interpretation’ is used here to refer to the recovery from an image of a symbolic description of the scene depicted in the image. The symbolic description consists of statements about the objects in the scene and how they relate to each other. For example, a symbolic model of the anatomy of the abdomen would consist of statements about the different organs and how they are arranged relative to each other. Mapping from an image of the abdomen, for example on a CT scan, involves using image processing to identify the different organs. This means that each organ, or rather the symbol representing each organ, must be associated with a description of the organ, a description derived from the output of the image processing. Three approaches used in the systems described above are to describe anatomical features in terms of properties of the axes, the boundaries or regions associated with their appearance in image data. Robinson [Robinson1992] uses an axial description, Niemann [Niemann1985a] and Orphanoudakis [Orphanoudakis1994] use boundary

descriptions. The commonest technique is to use region-based descriptions: [Dhawan1990, Li1993, Li1994, Stewart1994, Kobashi1995, Kulikowski1995].

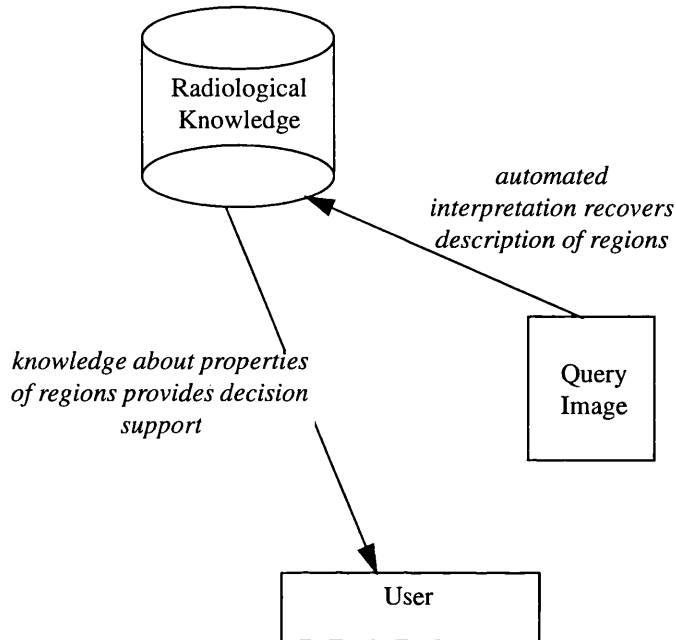


Figure 13: in AutoMEX [Stewart1994] processing of the query image results in a set of measurements which describe the image regions corresponding to abnormalities. These measurements are combined with the symbolic knowledge to provide inferences about the kind of abnormality likely to be present in the image.

To recover from an image a description which can be matched to a description stored with the image, it is generally necessary to segment the image into distinct regions and derive a description of each region. It is important to note that the selection of a particular segmentation technique – the detection of boundaries between regions or the separation of the image into regions of uniform brightness – does not necessarily imply a commitment to a particular technique for representing the object depicted in the region. A region-based segmentation technique could be used to identify regions on an image, but the stored description of the image doesn't also have to be in terms of properties of regions. Once a region is identified, its boundary and its axis can be recovered. On the other hand, an edge-based segmentation algorithm could recover enough information about the boundaries in an image to allow objects to be described

as a collection of boundary segments, but which could be too incomplete to allow the region to be identified.

The three representation formalisms mentioned above are the only ones used in the systems described in this chapter. There are, however, other techniques not described in the literature on systems integrating signal and symbolic representations. These include Active Shape Models which contain an explicit model of the variability of objects of the kind which are found in medical images [Cootes1994] and scale-space approaches which search for variation or stability in an object's appearance as image resolution varies [Fritsch1994] or the use of symmetries to describe an object's shape [Walker1990].

3.4.2 The representation of signal data

In each of the systems which provide a mapping between symbolic knowledge and signal data, symbolic terms are associated with a stored representation which can be matched with signal data, or with a representation which can be recovered from signal data. The stored representations associated with symbols in the work described above include real images and idealised images, region-based descriptions, boundary-based descriptions and axial descriptions.

A description of a symbol in terms of image data simply means associating the symbol with an image, or a region of an image. In the same way, mapping from a piece of signal data to a symbol means matching the query image with the stored image. A description of a symbol in terms of regions means associating the symbol with a vector of values (or ranges of values) for properties taken by the region: size, shape average brightness, relative position and so on. Mapping from signal to symbol means dividing the image up into its component regions, measuring the properties of these regions and matching them against the values stored in the feature vectors. In the case of boundary

and axial-based descriptions the same process of image segmentation is required, but the matching uses vectors which are values of boundaries or axes.

In choosing the appropriate representation for signal data, it is necessary to consider four questions. The first two concern the applicability of the description method to the domain covered by the symbolic representation: *coverage*, that is, can all the required symbols be described, and *uniqueness*, that is, do all the different symbols have different descriptions? The other two concern the applicability of the method to the signal data: *recoverability*, that is, how easily can the description of an image be obtained, and *sensitivity*, that is, how reliably can the recovered description be matched with the stored descriptions associated with symbols?

Consider how the different categories of representation fare when judged against these criteria.

real or idealised images

- *Coverage and Uniqueness*: one of the principal problems with this approach is that objects which are included in the symbolic representation may have many different appearances. Complete coverage of the required domain would thus require many different images to be associated with a single symbol. The technique is therefore only suitable in situations where the symbols which must be associated with signal data have a predictable appearance in the signal data.
- *Recoverability and Sensitivity*: the great advantage of using the image data itself as the representation is that the representation does not then need to be recovered from the image, although in practice some form of normalisation will generally be required. The real difficulty with this approach is in reliably matching the input and the stored image data. The matching process must

take into account not only the normal variation between different patients but also the variation due to the image formation process.

region-based descriptions

- *Coverage and Uniqueness*: this is the approach used in the majority of the above applications; it seems possible to describe most features of interest in many classes of image in terms of their shape.
- *Recoverability and Sensitivity*: one of the difficulties in this approach lies in obtaining an accurate segmentation of a region. This is a key problem in image-processing and many approaches have been developed, some of which will be adequate for the applications considered here. Any lack of sensitivity in the matching of stored region representations with those derived from an image is likely to stem from the need to accommodate inaccuracies in the segmentation.

boundary-based descriptions

- *Coverage and Uniqueness*: this approach seems most appropriate for rigid shapes and therefore unsuitable for most medical images. Another problem is that it is difficult to apply hierarchically, it is difficult to decompose a non-rigid shape into sub-regions on the basis of boundary representation. This approach is perhaps appropriate in images where there are clear boundaries and long thin structures.
- *Recoverability and Sensitivity*: obtaining a boundary-based description of an image requires an accurate segmentation, so the difficulty of recovering a boundary-based description is similar to that of recovering a region-based description. Such methods are often described (e.g. [Robinson1992]) as vulnerable to small changes in outline or errors in segmentation.

axial descriptions

- *Coverage and Uniqueness:* axial-based descriptions are most suitable for shapes which have a clearly defined axis. Many biological and man-made shapes are appropriately described in this way, others are not. This is not a promising approach for the representation of ill-defined shapes, such as abdominal tumours, or for objects which are generally small and round, such as chest lesions and many mammographic abnormalities.
- *Recoverability and Sensitivity:* one reason for the popularity of axial descriptions is that there are a number of different techniques for recovering the axis of an object. An advantage is that finding the axis of an object is a step towards an object-centred (i.e. image-independent) representation.

3.5 Conclusion

The aim of this thesis is to investigate how information represented in different forms can be combined to provide decision support for the interpretation of medical images. This chapter has considered how information in a symbolic knowledge base can be combined with information in digital images. The various approaches which have been taken to this problem have been reviewed and the different techniques used for mapping between symbolic representations and signal data have been assessed.

This review, more tightly focused than that in Chapter Two, has identified a variety of systems which employ both signal and symbolic information and which could function as decision support systems. Several of these systems did not provide a mapping between signal data and symbolic information, allowing only a relatively weak form of decision support – such as facilitating access to a database – or requiring considerable input from the user – as in the consultation systems which provide an intelligent interface to image-processing software. The systems which did provide

some form of mapping between the two kinds of representation were all based on the association of a symbolic term with a representation which could be matched with signal data, or with a representation recovered from signal data. Different systems used different representations, some based very closely on the original signal data, some requiring a complex process of abstraction from it.

The less abstract representation schemes are cumbersome because large amounts of image data have to be considered and normal variations between different instances of a class of objects, or different depictions of the same object, are ignored. The more abstract schemes are vulnerable because of the difficulty of abstracting information from images. Segmentation is a particular difficulty. Some schemes may be able to overcome these difficulties, for example flexible models which capture the variability in image data. These may prove a useful development.

The approach taken in the rest of this thesis has been to avoid dependence on any one scheme – since all are likely to prove inappropriate in some circumstances – and to concentrate first on developing a framework to decision support within which different approaches to the representation of signal information can be combined, and then to consider carefully the suitability of different approaches to the relevant applications. The next chapter describes the framework adopted.

CHAPTER FOUR

A FRAMEWORK FOR DECISION SUPPORT

Science does not aim, primarily, at high probabilities. It aims at a high informative content, well backed by experience.

Karl Popper

4.1 Introduction

In Chapter Two I reviewed the computer systems used to aid decision making in radiology and identified four distinct kinds of system, each based on a different way of storing knowledge or information in a computer and then retrieving it to assist in the interpretation of medical images. No one kind of information source or single class of decision support system is likely to prove pre-eminent and yet a proliferation of systems would involve redundancy and be confusing and daunting for users. There is, therefore, a strong argument for developing an integrated decision aid which is able to provide information of different kinds on request, or in response to different problems. In particular, as I have argued, there is a need for systems which combine the kind of signal information contained in images and analysed in image processing with the kind

of symbolic information represented in knowledge bases, or in clinical databases and electronic patient record systems.

In Chapter Three I considered the various approaches which researchers have taken in building systems that combine information represented in these different forms and argued for one which allows a certain degree of flexibility in the association of symbolic terms and image data.

In this chapter I present an approach to the integration of image processing in decision support systems motivated, in part, by a conviction that a strong organising metaphor or model is necessary in a system which hopes to enable a user to make sense of a variety of different pieces of information. Stated simply, my approach is to present information in the form of arguments for and against different potential solutions to a decision problem.

Shortliffe [Shortliffe1991] identifies three methodologies used in the development of clinical decision support systems: Bayesian probabilities, decision theoretic methods (which could be considered as an augmentation of the Bayesian approach) and artificial intelligence or expert systems. In Chapter Two I identified another two which have been used in computer aids for image interpretation: image databases and image processing. The methodology adopted here is firmly in the artificial intelligence category, and has been employed for two reasons. First, it was developed to allow the use of uncertain information in decision support without any prior assumptions about the nature of the uncertainty or the formalism used to represent it and this flexibility allows a system to present information of very different types. The second reason is that it is based on an explicit model of decision making which, it is hoped, will provide a transparent mechanism for organising the display of information, so that users will understand the basis on which the advice is given.

In the next section I describe the symbolic model of decision making which forms the basis for this approach, and in the following sections I describe how this model is extended to form a model of image interpretation and how this extended model can be applied to a range of problems in radiology. It should be stressed that the importance of the model is not as a theory of how people actually make decisions, or of how radiologists actually interpret images. It is rather a hypothesis about how relevant information may usefully be presented to people who have to make decisions.

4.2 The Symbolic Decision Procedure

The Symbolic Decision Procedure used here has been developed from work carried out at the Imperial Cancer Research Fund by Fox and co-workers [Fox1988, Fox1990, Huang1993, Krause1995]. There have been three distinct stages in this research. Originally, part of the impetus was an interest in psychologically valid theories of decision making. The subsequent development of the symbolic decision procedure, as a practical basis for decision support systems, was largely determined by the requirements of the Oxford System of Medicine a project to develop a decision support system for general practice. The third, and most recent, element of the work has provided a formal basis for this approach.

The Symbolic Decision Procedure is composed of four types of knowledge:

- *generic decision rules*: a set of logical rules describing the process of decision structuring and decision making
- *task-specific knowledge*: a set of facts about different kinds of decision (e.g. diagnosis, the selection of a treatment)
- *domain knowledge*: a set of facts describing the domain in which the decision is to be made (e.g. knowledge about a medical problem)

- *case-specific knowledge*: a set of facts describing what is known about a particular case (e.g. an entry on a patient's record)

The generic decision rules represent a very abstract, very general model of the processes that are involved when people make decisions. In the Clark et al. [Clark1990] presentation of the Symbolic Decision Procedure, five processes are identified:

- *proposal*: the dynamic proposal of candidate solutions
- *argumentation*: the construction of arguments for and against solutions
- *annotation*: annotating arguments with an indication of their force
- *relation*: recording patterns of combination between arguments
- *aggregation*: to produce a ranking or partial ordering of candidates

For these generic rules to be applied in making a particular decision, knowledge about the kind of decision to be made is necessary so that the procedure can identify the elements of the domain which are to play the required roles in the decision procedure, to serve as the candidates and arguments and so on. This task-specific knowledge consists of specifications for various classes of decision. The specification of diagnosis might include the information that the decision problem concerns the symptom the patient presents with, the candidate solutions are the various diseases which can cause this symptom and the arguments for and against are based on the other signs and symptoms of these diseases. If the general rules are thought of as schemata, with variables or place-holders left empty, the task-specific knowledge instantiates some of the variables, allowing the general rules to operate as more specific rules, describing particular kinds of decision.

The third kind of knowledge used in the decision procedure is that which describes the relevant domain – the knowledge base from which the decision support is generated. In a clinical decision support application this would be a subset of general medical knowledge, for example, representing facts that enumerate all the diseases which are possible causes of the symptoms that may be entered as problems as well as all of their signs.

The final element in the decision procedure is the case-specific knowledge. In this context, this would be the information about a particular patient and his or her medical problems.

In order to understand how the Symbolic Decision Procedure operates, it is worth considering a simple example, the diagnosis of a woman who presents with a lump in the breast. Here the clinician must consider the different possible causes of breast lump. For the purposes of this example, let us say that a breast lump can be caused by one of two processes: benign change, or breast cancer. The clinician must consider the arguments which would support these two diagnoses: are any of the risk factors associated with breast cancer present and are there associated clinical signs? Some of the possible arguments require evidence which must be obtained from an investigation, such as an ultrasound scan or a mammogram.

To apply the Symbolic Decision Procedure in this case, a generic rule which describes how candidates are proposed is used in conjunction with task-specific knowledge describing diagnosis (which would include the fact that in diagnosis the candidate solutions are the possible causes of symptoms), and the domain knowledge (which would include facts that the causes of breast lumps include benign change and breast cancer) to allow the inference that benign change and breast cancer are candidate diagnoses.

The rule which describes how arguments are constructed is used in conjunction with task-specific knowledge (which would include the fact that in diagnosis arguments are based on the signs and symptoms of candidate diagnoses) and the domain knowledge (including the fact that calcifications are a positive sign of breast cancer) and also the case-specific facts (including, for example, that microcalcifications have been detected) to infer that there is an argument in favour of the diagnosis of cancer.

Task-specific knowledge which identifies positive signs as supporting but inconclusive arguments would be used by the generic rule for annotation. The rules provided for evaluation might include one which, in the absence of a conclusive argument, establishes a preference ordering among supported candidates on the basis of the number of supporting arguments, for example. More complex rules could indicate what action, if any, should be taken if a mix of arguments, for and against a candidate, was obtained.

The Symbolic Decision Procedure, as described above was developed to meet the requirements of a system called the Oxford System of Medicine (OSM) [Fox1990]. The OSM was intended as a decision support tool which would support a range of decisions in all areas of general practice. This meant that the decision procedure had to be generic, since it would be applied to different kinds of decision; that it had to be able to operate in the absence of statistical information about uncertainty, since this information would not always be available, and that it had to be possible to modify the knowledge base without compromising confidence in the reasoning. All of these requirements are met by the symbolic decision procedure. The high-level rules guarantee genericity since decisions are defined at the level of the task-specific knowledge according to a common schema imposed by the high-level rules. The presentation of information in the form of arguments allows a flexible approach to the handling of uncertainty since arguments can be formed with or without an indication of their relative strength. The medical knowledge consists of atomic facts which can readily be updated and amended.

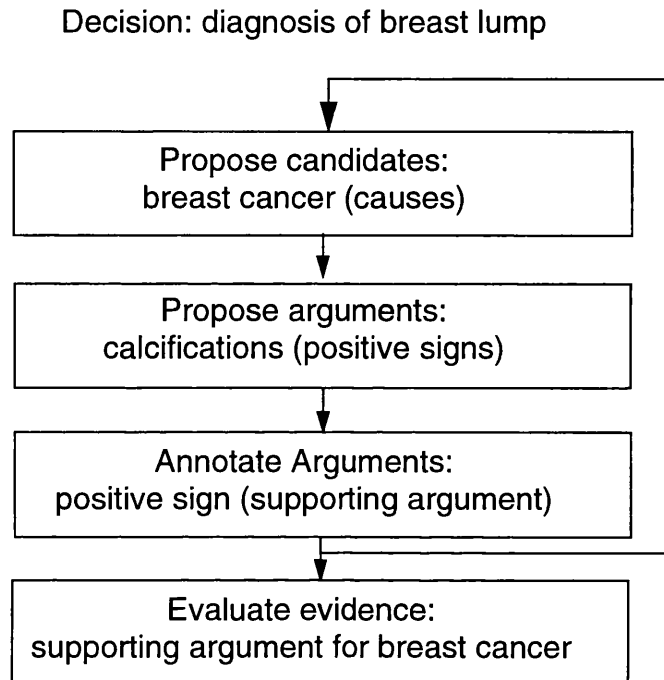


Figure 14: an example of how the decision procedure described above might be applied in the diagnosis of a breast lump. The candidates are the possible causes of breast lump, which include breast cancer. The arguments are provided by the signs and symptoms of breast cancer which include calcifications. The arguments are annotated with labels which indicate the force of the argument (positive signs support a hypothesis, pathognomic signs confirm it). The eventual decision support will include the evidence for the existence of calcifications. Once all the possible candidates have been considered, all possible arguments enumerated and all available evidence identified, any significant relations between arguments can be considered. Evaluation might involve removing all unsupported candidates and establishing a preference order of the remainder on the basis of the number of supporting arguments.

Later work on the Symbolic Decision Procedure [Fox1995] has included an attempt to support two other phases of decision making. The first occurs at the beginning of the cycle and involves detecting the problem, the need for a decision. This requires the notion of a goal to be attained or maintained. The second involves the notion of commitment to a preferred solution or course of action. These additional

phases are incorporated into a model of decision making in which the argumentation and annotation phases are elided, the relation process is now implicit and the evaluation is incorporated into the notion of commitment. This new model is presented using a graphical device which is known as the domino model.

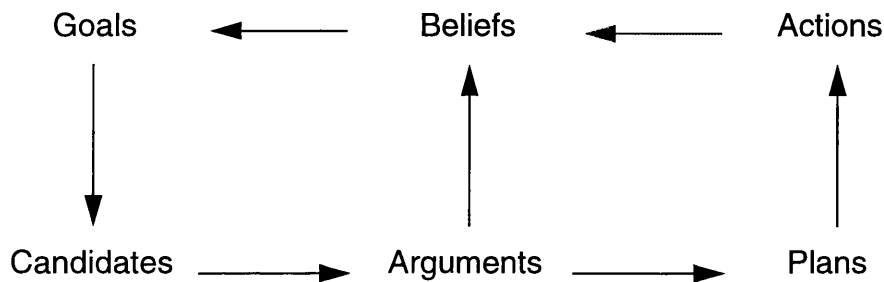


Figure 15: the domino model

A database of beliefs about the domain of interest is maintained in the system, certain beliefs can give rise to goals. Goals lead to problem-solving to generate candidate solutions and arguments for and against the candidates. It is intended that a system using the model should be able, on the basis of these arguments, to commit to a preferred candidate. This may represent a new belief or a plan of action depending on the nature of the decision. The plan will consist of a set of actions which are scheduled and performed. New information may arise as actions are performed, potentially resulting in another cycle of decision making.

The domino model develops the symbolic decision procedure in two ways. First, it supports a wider range of functionality than traditional decision support systems which simply present information about a decision problem to a decision-maker. Second in providing a sound theoretical basis for each of the steps. The step that is most relevant to the model of decision support adopted in this thesis is argumentation, and it is worth elaborating a little on the ideas behind it.

LA [Krause1995] represents an attempt to set the work of Fox et al. on the Symbolic Decision Procedure in a more rigorous mathematical framework. LA consists of a syntax for representing the arguments supporting propositions, a set of inference rules which can be used to construct the valid arguments and a set of techniques for reasoning about these arguments. LA is designed to handle propositions made up from terms connected by conjunction, implication and contradiction operators. The logic is augmented by labelling propositions with structures which provide a succinct representation of the arguments that support those propositions, making it possible to reason about the arguments themselves. LA provides a set of inference rules which can be used to derive arguments for complex propositions given arguments for simple propositions, and can derive arguments for conclusions given arguments for rules and for premises.

The rules of LA are applied to generate arguments for and against candidates. In addition to constructing arguments for propositions, it is necessary to reason about them. In LA this is done by attaching an indication of strength to each argument, that is, a measure of the capacity of that argument to alter our degree of confidence in the proposition. The symbols used to represent indications are drawn from one of a number of different 'dictionaries'. The term dictionary is used simply to indicate the collection of symbols used in a particular technique (or 'language', to continue with the dictionary metaphor) for describing uncertainty. Available techniques for describing uncertainty allow the force of the arguments to be formalised, as a probability, as a set of linguistic qualifiers or as a membership function of a fuzzy set. Different functions for describing LA consider three different dictionaries:

$\{0, +, ++\}$

$\text{strength} \in \mathbb{R}^{\infty}$

$[0,1]$

where 0 indicates ignorance, + support and ++ proof; where the set of natural numbers augmented with infinity would correspond to a simple counting of arguments with absolute proof mapping to infinity and the unit interval could correspond to a number of measures including probability. The terms in each of these dictionaries can be combined in different ways depending on the assumptions one wishes to make about the independence of the different arguments.

There are at least three ways of aggregating quantitatively weighted arguments. One is to consider arguments as based on semantically independent information, so that confidence in a conclusion is obtained by multiplying together all the confidences of the assumptions used to prove it and normalising the resulting values (as in probabilities). Another is to take a ‘weakest link’ approach, where the confidence in the conclusion is equal to the minimum of the confidences in the assumptions (as in fuzzy logic and possibility theory). The confidence in each proposition is the confidence of the strongest argument for it, the confidence in an argument is the minimum of the confidences in the assumptions used to construct that argument. A third method takes the strongest argument but multiplies together all the confidences in the assumptions. A more complex method takes the dependencies between arguments into account.

The approach adopted in this thesis has been based on a development of a version of the Symbolic Decision Procedure which corresponds most closely to that described at the start of this section. Neither the developments summarised in the domino model nor the formalisation in terms of a logic of argumentation are incorporated. Since the goal of this work is a decision support system which can organise the presentation of information from disparate sources, there was no need to consider the problems of scheduling and commitment. An aspect of the work which is of interest in relation to the domino model is that the extension to the decision procedure described later in this chapter could allow the kind of system envisaged in the domino model to update the database of beliefs on the basis of information obtained automatically from images. This topic will be discussed more fully in Section 4.6.

It was decided not to employ a strict version of LA in the present work, because a theorem-prover based on LA has considerable inefficiencies which stem from the need to support a greater variety of ways of constructing arguments – and ways of reasoning about them – than are anticipated in the applications considered here. The overriding advantage of the version of the decision procedure used here is its simplicity – a particular virtue since it means that the operation of the system is relatively transparent. A user simply has to familiarise him or herself with the specification of a decision to understand how candidates are proposed and arguments constructed and can also find out the scope and limitations of the knowledge in the knowledge base.

As a preliminary to the description of the extended decision procedure, the next section considers how radiological images are used in making clinical decisions.

4.3 Decision Making in Radiology

Radiologists are involved in making a wide variety of decisions, both within their own speciality and as part of a team making collegiate decisions. Such decisions might include:

- whether or not to recall a client of a screening programme
- whether or not further investigations are required
- selection of a treatment
- classification of a finding
- diagnosis of an abnormality
- staging of a disease
- assessment of response to treatment

To make each of these decisions a set of candidates must be considered, potential arguments constructed for and against each candidate and required pieces of evidence established. In the first example above, a radiologist inspects a previously unseen image to determine if an abnormality is present. The candidate solutions in this case would include the option of referring the patient, and the arguments would be based on the presence of abnormalities in the image and image data suggestive of an abnormality would provide the evidence required. In the case concerned with the staging of a disease an image is viewed, possibly having been previously interpreted in making a diagnosis, in association with other information to establish whether or not the tumour has certain visible characteristics indicative of its aggressiveness. Here the different stages are the candidate solutions, the staging criteria form the arguments and the characteristics are the items of evidence required.

Different decisions require different ways of interpreting images. I have identified three different interpretation tasks which radiologists have to perform in making decisions: detection, classification and measurement.

- *Detection:* given the class of images and the context in which the examination was requested, a set of possible abnormalities is known. The task of the radiologist is to assess the evidence that any of these abnormalities is present.
- *Classification:* given the context and a set of possible classifications for a particular piece of anatomy or pathology. The task of the radiologist is to determine whether the evidence meets the criteria for one of a set of possible classifications.
- *Measurement:* the radiologist is required to measure a piece of anatomy or pathology, e.g. to estimate the volume or greatest dimension of a tumour.

The notion of ‘task’ employed here is closely related to the idea of *generic tasks* developed by Chandrasekaran [Chandrasekaran1988]. Each of the three tasks is defined in terms of a set of abstract rules, together with sets of facts representing the knowledge employed in the task. The representation of the tasks is sufficiently abstract to be applicable across a range of image interpretation problems. The task image-processing specifications also make use of the notion of a generic sub-task, with component elements being shared between the three tasks.

4.4 The Extended Decision Procedure

I propose a set of rules which together form a model of the process of decision making in image interpretation. The model includes rules specifying the three generic image interpretation tasks which provide the evidence used in making decisions. The function of these rules is twofold: to provide a homogeneous representation of the different decisions and to control the construction of advice from the various information sources. These correspond to the declarative and procedural interpretations of the rules. In the next sub-section the rules describing the processes of candidate proposal and the construction of arguments are explained in both their declarative and procedural interpretations, and the formalism used to present them is introduced. In the following section the rules which describe the three interpretation tasks are explained.

4.4.1 Rules for proposing candidates and constructing arguments

Recall the example given in Section 4.2, where the decision procedure was used to generate two candidate solutions – benign change and breast cancer – given a problem of diagnosis where a women presents with a breast lump. Benign change and breast cancer were proposed because they were identified in the domain knowledge as possible causes of breast lumps and the task-specific knowledge about diagnosis

specified causes as appropriate candidates. The generic rule describing candidate proposal can be stated informally as follows:

Candidate Proposal: the candidates to be suggested include every item in the knowledge base which is linked to the focus (the initial dilemma the decision-maker faces) by one of the appropriate proposal criteria for the kind of decision.

The declarative reading of this rule provides an abstract definition of what constitutes a possible candidate. It can be stated formally, in predicate logic, as follows:

```
If    proposal_criteria( Decision, Criteria )  
and  medical_fact( Criteria, Focus, Candidate )  
Then proposed_candidates( Decision, Focus, Candidate )
```

An upper-case letter at the start of a word indicates a variable, and all variables are universally quantified. The rule states that the proposition in the conclusion, identifying the proposed candidates, is true for all values of Decision, Focus and Candidate for which the two propositions referring to the proposal criteria and the facts are true, so if the task-specific knowledge includes the fact that causes are appropriate proposal criteria for diagnosis and the medical knowledge includes the fact that breast cancer can cause a breast lump, the rule can be used to generate the inference that breast cancer can be proposed as a candidate in the diagnosis of a breast lump. This results in the variable Candidate being instantiated with the value breast cancer.

The next rule governs the construction of arguments. In diagnostic decisions signs and symptoms are used in constructing arguments. The task-specific knowledge will, therefore, record that signs and symptoms are the appropriate criteria for use in constructing arguments. So, if the knowledge base records that 'calcifications are a sign of breast cancer', then this is a possible argument in favour of 'breast cancer'. Abstracting from this, the generic rule is obtained:

Argument Construction: the possible arguments relating to each candidate include every item in the knowledge base which is linked to that candidate by one of the appropriate criteria for this kind of decision.

The argument in this example depends on the detection of calcifications. The calcifications are what is termed a ‘finding’, a fact about the patient for which evidence may be obtained. In the Symbolic Decision Procedure described at the start of this chapter, the evidence for arguments was either provided by the user, in response to a prompt of some kind, or was already available in the patient record. In the extension to the decision procedure described here, evidence is obtained from images. The rules which determine how this is done are called the *Interpretation Rules* and are described in the next sub-section.

Together, the rules for the proposal of candidates, the construction of arguments and the collection of evidence describe how decision support is provided. A general rule linking these three together provides – in its declarative interpretation – a statement of what information is provided as the decision support.

Decision Support: the decision support consists of all of the evidence for each of the findings used in arguments proposed for and against every candidate suggested from the specified focus.

Decision Support: the decision support (the complete set of information provided by the system) consists of all of the evidence for each of the findings used in arguments proposed for and against every candidate suggested from the specified focus.

```

If    proposed_candidates( Decision, Focus, Candidate )
and  argument_for_candidate( Decision, Candidate, Force, Finding )
and  evidence_for_argument( Decision, Patient, Finding, Evidence )
Then decision_support( Decision, Patient, Focus, Candidate, Force,
                      Finding, Evidence )

```

Candidate Proposal: the candidates to be suggested include every item in the knowledge base which is linked to the focus (the initial dilemma the decision-maker faces) by one of the appropriate proposal criteria for this kind of decision.

```

If    proposal_criteria( Decision, Criteria )
and  medical_fact( Criteria, Focus, Candidate )
Then proposed_candidates( Decision, Focus, Candidate )

```

Argument Construction: the possible arguments (with positive or negative force) relating to each candidate include every item in the knowledge base which is linked to a candidate by one of the appropriate criteria for this kind of decision.

```

If    arguments( Decision, Force, ArgumentType )
and  medical_fact( ArgumentType, Candidate, Finding )
Then argument_for_candidate( Decision, Candidate, Force, Finding )

```

Table 2: the top level of the Symbolic Decision Procedure – rules defining what the decision support consists of, how candidates are proposed and how arguments are constructed

In a procedural reading, each rule defines an executable procedure. The conclusion of a rule serves as the ‘procedure call’. In order to generate decision support for a specified patient, decision and focus, the procedure defined in the *Decision Support* rule is called with the first three variables instantiated:

```
decision_support( diagnosis, ms smith, breast lump, Candidate, Force, Finding, Evidence )
```

The rule functions as an instruction to call the other three procedures: *proposed_candidates*, *argument_for_candidate* and *evidence_for_argument*. Since the rule is universally quantified the procedures are called as many times as necessary to return

all possible values. The procedure `proposed_candidates` is called with `Decision` set to `diagnosis` and `Focus` to `breast lump`, and the *Candidate Proposal* rule functions as an instruction to call the procedure `proposal_criteria`, then `fact`. These procedures consist not of rules but of sets of facts, such as those given in Table 3. So `proposed_candidates` returns `breast cancer` and `benign change` as solutions and `argument_for_candidate` is called for each. If the knowledge base contained only the facts given in Table 3, only one argument would be found: `calcifications`.

Task-Specific Knowledge: sets of facts defining decisions to which the decision procedure could be applied.

`proposal_criteria( diagnosis, causes )`
`arguments( diagnosis, supporting, positive signs )`

Domain Knowledge: the set of facts describing the medical domain.

`medical_fact( causes, breast lump, breast cancer )`
`medical_fact( causes, breast lump, benign change )`
`medical_fact( positive signs, breast cancer, calcifications )`

Table 3: the task-specific knowledge and the domain knowledge required for the example of a women presenting with a breast lump

The three rules in this section generate all the relevant findings for a given decision: that is, all the pieces of evidence required by the possible arguments which could be constructed for and against each of the relevant candidates. The rules described in the next section are used to establish whether the available procedures for processing the available images provide evidence that the findings are actually present.

4.4.2 The interpretation rules

In the extension to the decision procedure described here, evidence is obtained from images. The rule for *Decision Support* calls a procedure `evidence_for_argument`,

which returns the evidence for a finding, given the decision and the patient. The evidence is obtained from the image by applying the rules which describe the three interpretation tasks – detection, classification and measurement – but first a fact identifying the appropriate task for a given decision must be consulted. Table 4 contains the definition of the rule which describes how the *Interpretation Rule* is chosen by consulting the facts stating which interpretation task is appropriate for a given decision.

Task Selection: evidence to justify arguments is found by applying the rules defining the appropriate image interpretation task for the decision.

```

If      find_evidence_for_argument( Task, Patient, Decision, Finding,
                                   Evidence )
and    decision_task( Decision, Task )
Then   evidence_for_argument( Decision, Patient, Finding, Evidence )

```

Decision Tasks: the facts recording which interpretation is required for each kind of decision.

```

decision_task( screening, detection )
decision_task( feature classification, classification )
decision_task( risk assessment, measurement )

```

Table 4: the facts that identify which of the three different image interpretation tasks is appropriate for each decision

The three generic image interpretation tasks described in Section 4.3 must be characterised using a set of logical rules, which – in their procedural interpretation – can be used to obtain evidence from images automatically. Consider first the information that must be represented in order to allow symbolic rules to control the application of image-processing operators and use the result of image processing in symbolic reasoning. The various findings depicted in images are represented as symbols in the medical knowledge database and the representation must include a mapping between these symbols and the values returned from the image-processing operators. If the rules are to be used to control the application of image-processing operators, then the representation must include a link between each finding and the image-processing operators. It would be possible to represent this directly: for example

by recording that a given operator is used in the detection of calcifications. However, image-processing operators do not respond to findings, but rather to characteristic properties of the ways in which findings are depicted in images. A clearer representation, therefore, will include a set of facts recording how the required medical findings are depicted in classes of image and which image-processing operators perform appropriate detection, classification and measurement operations for various depictions.

The rule describing detection will therefore use the facts about depiction to identify the depiction of a required finding, then use the facts about detectors to identify the appropriate image-processing detector for the required depiction and finally apply the detector to the appropriate image. The results of applying the operator will be used as the evidence for the existence of the finding.

Detection: the evidence for each argument may include the value returned by a detector applied to the image to detect a feature which depicts the required finding.

Similarly, the rules for classification and measurement will use facts about depiction to identify the depiction of a required finding, and facts about classifiers and measures to identify an appropriate image-processing operator for the classification or measurement of the required depiction and to apply that operator to the appropriate image. The results of applying the operator will be used as the evidence for the existence of the finding.

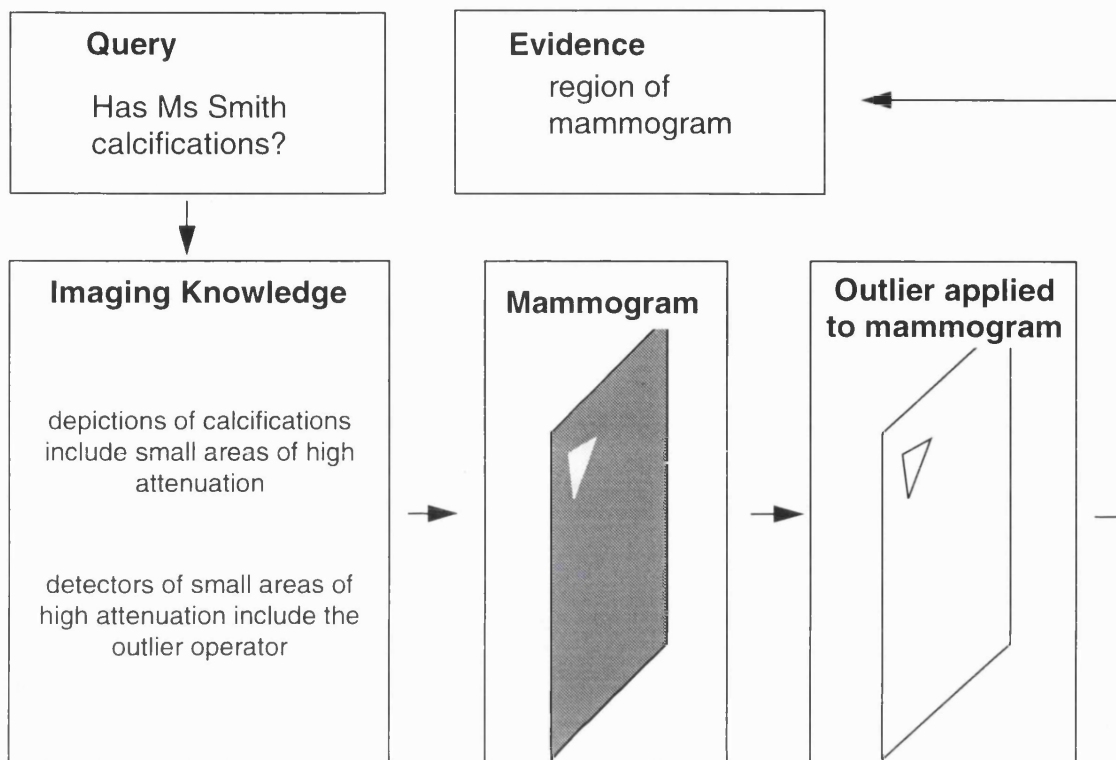


Figure 16: the procedural interpretation of the rule describing detection. In this example, the 'outlier' operator is applied to a mammogram and the region identified by the operator is used as evidence for the existence of calcifications.

The key difference between the detection task, on the one hand, and the measurement and classification tasks, on the other, is that in the classification and measurement tasks a representation of at least part of the relevant anatomy has to be constructed: if you want to measure or classify something, you must have to detect it. A *model* of anatomy is required which lists the structures that must be detected before a particular classification or measurement decision can be taken, as well as information about how the structures are depicted in the image and which image-processing operations can be used to detect these depictions. This permits the required detectors to be applied to the image and results in an *interpretation* of that image. Neither the model of

anatomy nor the resulting interpretation needs to be complete; they need only represent what is required for the particular task.

Region of Interest: the region corresponding to a finding is any region recorded in the interpretation of an image with features which match the depiction of the finding.

```
If    patient_has_images( Patient, ImageType, Images )
and  model( Decision, ImageType, Model )
and  interpretation( Images, Model, Interpretation )
and  depictions( Finding, ImageType, region(Region, FeatureSet ))
and  includes( Interpretation, region(ImageRegion, FeatureSet ))
Then region( Patient, Decision, Finding, region(Region, ImageRegion ))
```

Interpretation: the interpretation of an image according to a model is the list of regions identified by detectors which respond to the depictions of the anatomical features listed in the model.

```
If    first_item_of_list( Structures, Structure, RemainingStructures )
and  first_item_of_list( Regions, Region, RemainingRegions )
and  interpretation_of_item( Images, Structure, Region )
and  interpretation( Images, RemainingStructures, RemainingRegions )
Then interpretation( Images, Structures, Regions )

If    detector_applied_to_image( Image, Detector, Regions, Value )
and  detector( Depiction, Type, Detector )
and  depictions( Item, Type, Depiction )
Then interpretation_of_item( Image, Item, region( Item, Regions ))
```

Table 5: the rules describing the use of a model to identify a region of interest for the classification or measurement tasks. The model contains a list of items. The third rule performs the detection task for an item, the second ensures that this task is performed for every item in a list. The rule Region of Interest takes the list of regions returned when the interpretation of every item identified in the model has been performed, and identifies the region with properties corresponding to the depiction of the finding. This region is then used as the region of interest.

Region of Interest: the region corresponding to a finding is a region recorded in the interpretation of an image with features matching the finding's depiction.

Interpretation: the interpretation of an image according to a model is the list of regions identified by detectors which respond to the depictions of the anatomical features listed in the model.

In a classification task, facts are required which relate findings in the knowledge base to properties of the image, as well as a matching process which ascertains if the values obtained for these properties are suggestive of the presence of the finding.

Classification: the evidence for each argument may include the result of matching the features depicting a finding against the properties of relevant items in the interpretation of an image.

The rule for measurement is analogous except that the finding is defined in terms of a measure and this measure is applied to the interpretation of the image.

Measurement: the evidence for each argument may include the result of measuring a property of an item in the interpretation of an image.

In addition to the three rules describing the image interpretation tasks, the extended decision procedure includes another rule which allows information entered by the radiologist onto the patient record to be used as evidence.

Radiologist's Report: the evidence for each argument may include items recorded by a radiologist.

Detection: the evidence for each argument may include the value returned by a detector applied to the image to detect a feature which depicts the required finding.

```
If    patient_has_images( Patient, ImageType, Image )
and  depictions( Finding, ImageType, Feature )
and  detector( Feature, ImageType, Detector )
and  detector_applied_to_image( Images, Detector, Region, Value )
Then find_evidence_for_argument( detection, Patient, Decision, Finding,
                                [ Detector, Region, Value ] )
```

Classification: the evidence for each argument may include the result of matching the features depicting a finding against the interpretation of an image.

```
If    region( Patient, Decision, Finding, region( Region, ImageRegion) )
and  measure_of_property( Finding, Measure )
and  apply_measure( Images, Measure, ImageRegion, Result )
and  depiction_matches_interpretation( Measure, Result, Value )
Then find_evidence_for_argument( classification, Patient, Decision, Finding,
                                [ region( Region, ImageRegion ), Value ] )
```

Measurement: the evidence for each argument may include the result of measuring a component of the interpretation of an image.

```
If    region( Patient, Decision, Finding, region( Region, ImageRegion ) )
and  measure_of_property( Finding, Measure )
and  apply_measure( Images, Measure, ImageRegion, Value )
Then find_evidence_for_argument( measurement, Patient, Decision, Finding,
                                [ region( Region, ImageRegion ), Value ] )
```

Table 6: the rules describing the three image interpretation tasks

The extended decision procedure, therefore, consists of four elements from the original decision procedure:

- a set of logical rules describing the process of decision structuring and decision making
- a set of facts about different kinds of decision
- a set of facts describing the domain in which the decision is to be made
- a set of facts describing what is known about the particular case

augmented with:

- a set of logical rules describing the three image interpretation tasks
- a set of facts listing the structures to be identified in specific image interpretation tasks
- a set of facts describing how findings and other relevant structures are depicted in medical images
- a set of facts describing how image-processing operations are used to detect certain kinds of depiction

In this section I have introduced rules describing the process of decision structuring and rules describing the generic image interpretation tasks. In the next three sections I illustrate how these rules are applied in three different decisions, and give examples of the additional task specific and domain specific facts which are required.

4.5 Supporting Image Interpretation Tasks

In this section I give examples of how the extended decision procedure outlined above is used to control the application of image processing in a detection, a classification and a measurement task. The examples given are taken from the mammography system, described more fully in Chapter Seven.

4.5.1 Detection of calcifications

A great deal of research in the processing of medical images has been directed at the problem of detecting calcifications on mammograms, see Guillemet for a recent bibliography [Guillemet1997]. The standard approach to this problem is a two-stage process in which the image is first segmented to identify regions which are likely to

contain calcifications, and then measured in various ways to allow a classification which will sort true from false positives.

Recall the rule which described the detection task:

Detection: the evidence for each argument may include the value returned by a detector applied to the image to detect a feature which depicts the required finding.

To apply this rule to the detection of calcifications, for example in deciding whether or not to refer a woman from screening for further investigation, it is necessary to represent the way that calcifications are depicted in image data. This is done by listing the properties of an image region which could correspond to a calcification:

Facts about Depiction: calcifications are depicted as small regions of high relative density.

A detector specialised for small regions of high intensity is required. The segmentor which performed best in Kegelmeyer's study [Kegelmeyer1994] is called the 'outlier' method, so we must represent the fact that:

Rules describing detectors: 'outlier' is a detector for small regions of high relative density.

The definition of the 'outlier' detector is given by Kegelmeyer as:

A local window w around every point p is considered. If the grey-level of p is greater than a threshold defined as the mean of w plus three times the standard deviation of w , the point is considered to be extreme. Extreme points are grouped into segmented calcifications on a four-connected basis.

The 'outlier' operator is applied to the image by calling a function which returns the regions of the image identified by the 'outlier' detector. The outlier operator combines two generic classes of image-processing operation. In the first a window is

moved across an image and an output value calculated for the point at the centre of the window. In the second a binary image is converted into a list of regions corresponding to calcifications. The first operation requires three parameters: the window size, the statistics to be calculated for the window (the value at the centre of the window, the mean and standard deviation) and the output rule (if the value at the centre is greater than the mean plus three times the standard deviation the output is one, otherwise it is zero). The parameters for the second operation specify the class of pixels to consider (those with value one) and a size threshold. The following rule describes the operation of the 'outlier' operator as implemented in the system described in Chapter Seven:

Outlier Operator: the regions identified by 'outlier' are those returned by applying the region finder operator to return all regions of at least eight pixels in the image created by moving a 32-pixel window over the original image, computing the moments and applying output rule one at each stage. Output rule one states that if the value at the centre is greater than the mean plus three times the standard deviation the output is one, otherwise it is zero.

The rule for detection, if invoked as described, would be instantiated as follows:

```

If    patient_has_images( anthea andrews, mammograms, m001)
and  depictions( calcification, mammograms, region([ small, high_attenuation])
and  detector( region([ small, high_attenuation]), mammograms, outlier)
and  detector_applied_to_image( m001, outlier, region(...), 1)
Then find_evidence_for_argument(detection, anthea andrews, screening,
                                calcification, [ outlier, region(...), 1 ])

```

with the list of detected regions returned in the structure region(...).

Detection: the evidence for each argument may include the value returned by a detector applied to the image to detect a feature which depicts the required finding.

```

If    patient_has_images( Patient, ImageType, Image )
and  depictions( Finding, ImageType, Feature )
and  detector( Feature, ImageType, Detector)
and  detector_applied_to_image( Images, Detector, Region, Value )
Then find_evidence_for_argument(detection, Patient, Decision,
                                Finding, [ Detector, Region, Value ])

```

Facts about Depiction: calcifications are depicted on mammograms as small regions of high_attenuation.

```

depictions( calcifications, mammograms,
            region([ small, high_attenuation]))

```

Rules Describing Detectors: outlier is an appropriate detector for regions which show the properties 'small' and 'high_attenuation'.

```

If    includes( Properties, small )
and  includes( Properties, high_attenuation )
Then detector( region( Properties ), mammograms, outlier )

```

Outlier Operator: the regions identified by 'outlier' are those returned by applying the region finder operator to return all regions of at least eight pixels from the image created by moving a 32-pixel window over the original image, computing the moments and applying output rule one at each stage (output rule one states that if the value at the centre is greater than the mean plus three times the standard deviation the output is one, otherwise it is zero).

```

If    moving_window( Images, 32, OutputRule1, NewImages )
and  regions( NewImages, 8, Region, Value )
Then detector_applied_to_image( Images, outlier, Region, Value )

```

Table 7: the rules and facts used to represent the detection task, when employing the 'outlier' operator in detection of calcifications

4.5.2 The classification of calcifications

Calcifications, or apparent calcifications, are caused by a variety of processes, some are malignant, more are benign and a number are merely artifacts. In several recent papers authors have turned their attention to the problem of characterising calcifications, again Guillemet provides an account of this work [Guillemet1997]. The

measures described by these authors are used in the system described in Chapter Seven to provide information which allows a knowledge base of facts about the typical presentation of calcifications to provide decision support.

Recall the rule which describes the classification task:

Classification: the evidence for each argument may include the result of matching the features depicting a finding against the interpretation of an image.

First a representation of the model to be used in the classification task is required. The purpose of the model is to indicate the structures which have to be identified before the classification can be performed. Interpreting the image according to the model is a matter of identifying the image regions which correspond to the structures listed in the model. This involves a sequence of tasks – the sequence is analogous to the detection task – for each structure: the relevant depiction is first identified and then used to select an appropriate detector which can be applied to the image. The result of the interpretation process is a list of image regions associated with each of the structures in the model. This interpretation process is described in the rules in Table 5.

The knowledge base refers to properties such as shape, distribution, location and size. The only property which refers to structures other than the calcifications or their clusters themselves is location. Possible locations referred to in the knowledge base include 'close to nipple', 'inside fibroadenoma' and 'around a fibroadenoma'. The model therefore lists the nipple, calcifications, calcification clusters and fibroadenoma as items to be identified. The model required for the classification of calcifications is:

The Model: to classify calcifications we must identify the calcifications, calcification clusters and the surrounding normal tissue as well as significant landmarks such as the nipple and any benign tumour.

Classification: the evidence for each argument may include the result of matching the features depicting a finding against the interpretation of an image.

```
If    region( Patient, Decision, Finding, region( Region, ImageRegion ))
and  measure_of_property( Finding, Measure )
and  apply_measure( Images, Measure, ImageRegion, Result )
and  depiction_matches_interpretation( Measure, Result, Value )
Then find_evidence_for_argument( classification, Patient, Decision,
                                Finding, [ region( Region, ImageRegion ), Value ] )
```

Facts about image-processing Measures:

```
measure_of_property(round calcifications, high compactness)
```

```
If lower_bound( high compactness, LowerBound )
and greater_than( Result, LowerBound )
Then depiction_matches_interpretation( high compactness, Result, high )
```

The Model: to classify calcifications we must identify the calcifications, calcification clusters and other structures.

```
model( feature classification, mammography, [ calcification, calcifications,
        calcification cluster, nipple, benign tumour ] )
```

Table 8: rules and facts used to represent the classification of calcifications. The regions of interest which correspond to calcifications are first identified by applying the rules for Region of Interest and Interpretation (see Table 5). These rules, in effect, perform the detection task for each of the structures identified in The Model for this decision. The appropriate measure is applied to the regions and the obtained value tested in a classification rule.

This list of regions is then matched with the depiction of the finding. If the required finding is that the calcifications are round and the measurement of round shape is compactness, a rule which tests to see if the regions identified as calcifications have a high compactness is applied (compactness is defined as the ratio of area to perimeter).

4.5.3 The measurement of tissue density

One of the properties of mammograms measured using image-processing techniques is the density of normal tissue. There are a number of reasons for being

interested in this: in order to be able to identify asymmetries [Miller1992], because denser breasts are harder to read and may require a different interpretation protocol [Jackson1993] and because density may be related to risk [Oza1993].

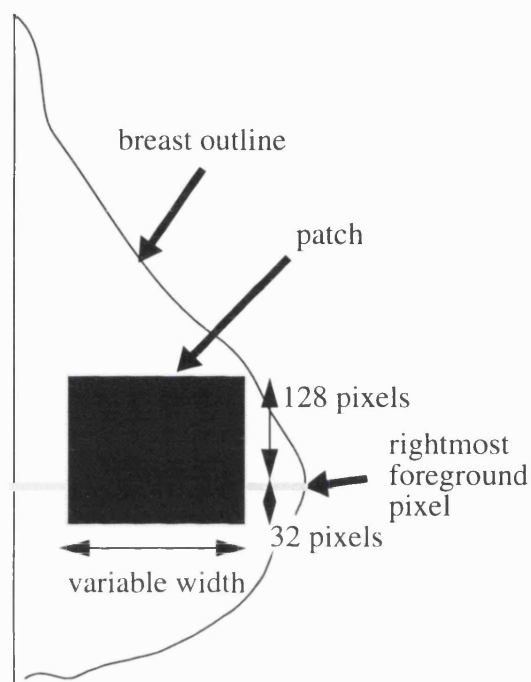


Figure 17: the method used to measure the density of tissue. Thresholding is used to give an estimate of the breast outline and then an automated procedure is used to identify a region of interest behind the nipple. The measure, based on the asymmetry of gray-level histograms, is applied in tiles over the region of interest. The diagram is not to scale.

Elsewhere [Taylor1994] I have presented a method for measuring the density of normal tissue in the breast. This method involves first locating a region of interest above and behind the nipple, where, if the breast is dense, dense tissue will normally be found and then measuring the asymmetry of grey-level histograms in non-overlapping 16x16 pixel windows within this region.

Recall the rule from the decision procedure describing the measurement task:

Measurement: the evidence for each argument may include the result of measuring a component of the interpretation of an image.

The model required in this task distinguishes the breast region from the background. The finding is depicted as the ratio of lighter to dark pixels and measured using an operator which constructs histograms of small regions within the breast area, measures the skewness – the asymmetry of the histograms – and takes the mean of the skew values. We require facts stating the model:

The Model: to assess risk we must identify a region within the breast.

A set of facts is also required, recording how the regions in the model are depicted, identifying detectors for those depictions, representing how ‘tissue density’ is depicted in the region of interest and how that depiction can be measured.

Measurement: the evidence for each argument may include the result of measuring a component of the interpretation of an image.

```
If      region( Patient, Decision, Finding, region( Region, ImageRegion ))
and    measure_of_property( Finding, Measure)
and    apply_measure( Images, Measure, ImageRegion, Value )
Then find_evidence_for_argument( measurement, Patient, Decision,
                                Finding, [ region( Region, ImageRegion ), Value ])
```

Model: to assess risk we must identify a region within the breast.

```
model( assessment of risk, mammography, [ breast region ]).
```

Facts Describing Detectors: the measure used is the mean of local skewness values.

```
measure_of_property( skewness_of_local_histogram( Bound ),
                    mean( local skewness ))
```

Table 9: the rules and facts used to describe the measurement of tissue density. The model identifies the region which must be detected; the measurement of tissue density is performed by applying the appropriate measure in this region.

Two image-processing procedures are applied. First a thresholding operation is applied in the interpretation of the image according to the model. Secondly the skewness operator is applied to measure the tissue density in the region identified in the interpretation.

4.6 Discussion

The purpose of this work is to solve a number of problems associated with the provision of decision support based on information obtained from two distinct information sources and represented in radically different ways: as digital images and as symbolic knowledge. This chapter has described an approach which incorporates information obtained from image processing in a knowledge-based decision support system. The approach is based on a symbolic model of decision making which is used to propose candidate solutions for a decision problem and then to generate arguments for and against them. This model has been extended to allow arguments to be based on the results of image-processing operations carried out on the image under the control of an extension to the decision procedure. This represents a different way of looking at image processing – as a source of information to help in making a decision – and a new framework within which to present the information which image-processing operations can obtain from images.

The model represents the processes of decision making and image interpretation at a very general level. The hope is that this makes possible the implementation of systems which support different decisions and different image interpretation tasks and consider different classes of image while sharing a common structure and consistent interface. As a demonstration of at least a measure of generality, two prototype systems have been developed using the model. These are described in Chapters Seven and Nine.

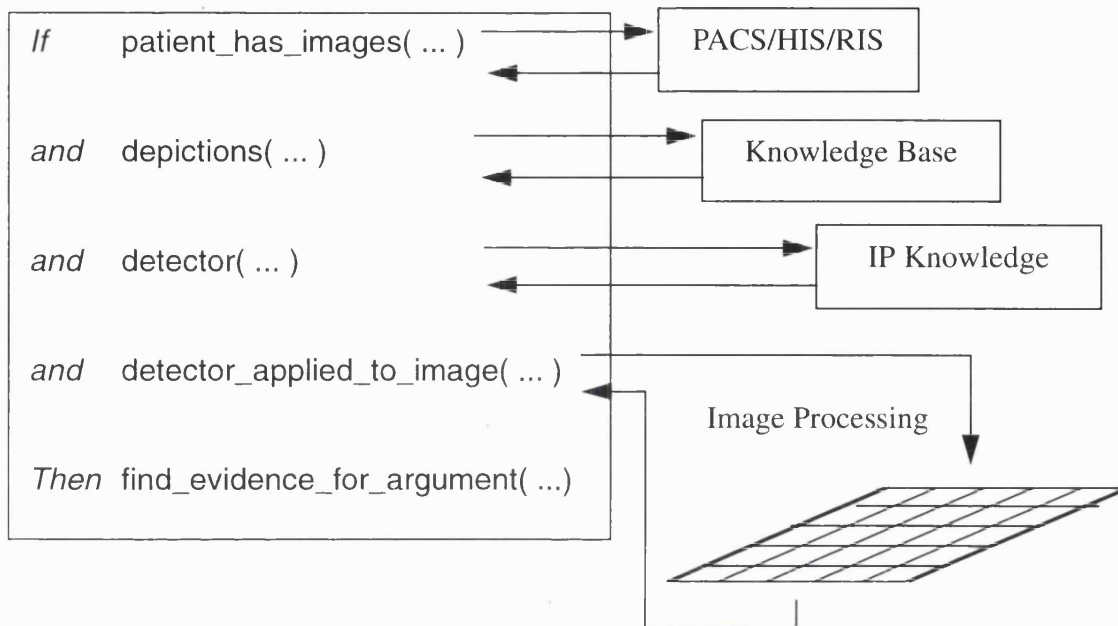


Figure 18: the operation of the rule describing the detection task. The first line represents a query to be addressed to an external information system – perhaps a PACS system of the kind described in Section 1.3 – and the second looks up the required depiction in the knowledge base. The third identifies the appropriate image processing operator. The fourth line invokes the operator, which is applied to the image. The result of the image processing operation then provides the evidence.

One feature of the model outlined that once decision support is requested the image-processing operators are applied automatically by the system. The model is intended to fit within a decision support system and the mode of operation of the system has been assumed to be passive: the system only provides decision support on request. It should, however, be possible to use a model of this kind within an automatic system of the sort described in the account of the domino model in Section 4.2.

Consider how such a system might be used in the measurement of tissue density. One reason for measuring tissue density is the hypothesis that more skilled radiologists perform better at the interpretation of dense mammograms, whereas the

interpretation of fatty mammograms can be adequately performed by a trained radiographer, or another less expensive member of the clinical team. If the protocol followed in screening programmes was based on this hypothesis a decision would have to be made, once the images were taken, as to who would interpret them. The system would have a knowledge base which included a representation of the protocol and would be able to apply the model to obtain the information required for the system to make a decision – the measurement of tissue density. This decision could then be taken automatically, which would then move on to the next activities in the plan and schedule these. This is a potentially interesting application of the model, but one which is radically different from the notion of a decision support system adopted in this thesis, where the emphasis is on assistance for decision-makers rather than autonomous decision making.

One reason for adopting the Symbolic Decision Procedure as a means of combining image processing and knowledge-based reasoning was that the decision procedure allows a flexible approach to the handling of uncertain information. Recall the approach taken in LA where arguments are associated with some measure of their force. This measure could be drawn from any one of a number of dictionaries and different approaches to aggregation could be taken depending on which representation was used and what assumptions could be made about the independence of different arguments. I have used the Symbolic Decision Procedure to organise the presentation of decision support information in the form of evidence for arguments; no facility is provided for combining numbers associated with different arguments, or with different pieces of evidence for an argument.

This is an important area for future research: the integration of measures of uncertainty associated with the application of image processing to digital images. Uncertainty arises through the fallibility of the image processing, the imprecision of the radiological descriptions it seeks to capture, the uncertainties of the image-forming process and the probabilistic association between radiological signs and the underlying

aetiology. Further difficulties stem from the problem of how uncertain information from different imaging modalities, or even different areas of a single image, should be combined.

The prototypes have to deal with uncertainty in a minimal way, however, because the procedure, as presented here, contains no threshold for what constitutes useful information. Thus, in the detection task, the relevant evidence consists of the response of the calcification detector at every point in the image. If this was presented to the user as a set of arguments, the result would be disastrous. This problem is handled by implementing an all-or-nothing calcification detector. In fact, most systems which attempt to detect abnormalities take this approach. Even where the detector returns a value which indicates the likelihood of an abnormality being present, the only information which is passed to the user is whether the obtained measure was above or below some threshold.

In the classification task, an elongated appearance is an argument in favour of the hypothesis of intraductal carcinoma: malignant calcifications. The rule used in the classification task controls the application of an image-processing operator which represents a measure of elongation. Whether or not a particular value for this measure means that a calcification is elongated or not is determined by a simple threshold and no other information about the degree of elongation is passed to the user.

A consequence of using the decision procedure to control the invocation of image processing is that computationally intensive image-processing measures may be invoked, slowing down the system in a way that is unacceptable if it is to be used interactively. The implementation of the extended decision procedure copes with this problem in two ways: by requiring a user who requests decision support to indicate a region of interest and by storing with each image a set of facts representing the results

of expensive computations. This cache of facts is added to as further processing is performed under the control of the decision procedure.

The aim of this chapter has been to describe the extended decision procedure which forms the basis of the approach to combining signal and symbol information. The next chapter describes a generic architecture for decision support systems built using this approach.

CHAPTER FIVE

A GENERIC ARCHITECTURE

The medium is no longer the message.

Nicholas Negroponte

5.1 Introduction

The previous chapter described a logical model of image interpretation which could be used for combining image processing and knowledge-based reasoning in a clinical decision aid. The model consists of a set of logical rules which describe the processes of decision making and image interpretation and sets of facts which characterise the different decisions and tasks and represent knowledge about the domain for which decision support is to be provided, including facts about how abnormalities are depicted in image data and which image-processing operations are required to detect them. These rules can be implemented as a logic program which would form the central element in a system providing decision support drawn from image processing and symbolic knowledge.

The aim of this chapter is to outline a generic architecture for such systems: a modular design for applications in different areas of medicine with different imaging requirements. Certain shared modules could be provided within this and a set of consistent rules laid down for the implementation of the remaining modules. First I describe the set of principles for the design of a radiologist's workstation.

5.2 Three Principles for the Design of a Radiologist's Workstation

This chapter provides a description of the architecture used in the two prototype systems. It is intended not just as a high-level description of the common elements of these two implementations but as a framework which could support a range of analogous systems. This architecture was developed following a number of design principles, which in turn stem from the arguments developed in previous chapters.

5.2.1 Integration of an appropriate range of functions in a single environment

It is argued [Greenes1991, Shortliffe1991, Swett1993] that decision support systems must form integrated components of wider computer systems which can carry out everyday data management tasks thus avoiding the need to re-enter data which is already held elsewhere. Decision support could be provided by a 'clinical workstation': a networked computer providing a consistent interface to software supporting various functions. Meinzer and Engelmann [Meinzer1996] argue that clinical workstations do not require highly sophisticated image analysis functions, but must be able to display images, provide access the full range of clinical data and be integrated with medical protocols. They suggest that the technology to permit this is available, but that there has been no attempt to design such a system.

The generic architecture described here forms an abstract design for a radiologist's workstation which would provide not only decision support but also functions for image reporting, protocol management and information retrieval. These functional components must be supported by a modular design in which elements of the workstation are designed to complement elements provided by other available systems: hospital information systems, electronic patient records and PACS systems. The model provided by the generic architecture includes elements of functionality which, it is expected, could be provided through network connections to such systems.

Design principle: the design must support functions such as decision support image reporting, protocol management and information retrieval. In order to do this it must provide a set of modules which can be connected to other clinical information systems.

5.2.2 Combination of image processing and symbolic reasoning

It was argued in Section 2.4 that, since a range of computerised information sources could be used to provide decision support, there was a case for developing a decision support system which was able to draw on different kinds of information source. The generic architecture described here is intended to allow the implementation of systems which provide decision support based on a combination of image processing and symbolic reasoning. An interface must therefore be defined between the image processing and the decision procedure, and between the knowledge base and the decision procedure, which specifies how information is to be passed between the modules. This interface can be viewed as a query language. In addition to image processing and symbolic reasoning, a number of other information sources could be of assistance to clinicians – resources such as image databases, stores of statistical data and clinical guidelines. Although the model used as the basis for the decision support tool does not draw on all of these, it would be desirable to incorporate them in the architecture, enabling the system to provide a unified environment for accessing different knowledge sources. This requires a design which caters for the different display needs of various kinds of information.

Design principle: for each knowledge source incorporated within the architecture, there must be a definition both of how information is to be displayed to the user and of the form of queries which are supported by the knowledge source.

5.2.3 A task-oriented interface

Meinzer and Engelmann [Meinzer1996] argue that the interface of a clinical workstation must be designed around medical questions. A radiological workstation must be designed to support the tasks performed by radiologists. These tasks include screening, diagnosis, pre-operative investigations and assessments of treatments. Each of these tasks involves an interpretation and making a report, it can also include making, or contributing to, a decision. Ideally a decision support tool provided as part of a radiological workstation will provide support for all the decisions taken in tasks carried out using the workstation. The user must be able to access the decision support while interpreting images, the images must be accessed through an interface which reflects radiologists' conceptions of the tasks to be performed.

Design principle: the design of user interface must reflect the tasks radiologists perform.

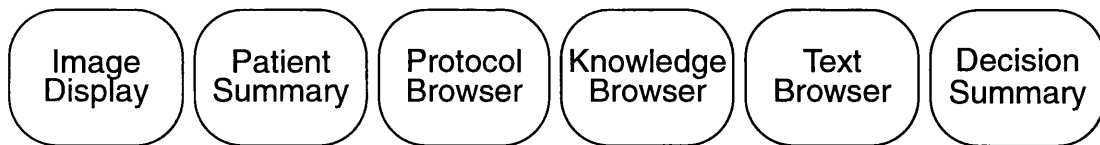
5.3 Components of the Architecture

The components of the system can be divided into three groups:

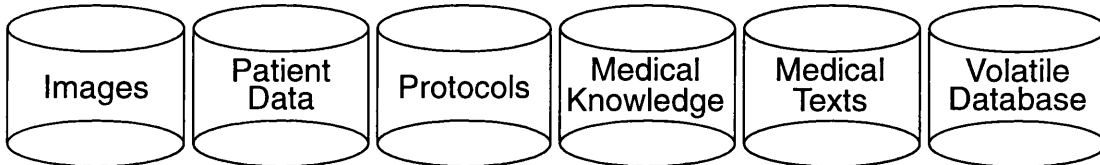
- user interface
- information sources
- information processing modules

These components fit together to provide a range of functions: image reporting, decision support, protocol management and information retrieval.

User Interface



Information Sources



Information Processing Modules



Figure 19: the components of the generic architecture. Rounded boxes represent the different displays which make up the User Interface. Information sources are shown as cylinders while square boxes represent information processing modules.

The modular design of the system reflects the desire to develop a generic architecture; the information sources are distinguished to allow knowledge bases describing other domains and other protocols to be used. The interface is also a separate system, since different imaging modalities have different display requirements. The design also embodies assumptions about how a workstation could be used in practice: the storage and management of patient and image information is handled by a self-contained subsystem which simulates an interface to external information systems – an electronic patient record system, a picture archiving and communication system and a radiology information system.

This section sets out the principles which have guided the design of the prototypes described in Chapters Seven and Nine. The principles constitute a framework which could be used in systems dealing with a range of medical imaging applications. Particular attention is paid to the definition of the interfaces between different modules.

5.3.1 User interface

The user interface of the system consists of a minimal set of controls by means of which the user controls the set of displays which allow him or her to access the information sources and control image processing.

The displays have different requirements, in part because of the variations in the type of information to be displayed and in part because of the different clinical tasks which need to be supported. Basic elements of the interface can be implemented in a way which will be common to all applications, indeed this is necessary if different clinical applications are to be given a consistent appearance. They must however be able to import application-specific functions for image processing and image display. These will include definitions of the display requirements of the different clinical tasks to be supported by the workstation. For example a workstation which is to support a radiologist making staging decisions must include a representation of the staging task which can ensure that the relevant images and patient data are available and can be appropriately displayed when required. These task definitions, which represent a problem-oriented definition of the user interface, are built up from generic functions for the display of images and patient data described in the next sub-sections.

5.3.1.1 Image display

When a user requests that information be displayed about a patient, information about the current clinical task is used to determine which images are displayed, at what

resolution and with which associated functions. Similarly, information about the task is used to create the menus which form the basic tool for reporting on images.

The image display module therefore provides an interface to the patient data which allows the current task to be identified and an interface to the medical knowledge base from which the menus are generated.

5.3.1.2 Patient summary

The display of patient data is organised around the representation of the protocol being followed in the case of a particular patient. It is intended to provide a concise summary of the actions performed or to be performed on the basis of the current plan. It also provides the user interface for the decision support function since, in order to request decision support, the user must first indicate a problem or focus for the decision procedure.

5.3.1.3 Browsers

In order to provide the best possible information retrieval facilities, browsers which allow an intelligible display of the contents of the knowledge bases are included in the architecture. These allow a structured display of the represented relationships between entities in the knowledge base and their properties. The architecture also includes a facility for the display of information stored in text files.

5.3.1.4 Decision summary

The Extended Decision Procedure described in the previous chapter generates assertions about candidates, arguments and the evidence underpinning the different arguments. These are stored in the volatile database and displayed using a decision

summary programme. They can be displayed in different ways according to the preferences of the user. Information about evidence that the decision procedure sought but didn't find can be provided to prompt user input, and a complete list of candidates can be displayed. The supported candidates can be ranked according to the number of arguments found in support of each candidate.

5.3.2 Information sources

Each information source is associated with a set of functions which defines an interface to that source. In the case of images and patient data, these functions provide an interface to programs simulating three external systems: a picture archiving and communication system (PACS), an electronic patient record (EPR) and a radiological information system (RIS).

5.3.2.1 Images

Images are grouped in 'packets' which represent the results of a single radiological investigation. In mammography, for example, the packet consists of a set of up to three views of each breast and might include additional sectional images. Each image is stored with a header containing information about the format. A separate file contains details of the investigation, a link to the patient's record and any radiologists' reports associated with the image packet. Packets are retrieved via the interface to the simulated PACS and radiologists' reports via the interface to the RIS.

Once images have been retrieved, they are processed by the functions which control the display of images in image windows and by another set of interface functions which control the application of image-processing operators and measures. The interface to the radiological report allows it to be recorded that an object exists on the image, and that it has certain properties.

5.3.2.2 Patient data

Each patient's data are retrieved separately from the simulated EPR system. Once the data have been retrieved, the information is accessed via a set of interface functions which allow it to be used in reasoning and updated as required. The definition of this interface requires the data to be recorded in a form which allows retrieval of information about the current protocol, the status of any component plan or action within that protocol, of any recommendation that has been made, any clinical entity which has been recorded (this might be a sign, a symptom or a test result) and any properties it might have. The interface also requires that every action or plan which is recorded on the patient record is associated with a unique identifier which allows the system to identify the protocol within which it was performed.

Plans and actions are recorded on the patient record as having a particular state. The states which are defined for actions are:

- *requested*: the next action to be performed
- *pending*: an action in the current plan which can't be requested until some earlier plan or action is performed
- *abandoned*: an action which is in the plan but will not be performed
- *performed*: an action which has been performed

The states which are defined for plans are:

- *relevant*: a plan which is part of the protocol but has not yet started and which the clinician could decide not to perform
- *pending*: a sub-plan of the current plan which can't be started until some earlier plan or action is performed, but which will normally be performed

- *started*: a plan whose first component is either a plan which has been started or a requested action which has neither been abandoned nor performed
- *abandoned*: a plan which is a part of the current plan but which will not be performed
- *performed*: a plan whose components have all been abandoned or performed

5.3.2.3 Protocols

A clinical protocol is a set of plans and actions. Plans are composed of decisions and other plans and actions, while actions are atomic. The protocol representation includes constraints on the order in which plans and actions can be performed and information about any indications or contraindications for a plan or action. It also includes facts recording which plans and actions are potentially alternatives and what recommendations could be made following each investigation. A diagrammatic representation of the protocol used in the mammography prototype is included in the appendices.

Knowledge about a protocol to be followed in the management of patients is stored in a distinct knowledge base. The interface to this information source supports three distinct kinds of query: is a fact true of a currently active part of the protocol? Is there a unique solution to a particular query? Is a fact true of any part of the protocol? The first query is used in updating the patient record when new information is provided, as is the second kind of query which is required to establish where decisions between alternatives are required. The third form of query is used only in browsing.

The representation is intended to meet the requirements of the decision aid, therefore information is represented at a level of detail which can be conveniently summarised on the display of patient data. This representation of the protocol also

determines the task definitions which form part of the user interface. A discussion of how the specific protocols are represented is included in Chapters Seven and Nine.

5.3.2.4 Medical knowledge

The medical knowledge which is used for decision support is represented in a knowledge base whose interface function allows two kinds of query: *Is there a relationship A between entity B and entity C?* and *Is there a relationship A between B and a C with a value D for property E?* For example:

```
enquiry( medical_knowledge, known( causes, calcifications, breast cancer )  
enquiry( medical_knowledge, known( causes, calcifications::shape=round,  
                                   benign breast change )
```

Knowledge about a domain must include all the entities relevant to the decisions to be tackled by the system. The task-specific knowledge represented in the extended decision procedure, described in the previous chapter identifies the relations between entities used to generate of candidates and construct arguments. Each of these relations must be represented in the knowledge base. The knowledge base also includes the information required to construct menus that allow users to enter reports of images.

The knowledge base must depict the domain in terms of entities which reflect the shared conceptualisation of users. This not a question for the generic architecture since it is clearly dependent on the nature of the domain for which decision support is to be provided. Discussion of how knowledge bases were developed for both the prototypes is included in Chapters Seven and Nine.

5.3.2.5 Texts

The medical texts, which are made available through the system as an additional information source, are stored as distinct text files. No indexing information

is associated with them and there are no links between words mentioned in the texts and terms used in the knowledge base or in patient data.

5.3.2.6 Volatile database

The volatile database is the term given to the store which contains information generated in the course of a session with the workstation which is not to be recorded on the patient record. For example, the information generated by the decision procedure is stored in a set of assertions which is not retained permanently. A new set of assertions is created every time the decision procedure is invoked with a new problem. The decision procedure works, as described in previous chapters, by first proposing candidates from a given focus, constructing arguments for and against, and then looking for the evidence required by the various arguments. Supported candidates are used to propose further candidates, normally sub-types or refinements of the initial candidate, and the process is repeated. The volatile database is used to record the initial candidate and the refinements. Where the evidence required for an argument is found, a fact recording the argument and noting the evidence is recorded in the volatile database. Where the required evidence is not found, the argument is not recorded but the system notes the existence of an 'other sign', a relevant piece of information which is not currently known.

For each candidate the system records the type of decision, the starting-point of the decision and the link which connected the candidate to the starting-point. For each argument the system records the type of decision, the kind of argument (confirming, supporting, opposing or excluding), the candidate and the evidence.

Whenever an image-processing operation is performed, the result – represented symbolically – is also stored in the volatile database together with a unique identifier for any new image created.

5.3.3 Information processing components

The third element of the architecture is the set of components which draws on the different information sources and generates new information, including that used to provide decision support.

5.3.3.1 Image-processing routines

Different applications will require different image-processing operations. The generic architecture, however, defines a consistent interface and specifies the functionality required to support the extended decision procedure. Image-processing routines are called via a pair of interface functions, one for the application of detectors or segmentors and one for the application of measures. The interface function for detectors takes as arguments identifier of the input image and the image-processing operator and returns the identifier of the output image and a representation of the detected region. The representation of the detected region may be a list of vertices, a list of sub-regions or a single sub-region. The interface function for measurement takes as arguments the identifier of the input image, a representation of a region and the name of a measure and returns the result of applying a measure to the appropriate region of the appropriate image.

The image-processing operators are applied by the extended decision procedure using information about images and operators, as described in the previous section. This ‘imaging knowledge’ includes facts about how normal and abnormal structures are depicted in images, what structures must be detected in a given image interpretation task and which image-processing operations are appropriate for the detection of various depictions and the measurement of various properties.

5.3.3.2 Task manager

The task manager consists of a set of state-transition rules, analogous to those making up the decision procedure, which define the conditions under which the states of the actions and plans making up the protocol should be updated: for example, when a plan is performed and whether actions are to be labelled as 'requested' or 'pending'. These rules are invoked whenever new information is entered on the patient record and provide a mechanism for automatically updating information about the protocol being followed in the case of a particular patient.

The rules which apply to a pending action, for example, use the scheduling constraints represented in the protocol to check whether there are actions which must be performed before this action can be requested. If there are none the state of the action is altered from pending to requested. Some state transitions require input from a user and cannot be updated automatically. For example, if a plan is dependent on a decision, then a user must explicitly record that it has been decided to perform the plan.

5.3.3.3 Decision procedure

The decision procedure contains a declarative definition of the rules making up the extended model of decision making, which is described in the previous chapter. The rules consult knowledge about the kind of decision being made, the domain, and the patient. The decision procedure therefore requires an interface to the decision specifications, the medical knowledge base, the patient record and the imaging knowledge.

Decision specifications describe the characteristics of the different decision tasks which are required by the generic decision procedure. Each decision specification indicates the nature of the focus, the criteria which can be used to propose candidate solutions and the criteria which are used in the construction of arguments.

5.4 Functional Components

The above components together support a range of functions including: image reporting, decision support, protocol management and information retrieval. In the following diagrams I explain how the architecture works to provide these functions. Each diagram shows the subset of components from Figure 19 employed in a functional component. The arrows represent the flow of information

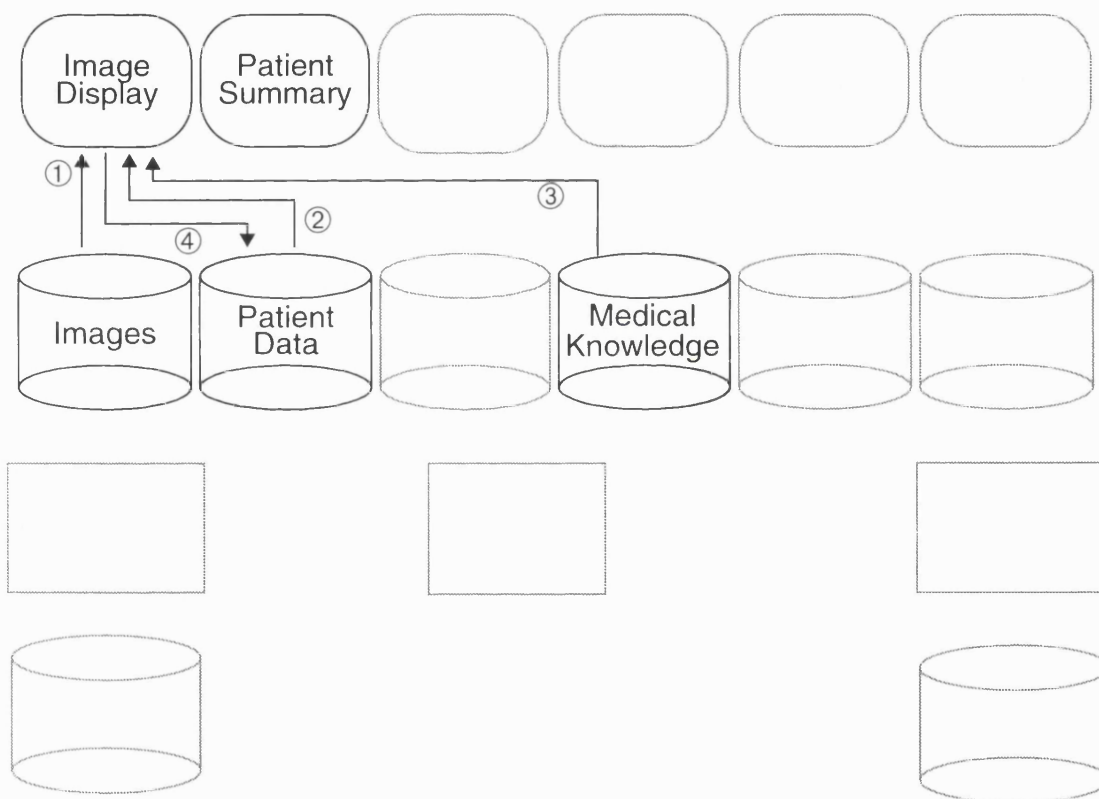


Figure 20: the image reporting function. When an image is displayed ①, the nature of the interpretation task is ascertained from the patient's record ②. The facts in the medical knowledge ③ appropriate to this task are then used to generate a set of menus which are displayed with the image. The user may request image processing, decision support or use the information retrieval function to help interpret the image, before using the menus to select the terms to be entered on the radiologist's report ④.

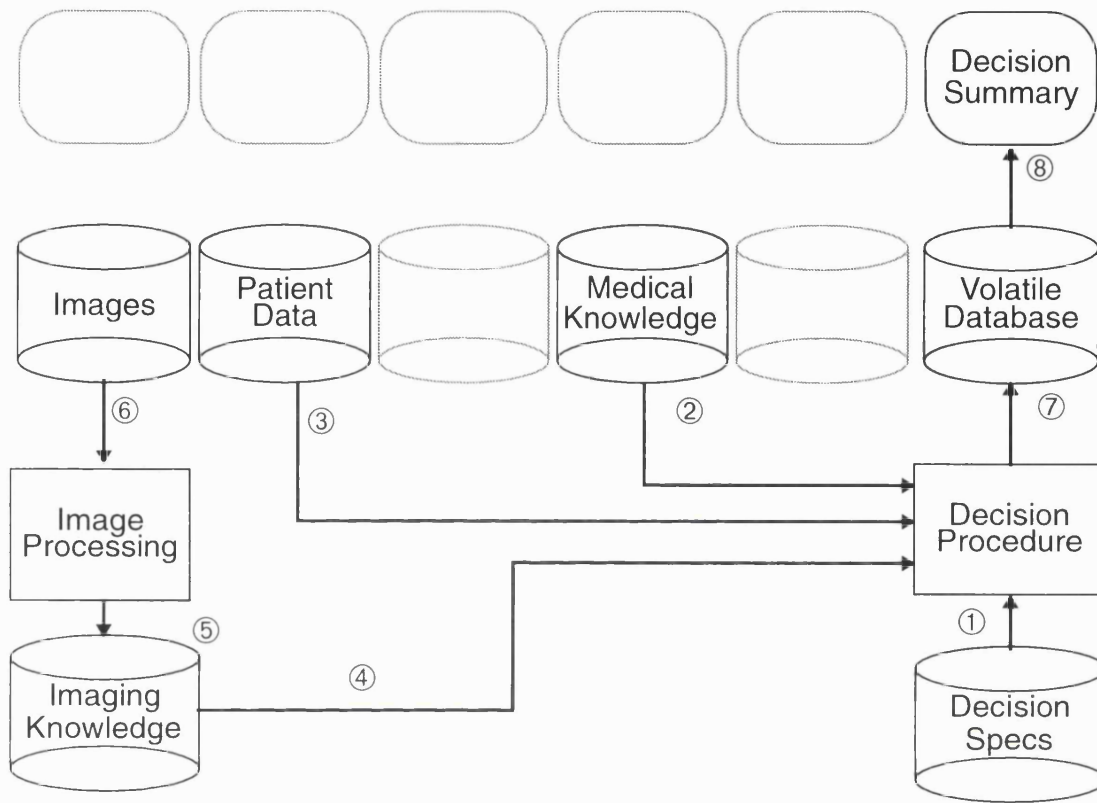


Figure 21: the decision support function. As described in the previous chapter, the decision procedure uses the information in the decision specifications ① to determine which of the entities represented in the medical knowledge base can be used as candidates and arguments, and these are then generated from the medical knowledge ②. The evidence required by arguments is sought on the patient record ③, on the report entered by the radiologist and also, using the information in the imaging knowledge base ④ and the image-processing operators ⑤, on the images ⑥. The results of the decision procedure are stored in the volatile database ⑦ and used to generate the decision summary ⑧.

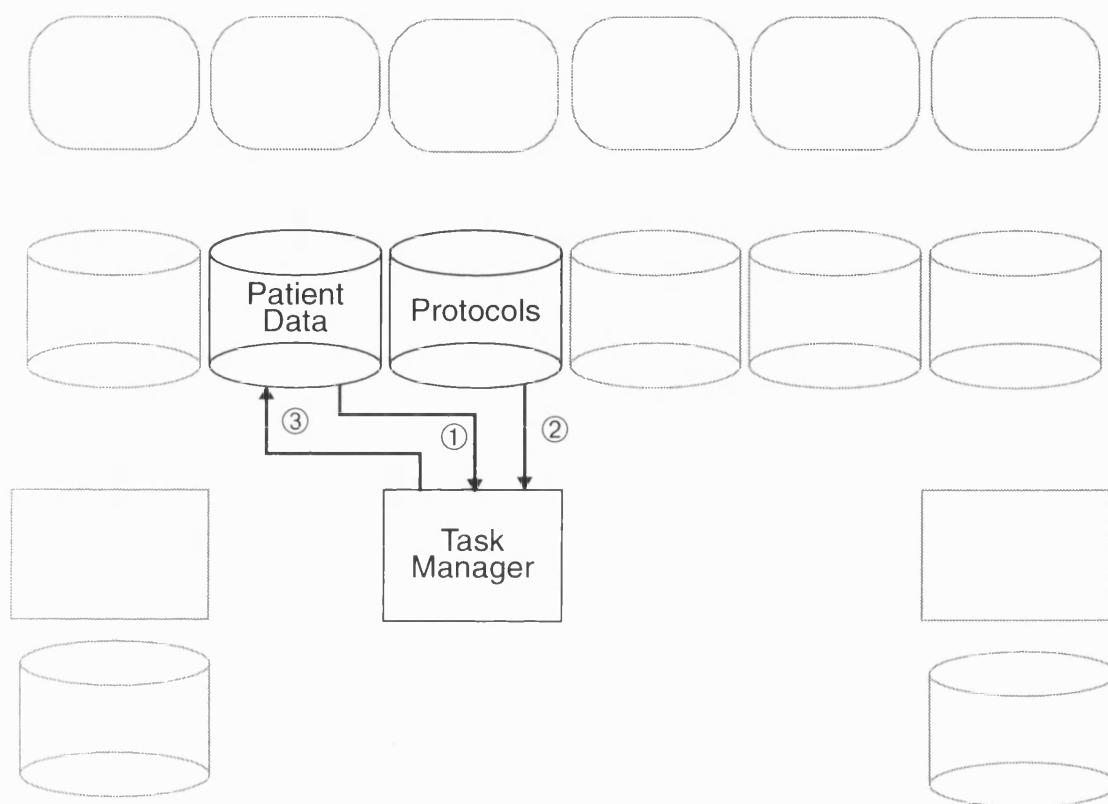


Figure 22: the protocol management function. When new information is entered as patient data ①, the task manager uses information from the representation of the protocol ② to update the representation of the patient record ③.

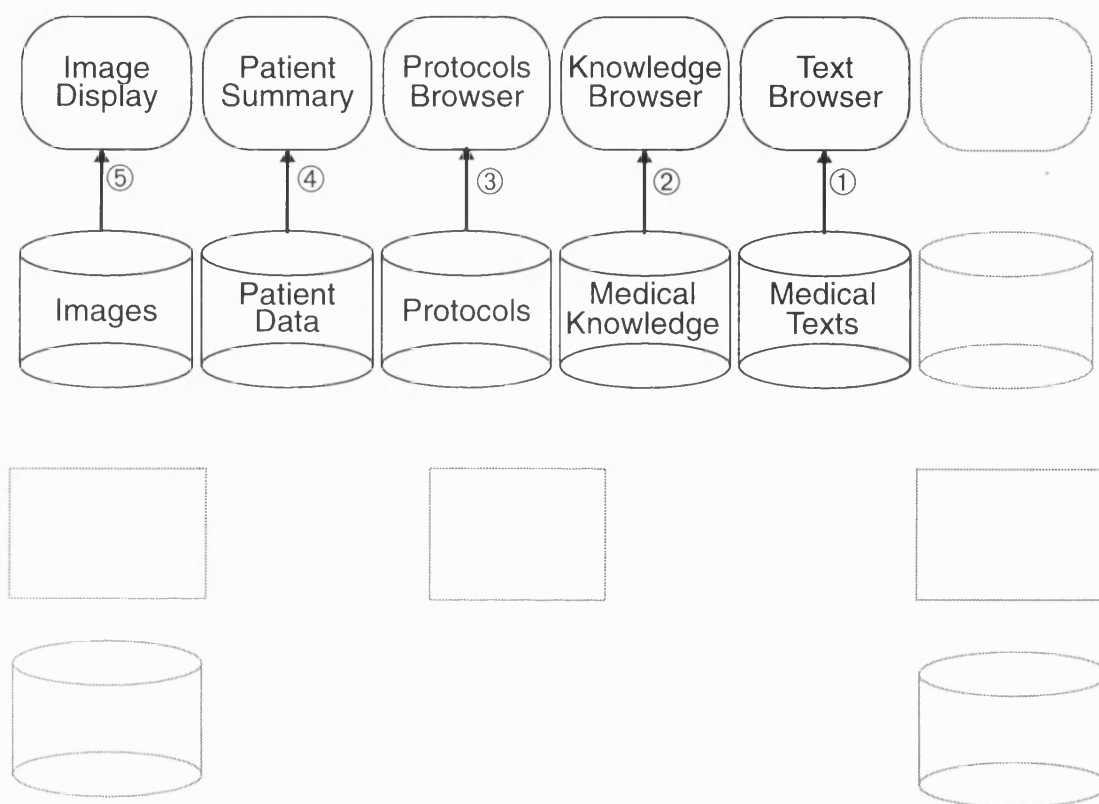


Figure 23: it is intended that the system should function as an information retrieval system, providing a uniform interface to a variety of information sources: medical text ①, medical knowledge bases ②, computer representations of protocols ③, patient data ④ and medical images ⑤.

5.5 Conclusion

An architecture within which knowledge-based decision support could be combined with image processing has been presented. This architecture provides a generic design for systems to support image interpretation within a computing environment which supports an electronic patient record, a picture archiving and communication system and a radiological information system. The belief is that the architecture offers a practical model for decision support systems which provide a

range of useful functions for radiologists. The model was used in the implementation of two prototype systems, which are described in detail in Chapters Seven and Nine.

The architecture is based on the principles outlined in Section 5.2, and on the argument presented in the conclusion to Chapter Two, that a decision support system for image interpretation could be based on a number of information sources, that it could use information represented in a symbolic knowledge base, in digital data or in other forms. Additional information sources could be added, for example to provide the kind of information employed in the numerical decision aids described in Section 2.3.2. The user interface, through which these functions are accessed, must link certain of these functions together, to provide the user with a way of accessing related information from different sources, so that a patient's record can be displayed at the same time as a set of images for example.

The interface functions through which the information processing modules access the different knowledge sources provide a high-level description of the information requirement of the different modules, specifying the range of queries which the information sources must support. Of particular interest is the interface between the image-processing functions and the images. Any image-processing operator in the system must be designed to be called by the extended decision procedure on the basis of the information contained in the imaging knowledge and must return information in the form required by the decision procedure. The information returned by the operator is either a measure, or an image region, or a set of image regions. An image region can be represented either as a set of vertices or as a set of pixels with value 1 in a binary image or as a set of images.

It is worth noting that if the system were to be used in a setting where, for example, mammograms from different X-ray machines or different digitisers were being analysed, then the images would have to be normalised or the image-processing

operators used would have to be able to adjust to cope with variations in resolution and quality. Note that the system does allow the results of computationally expensive image-processing operations to be stored with the image, this facility could also be used to store appropriate parameter values.

The three information processing modules provided include two required to provide support for the model of decision support described in Section 4.4: the decision procedure and image-processing modules. The third module, the task manager, is provided as an example of the kind of additional facility which could be incorporated into the architecture in order to support a range of clinical activities within a single system.

CHAPTER SIX

DECISION SUPPORT FOR MAMMOGRAPHY

6.1 Introduction

In this chapter I describe the work which informed the development of a prototype decision aid for mammography. First I provide an introduction to the subject of breast cancer and mammography. In Section 6.3 I discuss the decisions which radiologists make, or to which radiologists contribute, on the basis of their interpretations of mammograms. This material provides the basis for the knowledge base in the decision aid and is therefore surveyed in some detail. In Section 6.4 I review attempts to use image processing in the detection and classification of microcalcifications. This, again rather detailed, survey is used in the development of the image-processing operators for the decision aid.

6.2 The Medical Context

The female breast consists of a mixture of fatty tissue and secreting glandular tissue. The glandular tissue is divided into 15 to 20 lobes each containing hundreds of

lobules, connected by ductules which join together close to the nipple to produce the major lactiferous ducts, forming seven or so duct openings at the nipple. The breast is held between two thin layers of fibrous tissue, one under the skin and one over the chest muscle, which are connected by thin fibrous ligaments known as Cooper's ligaments. These connective tissues, or parenchyma, are the essence of the breast's radiopacity, and it is the fact that they shrink and wither with age that allows mammography to be used as a diagnostic tool in older women.

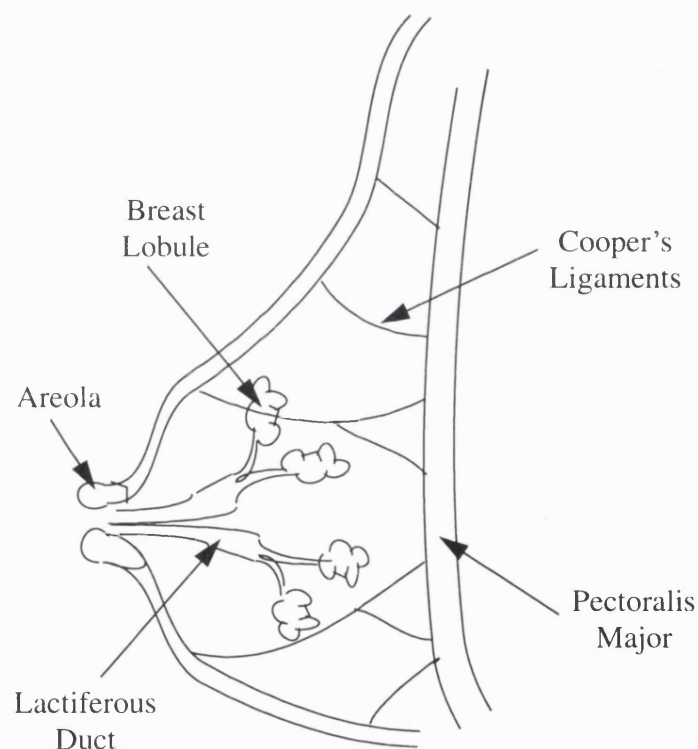


Figure 24: the anatomy of the female breast

Breast cancer can be lobular or ductal in origin and is termed 'in situ' or invasive depending on whether or not it has extended into neighbouring tissue. In addition to this four-way classification some sub-types of invasive ductal carcinoma are occasionally identified: mucinous, medullary and tubular. Cancer also, albeit rarely,

occurs in the breast as a result of metastasis from some other site, as a cancer of connective tissue (breast sarcoma) or within a benign tumour of the breast.

The preferred investigation for the detection and management of breast cancer is mammography. Programmes of regular mammographic screening are being adopted: since 1992 British women between 50 and 65 have been invited to attend for a mammogram every three years. Originally a single image (the 'mediolateral oblique' view) was taken of each breast, but now a second image is recommended (the 'cranio-caudal' view). With over a million women attending every year, the programme generates an enormous demand for image interpretation expertise. Radiologists must search every image for signs of cancer and decide whether the woman should be recalled. The number of images involved is such that inspection must be rapid and the decision must take into account not only the risks associated with failing to identify subtle lesions but also the consequences of recalling healthy women unnecessarily.

The screening programme has, not surprisingly, been the focus for the vast majority of the work on computer aids for mammography. However for women who are recalled the screening mammogram is only the first step in a process of diagnosis and treatment, throughout which the preferred investigation will normally be mammography. Mammography is also used in the investigation and management of breast cancer patients who present outside of the screening programme. The role of the investigation and the nature of the interpretation task is rather different for these patients: the expected incidence is different, the cost-benefit ratio of incorrect decisions changes and a greater variety of future courses of action is available.

A mammogram is a 2D image of the radio-absorbent properties of the tissue types in the compressed breast, normally recorded on a 16x24cm film. Fibrous and glandular tissues appear as white linear arrangements – or, in very dense breasts as a dense grey or white smudge – against a black background of radiolucent fat. Where

calcium has been precipitated it appears as small bright traces. The radiologist must search the image for five signs.

- *Breast masses*: this term is used to describe a variety of features, some of which have a benign interpretation. A breast mass can appear as a radiolucent shadow, an area of focalised density, a radio-opaque capsule or a stellate lesion which may not have a discernable centre. Terms such as circumscribed lesions, stellate or spiculated lesions and focal densities are used to describe breast masses of different types.
- *Microcalcifications*: small flecks of calcium which are caused by a variety of processes including normal change and benign disease as well as cancer
- *Asymmetries*: any asymmetry between the breasts is potentially, though not necessarily, a sign of malignancy, as is any change over time.
- *Parenchymal or architectural distortions*: sometimes the presence of a cancer is revealed through the way it distorts surrounding tissue.
- *Skin thickening*: it is possible to detect the skin on a mammogram and cancer can cause a thickening of the skin.

6.3 Decision Making in Mammography

The literature on decision making in mammography focuses on two issues: the differential diagnosis of breast masses and of microcalcifications. This literature is reviewed in the next two sub-sections. The potential for decision support in mammography is then assessed in a third sub-section.

6.3.1 The analysis of breast masses

Tabar and Dean's *Teaching Atlas of Mammography* [Tabar1983] is a classic description of mammographic signs. They consider breast masses under two separate headings: circumscribed lesions and stellate lesions. Circumscribed lesions can be sharply or poorly outlined; circular, oval or lobulated and can be solitary or multiple. Two signs are of primary importance in distinguishing benign tumours, such as cysts and fibroadenomas, from cancer: the presence or absence of a halo or capsule at the margins of the lesion and the density of the lesion. Two signs are of secondary importance: the orientation and form of the lesion and its size. These can help confirm a diagnosis which should have already been made on the basis of contour and density.

Sign	Cause
halo present	<i>fibroadenoma</i>
capsule present	<i>cyst, intracystic carcinoma, papillary carcinoma, carcinoma in fibroadenoma</i>
radiolucent density	<i>lipoma, oil cyst, galactoceles</i>
radiolucent and radiopaque density	<i>fibro-adeno-lipoma, galactoceles, lymph node, haematoma</i>
low density radiopaque	<i>fibroadenoma, cyst, giant fibroadenoma, atheroma, oil cyst, cavernous hemangioma, papilloma, papillomatosis, wart, abscess, cystosarcoma phylloides, papillary carcinoma, mucinous carcinoma</i>
high density radiopaque	<i>medullary carcinoma, ductal carcinoma in situ, invasive ductal carcinoma, breast sarcoma, metastases to the breast, cystosarcoma phylloides, abscess, cyst, haematoma, enlarged lymph nodes, sebaceous cyst</i>

Table 10: a classification of circumscribed lesions on the basis of two signs identified by Tabar and Dean as being of primary importance – the presence or absence of a halo or capsule and the density. Benign lesions are shown in italics.

Tabar and Dean characterise stellate lesions as radiating structures with ill-defined borders, and add that they can be very difficult to perceive, especially when small. They identify two key properties of stellate lesions: the presence or absence of a

central mass and two basic patterns of radiating spicules – sharp dense lines radiating in all directions or fine spicules bunched together. There are four possible diagnoses.

- *Invasive ductal carcinoma*: a distinct central mass with spicules radiating in all directions. The spicule length increases with tumour size and spicules may reach the skin causing retraction and localized thickening, which is often associated with microcalcifications. With small invasive ductal carcinomas the central mass may be imperceptible and spicules may form a lacy pattern only detectable through parenchymal distortion.
- *Radial scar*¹ (*sclerosing duct hyperplasia*) varies in appearance from projection to projection; there is no solid dense mass but translucent circular areas, the long spicules are very long, close to the centre spicules are bunched in groups, there are often radiolucent linear structures parallel to spicules, and even large lesions are not palpable.
- *Fat necrosis*²; there is no distinct mass but radiolucent areas in a loose reticular structure, the radiating structure varies with the projection, the spicules are fine and of low density and there may be localised skin thickening.
- *Hyalinized fibroadenoma with fibrosis*³ the pattern of spiculation changes with the projection.

A more recent review of research into the characterisation of breast masses [Feig1992] suggests criteria to be used in establishing if a lesion is really present, criteria to be used in establishing if a stellate lesion is malignant and to determine the management of a circumscribed lesion. Feig describes four possible artifacts which can

1. The term 'radial scar' is used to describe a lesion which is not, in fact, a scar but a pathological condition in which ductal elements radiate from a sclerotic focus.

2. Trauma, injury or medical intervention can result in the death or 'necrosis' of fat cells.

3. The retraction of tissue around a degenerating benign lesion causes retraction of the surrounding tissue which simulates a stellate lesion.

create the appearance of breast masses: normal dense tissue, skin lesions, nipple shadow and the sternal injection of pectoralis muscle.

- *Normal breast tissue density (summation shadow)*: this is not seen on all views, or at least not seen to be 3D; it is scalloped, not homogeneous; and there is a gradual change in density at the edge.
- *Skin lesion*: the density change at the margin may be sharp and gradual on opposite sides, it may have a cauliflower shape; and the location will correspond to skin.
- *Nipple shadow*: the location corresponds to the nipple.
- *Sternal injection of pectoralis muscle*: an elongated triangular shape located against chest wall and only visible on the craniocaudal view.

Feig describes the characteristics of 11 benign and four malignant tumours which can present as circumscribed lesions.

Benign:

- *Fibroadenoma* is the most common benign lesion, it has a smooth outline (although the margins may be partially obscured) with a round oval or lobulated shape. They are frequently multiple or bilateral and can be up to 30mm in size. They rarely grow or develop after menopause.
- *Cysts* are round or oval in shape with well defined margins, although these can be partially obscured. They are often multiple and bilateral and less common after menopause. They show rapid growth and rapid regression.
- *Intramammary lymph nodes* are characterised by a central lucency. They are typically less than 10mm in size.
- *Haematoma* is characteristically round or oval with smooth borders containing areas of fatty and glandular density in a fibrotic pseudo-capsule

The surrounding tissue is compressed. It can present with ill-defined borders and smaller lesions are less likely to have the characteristic appearance

- *Lipoma* is totally fatty and contained in a fibrotic capsule 1mm thick. It can be identified by the splaying of adjacent breast trabeculae.
- *Sclerosing adenosis* can appear as a stellate lesion, an irregular density, a circumscribed mass or a poorly defined tumour-like density.
- *Papilloma* usually occurs behind the nipple, is less than several mm in size and may be elongated or have a nodular shape.
- *Haematoma* can appear as circumscribed or poorly defined masses, may be dense or have areas of fatty and dense tissue.
- *Galactoceles* may have a centre with areas of fatty and of glandular density with a clearly demarcated fat/fluid level, or one which is totally lucent with a thin capsule.
- *Myxoblastoma* is round or oval with short spiculated extensions.
- *Giant fibroadenoma* has a smooth outline and is clearly defined with a round or oval shape. It is characteristically large and rapidly growing.

Malignant:

- *Circumscribed carcinomas* are, in Feig's terminology, ductal carcinomas presenting as a breast mass. Most have borders which are at least partially poorly defined. Papillary carcinoma appears as a cluster of nodules, or a single nodule; medullary carcinoma is characteristically lobulated.
- *Sarcoma* is well defined, or at least has minimal irregularity. Rapid growth is a feature.
- *Secondary involvement of lymphoma* is described as a diffusely increased density associated with skin thickening. Moderate or marked spiculation is not infrequent.

- *Metastases to the breast* present as discrete, usually solitary nodules which are mainly found in the upper outer quadrant. Only 25% of cases are bilateral.

Frequency	Stellate	Poorly defined/ irregular	Relatively well defined
Common	<i>postsurgical scar</i> invasive ductal carcinoma tubular carcinoma invasive lobular carcinoma <i>tat necrosis</i>	circumscribed carcinoma: (invasive ductal medullary, mucinous, papillary), invasive lobular carcinoma <i>fat necrosis</i>	<i>fibroadenoma</i> <i>cyst</i> <i>intramammary lymph node</i> ductal carcinoma in situ <i>haematoma</i> <i>lipoma</i> <i>sclerosing adenosis/focal</i> <i>fibrous disease</i> <i>lipid cyst</i>
Infrequent	<i>radial scar</i> <i>sclerosing adenosis</i> ductal carcinoma in situ	<i>sclerosing adenosis/focal</i> <i>fibrous disease</i> <i>haematoma</i> <i>abscess</i> <i>plasma cell mastitis</i> lymphoma	circumscribed carcinoma (ductal in situ, invasive ductal, medullary, colloid, papillary carcinoma) <i>fat necrosis</i> <i>plasma cell mastitis</i> <i>intraductal papilloma</i> <i>haematoma</i> metastases to the breast <i>cystosarcoma phyllodes/</i> <i>giant fibroadenoma</i>
Rare	lymphoma <i>extra-abdominal desmoid</i>	primary sarcoma <i>extra-abdominal desmoid</i> <i>granular cell myoblastoma</i>	<i>galactocele</i> intracystic papillary carcinoma lymphoma sarcoma

Table 11: spectrum of Breast Mass Appearances, adapted from [Feig1992]. Benign lesions are shown in italics.

In order to evaluate circumscribed lesions Feig suggests the analysis of nine characteristics which can distinguish benign from malignant lesions: fat, shape, margins, the presence of calcifications, multiplicity, change in size, appearance on ultrasound, mass density and the presence or absence of a halo.

Seven different causes of stellate lesions are described.

- *Invasive ductal carcinoma* presenting as spiculations radiating from a point.

- *Invasive lobular carcinoma* is radiographically indistinguishable from invasive ductal carcinoma but harder to detect because spiculations are of low density and may not be visible.
- *Post-surgical scar* can cause a mass-like density which will normally regress, characteristically it has no central density but a central lucency may be present due to fat necrosis. On a comparison of views it will be seen to have a planar configuration corresponding to the excisional plane.
- *Fat necrosis* may cause a central lucency with a thin fibrotic capsule
- *Breast abscess* normally occurs in young primiparous women during lactation and may present as a spiculated mass, normally behind the nipple. There will be associated swelling, pain and erythema.
- *Radial scar* is a term for a lesion 10-15mm in diameter, normally with no centre or a radiolucent core, which tends to occur in one plane.
- *Sclerosing adenosis* can appear as a small stellate tumour.

6.3.2 The classification of microcalcifications

An early attempt to classify microcalcifications, which is still widely used in France [Dilhuydy1994], was developed by Le Gal et al. [Le Gal1973] and considers five different morphological types, each associated with a different degree of risk.

- *Type I*: crescent shapes and round, regular discs with clear centres (0% risk).
- *Type II*: round, with regular outline but without clear centres (19% cancer).
- *Type III*: very fine microcalcifications, too small to have an identifiable shape (39% cancer).
- *Type IV*: polyhedral, irregular shapes, similar to grains of salt (59% cancer).
- *Type V*: elongate or branching microcalcifications (96% cancer).

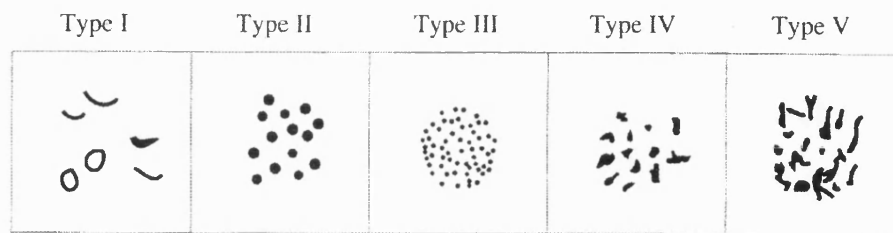


Figure 25: the five types of microcalcifications identified by Le Gal et al. [Le Gal1973]

Dilhuydy [Dilhuydy1994] suggests that the distinctions between types I and II and between IV and V may not be made consistently by different observers and that in practice the important question is whether microcalcifications are of a type where surveillance is possible (I or II) or where biopsy is required immediately (IV or V).

Tabar and Dean [Tabar1983] describe three categories of calcification:

- *ductal (or 'malignant type') calcifications*
- *lobular calcifications*: caused by a variety of processes: sclerosing adenosis, atypical lobular hyperplasia, cystic hyperplasia and blunt duct adenosis
- *miscellaneous calcifications of various types*: arterial, periductal mastitis, sebaceous cysts, small ring-like calcifications, large eggshell-like calcifications, papilloma and papillomatosis, calcified fibroadenomas, haemangiomas and warts

These are distinguished by analysing of the following properties of the calcifications: form, size, density, number and distribution. The characteristics of ductal, lobular and the miscellaneous types of calcifications are described in terms of these properties. The properties of ductal and lobular calcifications are summarised in Table 12.

Type	Form	Size	Density	Number	Distribution
ductal <i>granular</i> <i>casting</i>	tiny, dot-like fragmented, with irregular contours	variable, determined by the ductal lumen	variable	variable	clustered, often within one lobe
lobular <i>cystic</i> <i>hyperplasia,</i> <i>sclerosing</i> <i>adenosis</i>	homogenous, solid, sharply outlined pearl- like density	small and punctate, little variation in size	uniform and dense	numerous	scattered
<i>milk of</i> <i>calcium cysts</i>	crescent-like or smudged depending on the view	variable	depends on size	numerous	scattered

Table 12: the characteristics of ductal and lobular calcifications as defined by Tabar and Dean [Tabar1983]

Sickles [Sickles1986] presents a two-stage strategy for the analysis of microcalcifications. First, the distinctive characteristics of various benign causes of calcifications can be used to confirm a benign interpretation. Secondly, the characteristics of suspicious microcalcifications can be used to determine if a biopsy is required, or if continued surveillance is sufficient. Six benign causes are identified.

- *Eggshell calcifications* which are 1-4mm in size and, if multiple are usually scattered.
- *Arterial calcifications* are often bilateral and appear as two parallel linear calcifications.
- *Ductal ectasia* may cause calcifications of any shape from linear to oval or round, and which may have a bilateral, symmetrical distribution. Larger calcifications will have radiolucent centres, and be oriented towards the nipple, occasionally with a branching shape.

- *Calcified fibroadenoma* can result in calcifications which occur at the periphery of the mass, form large irregular shapes and sometimes have a distinctive ‘popcorn’ shape.
- *Post-surgical calcifications* present as large amorphous streaks or as eggshell shapes, or dense masses.
- *Milk of calcium cysts* cause multiple, scattered microcalcifications, which are often bilateral. The characteristic feature is that they are gravity dependent and when appear viewed from above – as in the craniocaudal view – appear as smudges.

Malignant microcalcifications are characterised by clustering (defined by Sickles as five calcifications in a cm²). Suspicious – as opposed to malignant – microcalcifications are characterised by their round shape. These clusters can be tiny and this may make it difficult to identify their shape.

The most detailed account of the differential diagnosis of calcifications is that of Lanyi, who classifies calcifications first by their place of origin and secondly by the process which causes them [Lanyi1987]. He then describes distinctive properties of calcifications within those categories for which differential diagnosis is possible.

Two properties stressed by Lanyi are ‘septation’ and cluster shape. Septation means separated by fine lines, which is typical of benign clusters caused by microcystic adenosis¹. Lanyi’s analysis of cluster shape showed that one of the following shapes is evident on at least one radiographic plane in 97% of breast carcinomas: triangular or trapezoidal, square or rectangular, bottle or club shaped, propeller or butterfly shaped, rhomboid or kite shaped and linear or branched. An analysis of 153 clusters found no

1. Microcystic or blunt duct adenosis is a process of benign change in which proliferative and regressive change in breast lobules can create microcalcifications.

round or oval clusters with malignant aetiology. Lanyi argues that cluster shape is a key sign distinguishing calcifications of ductal origin from those of lobular origin.

Origin	No.
Calcifications of lobular origin	
Calcified microcystic (blunt duct) or sclerosing adenosis	9
Milk of calcium cysts	
<i>Scattered</i>	27
<i>Solitary</i>	8
<i>Clustered</i>	23
<i>Total</i>	58
Calcifications of intraductal origin	
Calcifications of ductal carcinoma in situ	6
Invasive carcinoma with ductal calcifications	1
Calcified secretion as in plasma cell mastitis/comedomastitis	6
Calcifications outside the lobular and ductal system	
Calcified fibroadenomas	19
Calcified liponecrotic cysts	
<i>Macrocysts</i>	5
<i>Microcysts: scattered</i>	9
<i>solitary</i>	73
<i>clustered</i>	7
<i>Total</i>	94
Unsuspectious calcifications in scar tissue	7
Other innocent calcifications of unknown aetiology	
Follow-up necessary	33
Arterial calcification	37
Calcified sebaceous glands	
<i>Scattered</i>	5
<i>Solitary</i>	8
<i>Clustered</i>	15
<i>Scattered and clustered</i>	2
<i>Total</i>	30
Total	300

Table 13: the number of microcalcifications seen in 1,044 mammograms at the Gummersbach Radiology Institute [Lanyi1987]

Lanyi summarises his approach to the differential diagnosis of microcalcifications in the following questions:

- Are different types of microcalcification present?

- Where are the microcalcifications located? (Have they the appearance of dermal calcifications, if not, is the cluster shape indicative of a lobular or ductal location?)
- What is the shape of the microcalcifications? (Are they in both views? Is septation present? Are they polymorphous?)
- Is there unilateral or bilateral change?
- If the microcalcifications are adjacent to a soft shadow, are they calcifications of a benign lesion, is the shadow a carcinoma or do they reveal a carcinoma in proximity to a benign lesion?

Lanyi also points out that the differential diagnosis of microcalcifications requires history taking, comparison with any earlier images, clinical examination, assessment of technical image quality and, possibly, additional investigations in addition to the interpretation of the mammogram.

The American College of Radiology [ACR1992] distinguishes three kinds of calcification: typically benign, intermediate concern, higher probability of malignancy. The details of this classification are given in Table 14.

TYPES OF CALCIFICATION

Typically benign

- Skin
- Vascular
- Coarse or popcorn-like
- Large rod-like
- Round (< 1mm)
- Spherical or lucent-centred
- Eggshell or rim
- Milk of calcium
- Suture
- Dystrophic
- Punctate (< 0.5mm)

Intermediate concern

- Amorphous or indistinct (round or flake)

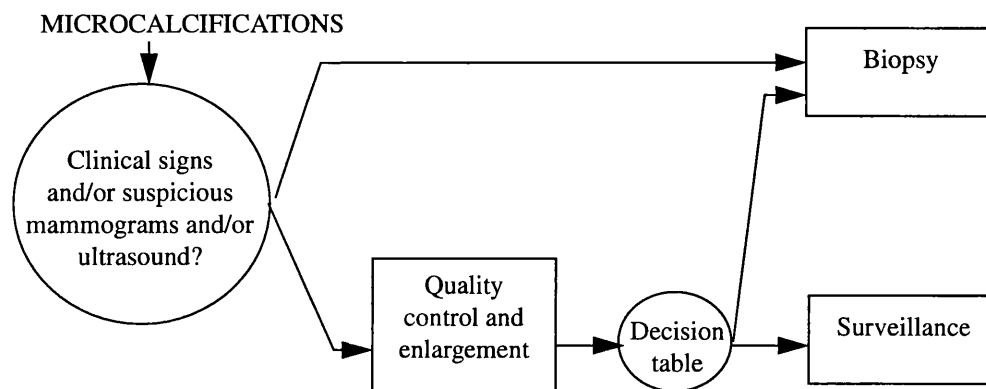
Higher probability of malignancy

- Pleomorphic or heterogeneous (granular)
 - neither typically benign nor malignant
 - more conspicuous than amorphous
 - varying sites and shapes (usually < 0.5mm)
- Fine and/or branching (casting)
 - thin irregular
 - appear linear
 - discontinuous
 - < 0.5mm in width

Table 14: classification of the American College of Radiology [ACR1992]

Bassett [Bassett1992] concentrates on describing of the appearance of benign calcifications, which are often larger, more likely to be round and monomorphic and may be scattered and diffuse. Coarse eggshell and gravity dependent microcalcifications are always benign. Bassett describes the following forms of benign calcification: calcified fibroadenoma, vascular calcifications, dermal calcifications, milk of calcium cysts, ductal ectasia, fat necrosis and artifacts. He also identifies the distinctive properties of malignant calcifications – grouped, clustered, pleomorphic, fine, branching and numerous – and, specifically, of ductal carcinoma in situ: casting, elongate, rod-shaped and branching in comedo form and punctate, bizarre, irregular and variable in cribriform and papillary types.

Dilhuydy [Dilhuydy1994] presents a 'decision strategy' for the analysis of microcalcifications. The aim is to identify those microcalcifications which can be appropriately managed by continued surveillance. The strategy is presented as a protocol in which a mammogram showing suspicious microcalcifications is followed by a review of the quality of the image and by the taking of additional, magnified images. The microcalcifications are then assessed on four criteria: topology (the signs identified by Lanyi), morphology (the signs identified by Le Gal), number and clustering and evolution. Additional factors are considered and a choice is made between biopsy and surveillance. If immediate biopsy is not indicated, a range of plans for surveillance are available, the default being for the patient to return every six months for two years.



Decision Table

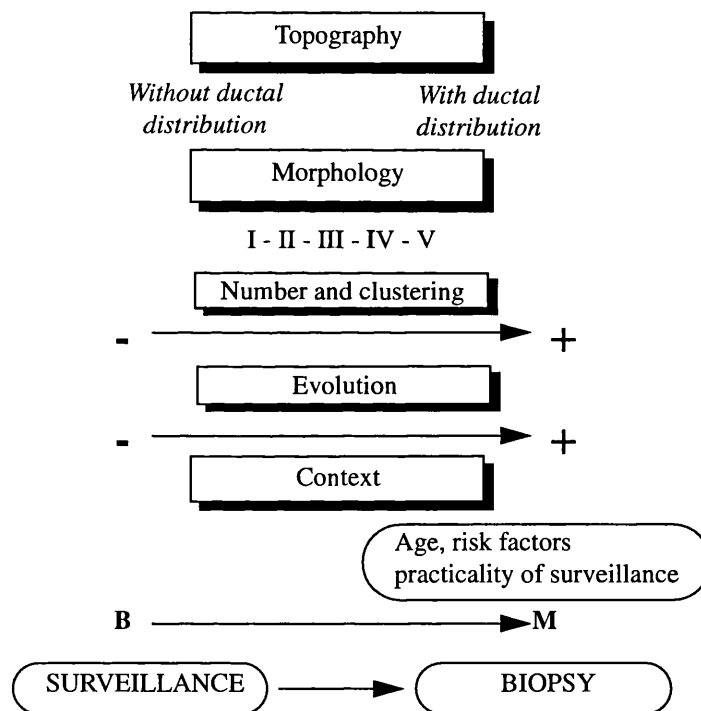


Figure 26: a protocol for the management of patients with mammograms on which microcalcifications have been detected. Decision are shown as circles, actions as rectangles. Adapted from [Dilhuydy1994].

Monsees [Monsees1995] considers two aspects of the work on the characterisation of microcalcifications, first noting two aspects of the distinctive appearance of microcalcifications associated with certain benign processes, as described by Bassett, and then reviewing attempts to stratify microcalcifications into groups with different

predictive values. Five attempts, identified by Monsees, to stratify microcalcifications into groups associated with different degrees of risk, are summarised in Table 15.

Descriptor	Number	Malignant	%	Benign	%
Clustering¹:					
Clustered	201	74	(36)	127	(63)
Scattered and Clustered	254	28	(11)	226	(89)
Total	455	102	(22)	353	(78)
Degree of Suspicion²:					
1	21	1	(5)	20	(95)
2	86	3	(3)	83	(97)
3	84	15	(18)	69	(82)
4	39	17	(44)	22	(56)
5	19	6	(32)	13	(68)
Total	249	42	(17)	207	(83)
Degree of Suspicion³:					
Moderate	157	4	(3)	153	(97)
Strong	106	59	(56)	47	(44)
Total	263	63	(24)	200	(76)
Appearance & No⁴:					
Minimal	33	2	(6)	31	(94)
Slight	46	7	(15)	39	(85)
Moderate	122	34	(18)	88	(72)
High	12	9	(75)	3	(25)
Total	213	52	(24)	161	(76)
No of Positive Criteria⁵:					
0	135	10	(7)	125	(93)
1	191	40	(21)	151	(79)
2	48	25	(52)	23	(18)
3	26	23	(88)	3	(12)
Total	400	98	(25)	302	(75)

Table 15: attempts to stratify the predictive value of microcalcifications with different characteristics, by Egan¹, Powell², Ciatto³, Hall⁴ and DeLefontan⁵

Egan [Egan1980] compared clustered with scattered and clustered microcalcifications taken from a series of biopsy-proven mammograms with microcalcifications not associated with a mass. A cluster was defined as three or more microcalcifications in an area of 0.5 x 0.5 cm. Powell [Powell1983] reviewed mammograms of biopsied lesions and classified them on a scale of one to five, on the basis of radiologists'

assessment of suspicion. Ciatto reviewed cases in which biopsies had been performed on non-palpable lesions. They divided the detected microcalcifications – into moderately or strongly suspicious – using the criteria identified by Sickles [Sickles1986]. Hall et al. [Hall1988] stratified 213 images for which microcalcifications had been the only sign which signalled a biopsy. These were classified primarily on the basis of the number of microcalcifications detected: three being considered minimally, four slightly and five moderately suspicious. The morphology of the microcalcifications was also taken into account, with benign appearance (larger, more rounded, better defined) downgrading suspicion and linear or branching morphology suggesting an upgrade. DeLafontan et al. [DeLafontan1994] evaluated 400 cases of clusters containing at least five microcalcifications. They used multivariate statistical analysis to identify five factors indicative of malignancy: the presence of vermicular microcalcifications, linear or branched appearance, irregularity of size, total number of calcifications and number per square centimetre.

Monsees suggests a protocol for the analysis of microcalcifications. Microcalcifications are classed either as definitely benign, definitely malignant or equivocal. If equivocal microcalcifications are detected further mammography is required, followed by a reassessment. If suspicious microcalcifications are found, either a surgical biopsy or a needle biopsy is indicated. If BCT (breast conservation therapy) is indicated mammography is required again to confirm that there are no residual microcalcifications. This protocol is shown in Figure 27.

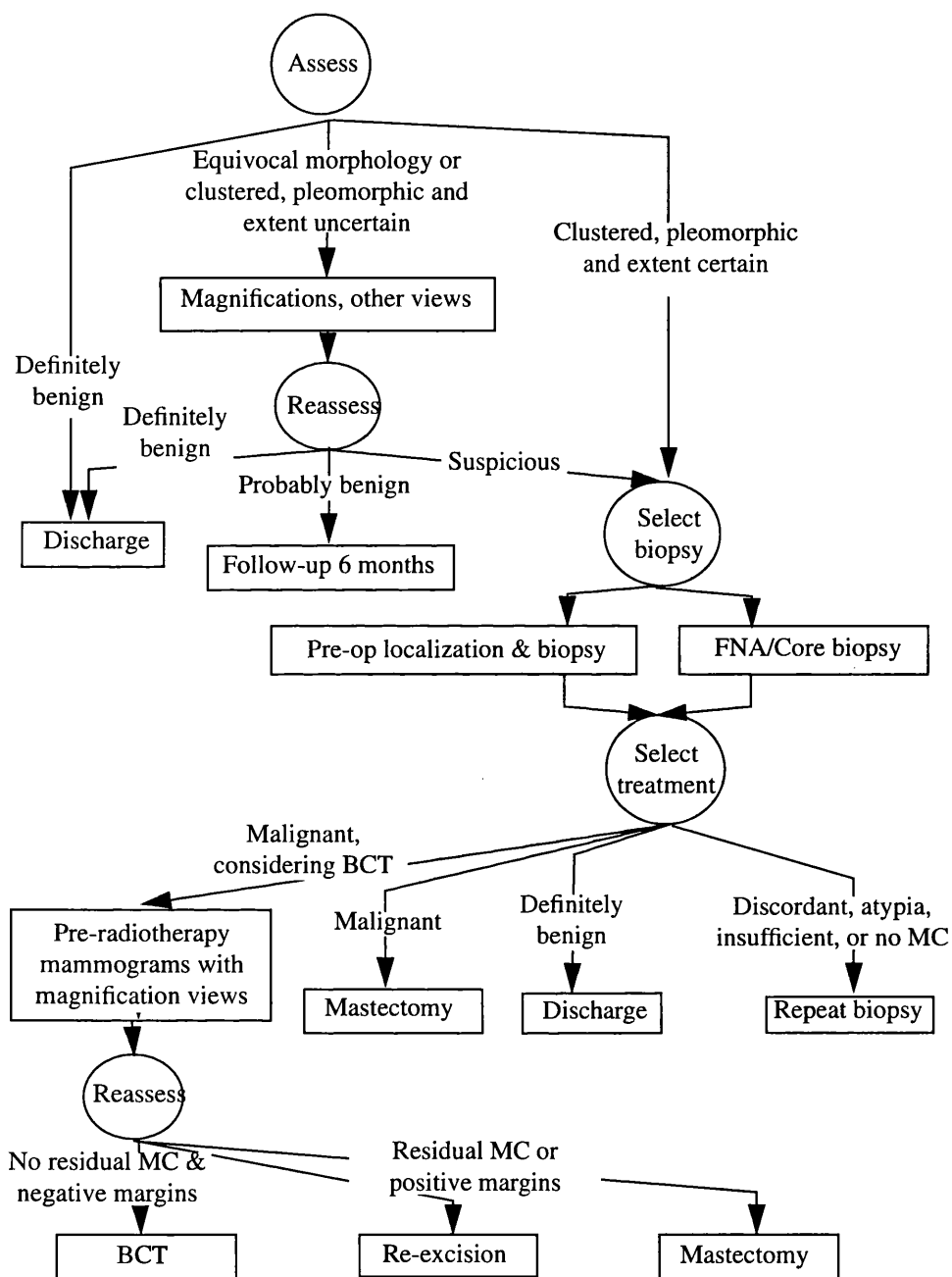


Figure 27: a protocol for the management of patients with mammograms on which microcalcifications have been detected. Decisions are shown as circles, actions as rectangles. Adapted from [Monsees1995].

6.3.3 Assessment of tissue density

In addition to the decisions routinely made by radiologists in the interpretation of mammograms, there is another perhaps should be taken into account. It has been shown [Warner1992] that the amount of dense tissue revealed in a mammogram is related to the risk of breast cancer. There is, however, no generally agreed method for classifying tissue density. The most frequently used system is that proposed by Wolfe [Wolfe1976] which defines four different kinds of pattern in breast parenchymal tissue. Another method is to measure the amount of dense tissue in the image, either by subjective estimation or by tracing areas of density in the image and measuring the traced area.

It has been suggested, [Taylor1994], that the expertise needed to interpret mammograms is actually used only in the interpretation of mammograms with considerable amounts of dense tissue, or ones in which there is an abnormality. If this hypothesis is correct, one acceptable way to relieve the pressure on radiologists' time would be to filter off those mammograms which do not require the attention of the radiologist but could be interpreted reliably by a trained technician or a computer. In order to establish the truth of this hypothesis, and indeed before it is possible to test whether this filtering is feasible, it is necessary to establish that radiologists can agree on an assessment of whether mammograms are fatty or dense.

Another reason for measuring the amount of dense tissue in an image is to test for asymmetry between the breasts, which may be a sign of breast cancer [Miller1992].

Although the assessment of the amount of dense tissue in a mammogram is not a task which radiologists are called on to perform as a matter of course, it is one which may be of value in the future and which radiologists currently find difficult.

6.3.4 Conclusions

In order to assess the potential for decision support in mammography, it is first necessary to identify the different decisions which are being made in the interpretation of mammograms. At one level the radiologist's decision is to do with the management of the patient: "Should she be recalled for a biopsy or further investigation or can she be safely returned to normal screening?" The same question can be cast in different terms: "Is the most likely diagnosis one of cancer or of benign disease?". Authors writing about mammography sometimes assume that the radiologist's task is simply to evaluate the information in the mammogram, sometimes it is assumed that the radiologist has access to additional information about the patient. It is, therefore, possible to identify three different, but related, decisions which radiologists are called upon to make, or to which radiologists contribute: the classification of an abnormality on an image, a preliminary diagnosis based on all available information and the selection of a course of action. The radiological knowledge required to make these decisions is the same. An interpretation of an image could result in a single set of observations which would serve as the basis of arguments to be used in support of a classification of intraductal microcalcifications, a diagnosis of cancer and a recommendation to biopsy. The framework proposed in Chapters Four and Five for decision support would allow a knowledge base and a set of image-processing measures to be developed which could be used with different decision specifications to support each of these three decisions.

In assessing the potential for a decision support tool based on image processing and symbolic reasoning, the following criteria must be considered:

- the difficulty of the decision
- the importance of the decision
- the role to be played by image processing
- the role to be played by symbolic reasoning

In the analysis of breast masses, for example, the cases of stellate lesions and circumscribed lesions are rather different. Stellate lesions are often difficult to detect but, if found, the risk of misclassification is small, since there are only a few possible benign interpretations. Where the possibility of confusion does exist, in the case of radial scar (a benign lesion similar in radiological appearance to a stellate lesion), biopsy is recommended [Feig1992], that is to say the radiologist does not even attempt to decide if the lesion is benign or malignant, and so support for the decision is not required.

The number of possible diagnoses of circumscribed lesions is somewhat greater. It might be argued that the decision space is defined not by the number of different diagnoses, but by the number of possibilities for management, which remain limited to surveillance, further investigations and biopsy. However the choice of management strategy is determined by the diagnosis. The decision here is important, since circumscribed masses are so common and have such a low specificity for cancer that to biopsy all of them would be impractical [Feig1992], and so radiologists are obliged to try and identify those lesions which can be managed by surveillance.

A number of authors have discussed the detection of breast masses by means of image processing [Lai1986, Yin1994, Chan1995a, Gupta1995, Zheng1995], but less attention has been paid to the characterisation of masses. Giger [Giger1994] and Huo [Huo1995] used measures of spiculation to distinguish benign and malignant masses, on the assumption that all stellate lesions should be treated as malignant and that circumscribed lesions need be treated as malignant only if judged to be so by a radiologist. Claridge and Richter [Claridge1994] measured edge blur, Stewart et al. [Stewart1994] measured density, form, contour and size to assess circumscribed lesions.

The knowledge base used to drive symbolic reasoning about breast masses would be rather limited since there are relatively few properties to be considered. Tabar and Dean list two which are of primary importance and two of secondary importance, Feig lists nine. One group have developed a knowledge base for the characterisation of circumscribed lesions, to be used in association with the image-processing measures described in the previous paragraph [Stewart1994].

In the case of calcifications, the decision is clearly difficult. There are a number of competing explanations for the cause of calcifications and, in some cases, reasonably complex descriptions of their characteristic properties. The radiologist's interpretation is important since it is possible to establish that some calcifications are benign on the basis of their radiological appearance, and in addition to this there is an identifiable category for which surveillance is a better option than biopsy.

A number of authors have used image-processing techniques as an aid to the classification of microcalcifications; and this work is summarised in Section 6.5. A considerable amount of information about the differential diagnosis of calcifications is contained in the articles reviewed earlier in this section, suggesting that a worthwhile knowledge base could be constructed describing the various types of microcalcification and their diagnostic significance.

On the basis of the above assessment, I decided to develop a knowledge base that would support the following decisions: feature classification, selection of next action and assessment of tissue density. I concentrated on the problem of the classification of calcifications since the radiological literature suggested that there was a significant body of knowledge on this problem and the literature on the application of image processing to this area suggested that there were a number of automated techniques which could be used. These two factors suggest that this was a promising area for the combination of image processing and symbolic reasoning.

6.4 Image Processing for the Detection of Calcifications

If image processing is to be used to assist in the characterisation of microcalcifications, they must first be detected, either automatically by the computer or by the radiologist. Obviously automatic detection is preferable. A great deal of research in the processing of medical images has been directed at the problem of detecting microcalcifications [Chan1988, Fam1988, Davies1989a, Karssemeijer1992, Shen1993b, Chan1995b, Qian1995, Guillemet1996]. I do not propose to provide a comprehensive review of this work: seven papers included in a recent workshop on digital mammography provide a representative sample of the approaches taken [Brett1994, Chitre1994, Kegelmeyer1994, Nishikawa1994, Parker1994, Strickland1994, Valatx1994]. The problem is a difficult one, and a variety of different techniques have been applied, most of which share a common structure. First the image is enhanced in some way, to remove noise or to filter out the components of the signal related to background tissue, secondly there is a detection phase and thirdly the results of the detection are improved by applying criteria which discriminate between true microcalcifications and false positives. In this section I summarise only the techniques used in the second stage, since techniques of noise reduction are outside the scope of this thesis and the decision criteria employed in the third phase overlap with those described in the next section for the classification of calcifications.

The techniques used to detect microcalcifications fall into two categories, those based on thresholding and those based on matched filters. Thresholding is employed by Chitre et al., Nishikawa et al. and Valatx et al. Chitre uses a threshold on the standard deviation in pixel values in a surrounding region. Nishikawa applies a combination of local and global thresholds on the grey-level of pixels. Valatx applies a local threshold to each image in a stack of images of different resolution.

A variety of filters are used. Brettle et al. use matched filtering in the frequency domain. Kegelmeyer applies a set of filters that have been used by other investigators and uses the results as input to a binary decision tree. Parker describes a filter based on the idea of hysteresis: line profiles are taken through the image at different orientations and then analysed by a filter which is analogous to passing a loop of wire over a one-dimensional signal. The loop will detect all local maxima where the difference in grey-level is larger than the loop. Strickland et al. use a 'wavelet' transform which acts as a bank of multiscale matched filters.

There is no easy way to compare the performance of these different techniques on the basis of the published articles. The results are based on different sets of data, the criteria used to define successful detection vary and the results are reported in quite different ways. Since the detection of calcifications is not of intrinsic interest for this thesis, I decided not to attempt to identify the best detection technique but to re-implement the filters described by Kegelmeyer and to test whether these could form the basis of a detection algorithm which would allow the application of measures for the classification of calcifications. I decided to take advantage of Kegelmeyer's work since this in turn built on techniques developed by other authors and therefore seemed likely to be generally applicable. Details of the implementation are provided in Chapter Seven.

6.5 Image Processing for the Classification of Calcifications

The aim of this work is to provide a decision aid for the classification of calcifications which is able to provide information based not only on knowledge-based reasoning but also on data derived from image processing. Nine sets of authors have recently published papers on the use of various kinds of computer system to help distinguish benign from malignant microcalcifications. In this section I summarise the

approaches taken, and then consider how elements of this work could best be incorporated into a decision-support tool for image interpretation.

In the following nine sub-sections I give outline summaries of these papers. There is a common structure to the approaches taken, and this determines the form of the outline summaries. First, each system has some mechanism for detecting calcification data in images, looking either for individual microcalcifications, clusters or regions of microcalcifications. Different measures are then taken of the data and these measures are used as the basis for some form of classification rule. Different techniques are used to generate the classification rule and in some systems a subset of useful measures is first identified. In the next nine sub-sections, the aim of the research is stated as well as the method used to identify the calcifications and their clusters, the set of measures considered, the set of measures finally adopted – or described as having been most useful – the technique used to identify the classification, and the results of any test of the classification. Detailed definitions are provided in the Appendices.

6.5.1 Magnin [Magnin1989]

Aim

The aim was to distinguish two types of microcalcification: lobular and intraductal. The value of this distinction is that lobular calcifications are more likely to be benign. The paper deals with the properties of individual microcalcifications and not those of clusters.

Detection of microcalcifications

The microcalcifications were identified automatically from a pre-processed image using a combination of thresholds, one computed by adaptive grey-level thresholding (adaptive in the sense that the threshold is computed for each pixel on the basis of statistics in a local region around the pixel) and the other based on a residual energy

measure (the sum of the squared differences between pixel values and a threshold for the region).

Measures considered

Seven measures were considered:

- perimeter
- area
- compactness¹
- eccentricity
- convexity
- lengthening rate
- mean ray vector

Classifier

No classifier was developed.

Measures used

No formal analysis was done but it is suggested that four measures can be used to distinguish the two types of microcalcification:

- compactness
- convexity
- lengthening rate
- mean ray vector

1. Compactness is a measure of the degree to which a shape resembles a circle; its definition is given in the Appendices.

Test

No test was performed.

6.5.2 Lefebvre [Lefebvre1991]

Aim

The aim was to identify clusters of benign microcalcifications. The misclassification of a benign cluster as malignant is taken to be less costly than the reverse error.

Detection of microcalcifications

Microcalcifications were identified manually by a radiologist and segmented automatically. The segmentation procedure measured statistics in a window around a pixel and identified pixels which were three standard deviations above the mean. Regions were then grown around these seed points using the same threshold. The threshold was modified according to the size of the region and also in order to prevent the merger of adjacent regions.

Measures considered

Sixteen measures were computed:

- number of microcalcifications within a cluster
- number per unit area
- coefficient of variation¹ of pixel value
- coefficient of variation of compactness
- the three lowest compactness values
- compactness of the convex hull² of the cluster

1. The coefficient of variation is the ratio of the standard deviation and the mean.

2. The convex hull of a shape is the smallest convex shape which will contain it.

- mean distance between calcifications
- moment of inertia
- moment relative to principal axis of inertia
- Hu's seven invariant moments¹

Four clinical features were also recorded: age, number of children, personal and family history.

Classifier

A classification and regression tree method was used to construct a decision tree. This process involved the construction of a tree of nodes; each of which represented a set of clusters of microcalcifications. At each step in the process a set of clusters in the parent node was divided in two, creating a pair of sub-nodes. To make these divisions all the possible thresholds of each of the measures were considered and thresholds selected in order to maximise a measure based on the ratio of benign and malignant clusters in each sub-node. The process continued until all the terminal nodes contained either only benign clusters or only malignant clusters, at which point the tree could be pruned.

Measures used

All the image measures were used in the decision tree, the principal ones being:

- coefficient of variation of optical density
- compactness of the cluster
- moment relative to the principal axis of inertia
- moment of inertia

1. The moment of an object in an image is defined by analogy with the physical property, with pixel value playing the role of mass. Hu used the moments about the centre of an object, normalised for area, to calculate seven moments which are invariant to rotation. These are defined in the Appendices.

Test

The system was developed using 240 cases of which 163 were benign. The tree produced seven terminal nodes which contained only benign cases, representing a total of 75 benign cases. This means that the classifier was able to identify as benign, a significant subset of the benign images (75 out of 163, or 46%).

6.5.3 Patrick [Patrick1991]

Aim

The aim was to classify clusters as benign or malignant.

Detection of microcalcifications

Microcalcifications were identified by “segmentation”. No details are given.

Measures considered

The measures of individual calcifications and of clusters were computed on both craniocaudal and mediolateral views. Individual calcification features:

- length of the major axis
- aspect ratio (ratio of major to minor axis)
- ratio of the sum of the first moments of boundary points about the major and minor axes
- standard deviation of the distance of the boundary points from the centre of gravity
- compactness (based on perimeter)
- compactness (based on convex hull)
- ratio of the two compactness measures
- gravitational potential of the other calcifications within the cluster

Cluster calcification features:

- number of calcifications
- mean aspect ratio
- standard deviation of aspect ratio
- maximum figure values
- standard deviation of calcification areas (normalised by the mean area)

A number of other features were used, including the incidence of bilateral calcifications and the location of the microcalcifications by quadrant. This data was termed 'clinical data'.

Classifier

The classifier used was a network of five expert learning systems. The expert learning systems used proprietary software (Outcome Advisor) and details of the process by which a classification was 'learnt' were not given. Each system was fed inputs from only one of the five subsets of data: cluster data from the mediolateral view, cluster data from the craniocaudal view, individual calcification data from the mediolateral view, individual calcification data from the craniocaudal view and the clinical data. The training phase, in which the classification is learnt, was performed using 108 cases. The five expert learning systems were tested in five different configurations.

Measures used

No selection of measures was performed.

Test

The five different configurations were tested on 20 images, ten of which contained a total of 15 benign clusters and ten of which contained malignant clusters.

The most successful of the five recognised 9/10 malignant clusters and 9/15 benign clusters (72%). The probability of the system performing at random and classifying 72% of images correctly is 0.02.

6.5.4 Shen [Shen1993a, 1993b]

Aim

The aim was to classify microcalcifications as benign or malignant.

Detection of microcalcifications

A region-growing process was employed in which every pixel with a value above the mean for the entire breast area was used as a seed. The criterion used in this process had three parameters: the maximum and minimum pixel values of the region being grown and a tolerance value. Three features (compactness, size and centre of gravity) were computed for a range of tolerance values and the change in values of these features used to determine the final tolerance value. The resulting region was treated as a microcalcification only if size and contrast criteria were met.

Measures considered

Three shape descriptors were used:

- moment descriptor
- Fourier descriptor
- compactness

Classifier

The classifier was a three-layer back-propagation neural net, with three input nodes and two output nodes. The number of nodes in the hidden layer and the 'momentum' and 'gain' parameters for the learning equations were set following a

leave-one-out testing process. The chosen number of hidden nodes was ten. The training phase used 143 calcifications from 18 images.

Measures used

All three measures were used.

Test

The test used four images, containing a total of 58 benign and 241 malignant calcifications. The classification rates were 94% for benign and 87% for malignant.

6.5.5 Wu [Wu1993]

Aim

The aim of the network was to distinguish benign and malignant lesions.

Detection of microcalcifications

The ratings were entered by the radiologist.

Measures considered

An expert radiologist identified 43 features from a textbook [Tabar1983]. The features pertaining to microcalcifications were:

- number in cluster
- presence of microcalcifications elsewhere in breast
- rounded shape
- linear shape
- branching shape
- uniformity in shape

- uniformity in size
- uniformity in density
- diffuseness of cluster
- geometric shape of cluster
- radiation along ducts
- presence of associated macrocalcifications
- macrocalcifications elsewhere in breast

Classifier

The classifier was a three-layer feed-forward neural network with five hidden nodes, trained on the 133 cases in the textbook and tested on 60 clinical cases.

Measures used

It was found that there was little or no difference in the average values of some features between benign and malignant cases. A reduced feature set, containing the features that two experienced mammographers considered to be among the most important, was identified from an analysis of the training and clinical cases:

- presence of microcalcifications elsewhere in breast
- rounded to irregular
- presence of linear microcalcifications
- presence of branched microcalcifications
- shape of cluster, geometric to regular
- radiation of microcalcifications along ducts

Test

The network was better at merging feature ratings to decide whether microcalcifications were benign or malignant than non-specialist radiologists were. A_z (the area under the ROC curve¹) for the reduced feature network on 60 clinical cases was 0.89. On the same images the average score of five attending radiologists was 0.84.

6.5.6 Chitre [Chitre1994]

Aim

The aim of this work was to classify difficult-to-diagnose clusters.

Detection of microcalcifications

Segmentation of the microcalcifications was performed by computing the local contrast image about each pixel using the local deviation of gray levels as a threshold. The resulting pixels were closed by a morphological operation and regions of fewer than five pixels discarded.

Measures considered

Three classes of measure were considered. Cluster features:

- number of microcalcifications
- average area
- average gray level
- standard deviation of gray levels
- standard deviation of gray levels per pixel

1. Receiver Operating Characteristic curves are used to assess the power of a diagnostic test by plotting the sensitivity (proportion of disease cases correctly identified) against 1 - sensitivity (proportion of disease-free cases correctly identified) for different decision thresholds. The greater the area under the ROC curve, the better the test.

- average distance between calcifications
- standard deviation of distance between calcifications
- average distance between calcifications and centre of mass
- standard deviation of distance between calcifications and centre of mass
- potential energy using product of area and average gray level to indicate mass

Image structure features (of second order histogram or co-occurrence matrix¹):

- entropy
- contrast
- angular second moment
- inverse difference moment
- correlation
- mean
- deviation

Image structure features (of a “difference histogram” derived from the second order histogram):

- entropy
- contrast
- mean

Classifier

The image structure features and the cluster features were used as the inputs to six different neural net architectures.

1. The second order histogram, or co-occurrence matrix, represents the distribution of the probability that pairs of grey levels occur separated by each of a set of distances.

Measures used

A k-nearest-neighbour¹ procedure was used to identify the best five features based on image structure (the higher order features) and the five best cluster features.

Image structure features (based on second order histogram):

- angular second moment
- correlation
- mean
- deviation
- angular second moment (first order histogram)

Cluster features:

- number of calcifications
- standard deviation of gray levels
- average distance between calcifications
- average distance between calcifications and centre of mass
- standard deviation of distance between calcifications and centre of mass

Test

The system was tested on a database of difficult-to-diagnose images. The best performance came when using the combination of second-order and cluster features on a network trained to contain RMS error. The area under the ROC curve was given as an average over different divisions of the data into training and validation sets and varied between 0.57 and 0.61 over the different network topologies.

1. In the k-nearest-neighbours technique, an object is measured on a set of dimensions and compared to the k most similar objects and assigned a classification based on a consideration of the classes assigned to the k nearest neighbours.

6.5.7 Nishikawa [Nishikawa1993, Nishikawa1994]

Aim

The system was designed to identify malignant clusters of microcalcifications.

Detection of microcalcifications

The breast area was segmented and the background texture suppressed before global and local thresholding were used to identify microcalcifications. Features such as contrast, area and texture were used to eliminate false positives. The identified clusters were then analysed by a shift-invariant neural network to further reduce the number of false clusters.

Measures considered

In a 1993 paper [Nishikawa1993] the following properties were measured:

- area
- effective thickness¹

and the following features calculated:

- the mean effective volume²
- standard deviation of mean effective volume
- standard deviation of mean area
- standard deviation of mean effective thickness

It was found that a combination of thresholds on these measures separated 14 of the 29 benign clusters while correctly classifying all ten malignant clusters.

1. The effective thickness is computed from the local contrast.

2. Since Nishikawa takes contrast to be a measure of the effective thickness of a microcalcification, the effective volume is computed by multiplying the area by the effective thickness.

Classifier

Eight properties were used as input to an error-back-propagation neural network with eight input nodes and six hidden nodes; the output was related to the probability of a malignant cluster.

Measures used

The 1994 paper [Nishikawa1994] lists the following eight properties:

- cluster area
- cluster circularity
- number of microcalcifications per cluster
- mean effective thickness of microcalcifications in a cluster
- standard deviation of effective thickness of microcalcifications in a cluster
- mean area of microcalcifications in a cluster
- mean effective volume of microcalcifications in a cluster
- second highest irregularity measure of microcalcifications in a cluster

Test

The system was evaluated using a database of 100 images, from 53 patients containing 107 clusters. In all the benign clusters the probability of malignancy had been sufficient to warrant biopsy. The results are presented on both a per cluster and per patient basis. All 19 patients with cancer were correctly classified as having malignant clusters while 16 of the 34 with benign disease were correctly classified (A_z , the area under the ROC curve, was 0.92). A_z for the per cluster analysis was 0.83.

6.5.8 Lo [Lo1995]

Aim

The aim was to predict biopsy outcome.

Detection of microcalcifications

Feature values were entered by the radiologist.

Measures considered

There were 29 measures, of which the following related to calcifications:

- calcification morphology
none, milk of calcium-like, rim, skin, vascular, spherical or suture,
coarse, large rod-like, round or dystrophic, punctate or indistinct,
fine branching
- calcification density
none, dense, mixed, faint
- calcification distribution
none, scattered, intermediate, clustered
- calcification number
none, fewer than 5, 5 to 10, more than 10

Measures used

The features were compared on a leave-one-out¹ basis. Calcification density and calcification morphology were the first and third most important features considered (other important features were related to breast masses and patient age).

1. A leave-one-out procedure involves computing a performance measure for the classifier based on all but one of the measures, and repeating the computation as many times as there are measures, leaving one out each time. The impact on the performance of leaving each measure out can then be determined.

Classifier

Two classification techniques were considered: back-propagation neural networks and Linear Discriminant Analysis (LDA). The number of hidden nodes used by the neural networks was varied between five and 25 to find the optimum, but only minimal differences were found. LDA performed significantly worse than the neural network.

Test

The system was developed using 260 cases of which 108 had calcifications. Since the values had to be entered by the radiologist, it was thought important to minimise the number of features used by the network. On those 108 images the network, using only age and density and morphology as measures, out-performed expert radiologists, with an A_z of 0.95.

6.5.9 Parker [Parker1995]

Aim

The aim was to distinguish calcifications arising from Comedo and Non-Comedo Ductal Carcinoma in Situ (DCIS), since Non-Comedo DCIS may be biochemically and prognostically distinct from DCIS.

Detection of microcalcifications

Microcalcifications were identified manually by the radiologist and then segmented manually. Clusters were identified automatically as being sets of microcalcifications which had a minimum of five members and where the nearest neighbour was less than 70 pixels (1.2mm) away.

Measures considered

Measures of individual calcifications and of clusters were considered.

Measures of individual microcalcifications:

- area
- foreground
- background
- difference between foreground and background
- difference between foreground and background/background
- compactness
- entropy
- edge-strength
- variance
- directionality (ratio of sides of bounding rectangle)
- ratio of object area to bounding rectangle
- eccentricity₁ (ratio of maximum and minimum distances from the centre of mass)
- eccentricity₂ (ratio of axes of the best fit ellipse)
- eccentricity₃ (ratio of moments of inertia of the best fit ellipse)
- ϕ_1 (Hu's first invariant moment)
- ϕ_2 (Hu's second invariant moment)
- radius of gyration
- pair (ratio of directionality with that of the nearest neighbour)
- whether the centre of the mass lies outside the object

Measures of microcalcification clusters:

- number in cluster
- fraction in cluster
- perimeter of cluster
- compactness of cluster
- density of cluster

The range, mean and variance of the individual measures were also calculated.

Classifier

A k-nearest-neighbours technique was used. The large number of features was reduced by two techniques. First cross-correlation was used to remove redundant measures, and second the classifier was tested with each measure individually – any measure not achieving a classification better than chance was removed.

Measures used

Measures of individual microcalcifications

- difference between foreground and background
- entropy
- eccentricity₁ (ratio of maximum and minimum distances from the centre of mass)
- eccentricity₂ (ratio of axes of the best fit ellipse)
- ϕ_1 (Hu's first invariant moment)

Measures of microcalcification clusters

- compactness of cluster

- variance in compactness of microcalcifications
- variance in edge strength of microcalcifications
- range of radius of gyration
- mean of radius of gyration
- ϕ_4 (Hu's fourth invariant moment)

Test

The system was developed by allocating 42 cases (of which 11 were non-comedo) to training and reference sets in six different distributions in order to maximise the data. An area under the ROC curve of 91% was achieved.

6.5.10 Discussion

All of these papers followed the same procedure: they first obtained measures of microcalcifications and microcalcification clusters from a set of training images and, used this data to develop a classification rule which was tested on a different set of images. A variety of methods were employed to analyse the data used to establish the classification: neural nets of various kinds, decision trees, expert learning systems and kNN (k-nearest neighbours) classifiers. Some authors report that they tried a large set of measures and then identified a smaller set of useful measures; others report which measures were assigned the strongest weights.

Table 16 summarises the different image-processing measures described in this chapter. Only measures which can be defined in terms of image-processing operations and which attempt to describe either the microcalcifications or their clusters are included. Measures based on radiologists' ratings (such as those of Wu et al. and Lo et al.) or on the texture of regions containing microcalcifications (the second order

measures used by Chitre et al.) are excluded. In addition, measures not considered by investigators to be among the most useful, are discarded.

The purpose of this comparison is not to establish which of the above systems performs best. The question is inappropriate since the systems were designed with different aims: two concentrated on distinctions which are tangential to the purpose of this thesis [Magnin1989, Parker1995]. Some of the systems looked at individual micro-calcifications, others at clusters and regions. Even among those systems designed to classify clusters as benign or malignant there are important differences of emphasis. A system designed to perform well on difficult-to-diagnose clusters may well misclassify some of the easy-to-diagnose cases handled by another system, and yet the latter system might well perform little better than chance when tested on the harder cases.

		Magnin	Lefebv.	Patrick	Shen	Chitre	Nishik.	Parker
<i>cluster shape</i>	compactness		X				X	X
	area						X	
	4th invariant moment							X
<i>density</i>	mean gray level						X	
	contrast							X
<i>variation in density</i>	s.d. of gray level		X			X	X	
	s.d. of edge strength							X
<i>distribution</i>	mean separation		X			X		
	number per unit area		X					
	mean dist. from centre					X		
	s.d. dist. from centre					X		
	grav. pot. of other mc.			X				
<i>number</i>	number		X	X		X	X	
<i>shape</i>	perimeter	X						
	compactness	X		X	X			
	compactness of hull			X				
	ratio of compactnesses			X				
	lowest compactness		X					
	convexity	X						
	lengthening rate	X						
	mean ray vector	X						
	moment descriptor				X			
	moment of inertia		X					
	1st invariant moment		X					X
	invariant moments 2-7		X					
	moment w.r.t. axis		X					
	ratio of moments			X				
	2nd highest irregularity						X	
	entropy (texture)							X
	eccentricity/aspect ratio	X		X				X
	eccentricity ²							X
	mean gyration							X
	Fourier descriptor				X			
	length of major axis			X				
	s.d. of boundary			X				
<i>variation in shape</i>	s.d. of compactness		X					X
	range of gyration							X
	s.d. of aspect ratio			X				
	max. figure value			X				
<i>size</i>	mean gray level x area						X	
	mean area	X					X	
<i>variation in size</i>	s.d. of area			X				

Table 16: measures used in the classification of calcifications

Several points are suggested by a comparison of the systems. There are an enormous number of measures which one could consider using in the classification of microcalcifications – the Parker et al. system alone considers over 77. Very few of them are used by more than one investigator. It appears that an acceptable performance can be obtained with a system which employs relatively few measures: in the Lo et al. system, for example, only three measures were required for radiologists' performance to be improved.

The important question, in this context, is how image processing can provide a radiologist with information that will assist in the differential diagnosis of microcalcifications. Measures which are susceptible to an intuitive interpretation in ordinary language are perhaps more likely to be useful. The following measures are chosen on the grounds that they have been applied successfully by two or more authors:

- compactness of cluster shape
- mean separation of microcalcifications
- number
- compactness of shape
- eccentricity of shape
- standard deviation of compactness
- mean area

Hu's first invariant moment is excluded from this list. It is a measure of shape, but not one which corresponds to a property radiologists might ascribe to a shape. The following measures have been successfully employed by only one author, but have an intuitive mapping to properties used by radiologists in order to characterise the difference between benign and malignant calcifications:

- mean grey-level
- standard deviation of grey-level
- standard deviation of area

Since eccentricity was included as a measure of shape, it makes sense to include its standard deviation as an additional measure of variation in shape. This measure and the ten listed above were all considered for inclusion in the system described in Chapter Seven.

6.6 Discussion

The detection and treatment of breast cancer is a major healthcare problem in the UK and other developed countries. The investigation of preference for the detection and assessment of breast cancer is X-ray mammography. Mammography screening programmes generate an enormous number of images, and an enormous demand for radiological expertise. Attempts to develop computer aids for mammography have so far been aimed at the screening programme, where large numbers of images are very rapidly inspected, the speed involved places strong constraints on the kind of decision aid which could be provided.

The treatment of women referred from the screening programmes also creates a burden on the healthcare system. In the UK during 1994/1995, 1,207,316 women were screened, of whom 63,925 were recalled for further mammography [Patnick1996]. The decisions made in these cases would very often have turned on an assessment of whether the appearance of microcalcifications was suggestive of malignant disease – microcalcifications account for over 50% of non-palpable lesions detected through mammography [Ciatto1987]. There is therefore considerable scope for a decision aid, if it could improve assessment at this stage.

The above review of the literature on the classification of microcalcifications suggests that there is a substantial amount of information which could be represented in a knowledge base. A review of the research on the application of image processing to the detection and classification of microcalcifications suggests that a number of techniques could be employed to provide additional information.

CHAPTER SEVEN

THE MAMMOGRAPHY WORKSTATION

7.1 Introduction

In this chapter I describe a prototype system to provide decision support for the interpretation of mammograms. The system is implemented according to the generic design described in Chapter Five. Five information sources each have a separate display function – the set of display functions makes up the user interface – and three information processing modules – the decision procedure, task manager and image processing operators – draw on the information sources and generate new information. The decision procedure and task manager are as described in Chapter Four; the image-processing operators are described in the next section. Later sections describe the knowledge bases representing the relevant medical knowledge and the protocol to be followed. The last section is an overview of the prototype.

7.2 Image-Processing Operators

The image-processing operators provided in the mammography system can be discussed under the following headings: the detection of microcalcifications, the measurement of microcalcifications and the measurement of tissue density.

7.2.1 Detection of calcifications

The important issue, as far as this thesis is concerned, is how image processing can provide a radiologist with information which will aid the differential diagnosis of microcalcifications. This, in practice, requires the automatic identification of microcalcifications. Kegelmeyer [Kegelmeyer1994] presents an approach to the detection of microcalcifications which involves six different segmentation operators, each of which had been found useful in earlier research. Three of these operators are termed ‘all-at-once’, because they return a single image in which every detected microcalcification is marked; three are termed ‘at-every-pixel’ because they find the best or most likely object centred at every pixel and so must return a set of sub-images, one for each pixel. One of the ‘at-every-pixel’ segmentors, the adaptive threshold, is also described by a number of authors, including Davies et al. [Davies1989], who implement it in an ‘all-at-once’ manner. Since the detection of microcalcifications is not the primary interest of this thesis, and in order to avoid the complication of handling multiple overlapping sub-images, I decided to see whether an adequate segmentor could be developed based on the three ‘all-at-once’ approaches described by Kegelmeyer and the adaptive threshold technique described by Davies.

The four segmentors were therefore as follows:

- *W (Woods)*: the difference between each pixel in the image and the mean of its surround is calculated. The resulting difference image is then thresholded with one of two values, depending on the maximum value in the surrounding

region.

- *O (Outlier)*: each point in the image is classified using a threshold calculated as the mean of the pixel values of the surrounding region plus three standard deviations.
- *B (Boundary)*: an edge detector is applied to the image, and the edge image is then dilated and eroded. Edge points lost in the dilation and erosion are restored and the enclosed regions considered as objects.
- *A (Adaptive threshold)*: overlapping regions are classified using a threshold computed for the region. A region is considered to be detected only if it is selected in three out five overlapping regions. The threshold is computed from the histogram, it is the minimum value between the two peaks in a bipolar histogram. A histogram is considered to be bipolar only if it contains two peaks separated by a minimum distance.

These four segmentors were combined. Kegelmeyer reports that the Woods segmentor produced the best approximation to the outline of microcalcifications. The combined segmentor, S , was therefore designed to return the regions identified by the Woods segmentor, but only if the region overlapped with a region detected by t other segmentors. The segmentor was tested with $t = 2$. Due to the computational inefficiency of the adaptive threshold segmentor, a segmentor, S^* , based on a combination of the other three was also tested, with $t=1$ and $t=2$. Regions of fewer than eight pixels in size were ignored, since it was judged that no useful shape information could be obtained from such a small area.

The segmentors were tested on images from a database distributed by the Mammographic Image Analysis Society (MIAS) [Suckling1994]. These images are mammograms from the UK screening programme digitised on a Joyce-Loebel scanning densitometer with a linear response in the range 0.0 to 3.2 at a spatial resolution of 50 microns and with eight bits per pixel. I discovered that a fault in the

digitisation meant that pixel values 64, 128 and 192 were suppressed, and these grey-level values were therefore removed. The position of lesions in the images was identified by radiologists and the nature of the lesions confirmed by histology. However, the radiologists had not indicated the outline of each microcalcification. For the purposes of this test I was therefore forced to use my judgement as to what constituted a microcalcification. The results are given in Tables 17 and 18.

Images	W			O			B			A		
	TP	FN	FP	TP	FN	FP	TP	FN	FP	TP	FN	FP
mdb209	26	20	>100	28	20	7	5	40	11	13	30	>50
mdb211	12	9	>100	7	14	15	6	15	11	5	16	>100
mdb213	27	5	>100	14	18	13	14	18	>50	6	26	>50
mdb214	1	0	>100	1	0	16	1	0	7	1	0	>50
mdb216	11	1	16	9	3	15	1	11	5	7	5	>50
mdb218	6	6	>100	8	3	14	7	5	15	4	8	>50
mdb219	14	0	>100	10	4	16	10	4	11	6	8	>50
mdb222	6	0	>100	6	0	9	2	4	3	3	3	>50
mdb223	20	13	>50	17	13	13	8	24	17	11	20	>50
mdb226	9	0	>100	8	1	25	4	5	30	1	8	>100
Total	132	54	>1000	108	76	143	58	126	160	57	124	>600

Table 17: results from testing the four individual segmentors on images from the MIAS database. TP are true positives, i.e. detected calcifications; FN are false negatives, i.e. missed calcifications; FP are false positives, i.e. returned regions which do not correspond to calcifications.

Images	$S^*, t = 1$			$S^*, t = 2$			$S, t = 1$		
	TP	FP	FN	TP	FP	FN	TP	FP	FN
mdb209	19	27	0	1	45	0	13	33	0
mdb211	3	18	2	0	21	0	2	19	1
mdb213	12	22	2	2	32	0	10	22	0
mdb214	1	0	1	1	0	0	1	0	0
mdb216	8	4	0	0	12	0	8	4	0
mdb218	3	9	1	1	11	0	2	10	1
mdb219	13	1	0	4	10	0	12	2	0
mdb222	3	3	1	2	4	0	3	3	0
mdb223	12	19	1	2	29	0	7	22	0
mdb226	7	2	0	2	7	0	5	4	0
Total	81	103	8	15	169	0	63	119	2

Table 18: results from testing the three combined segmentors on images from the MIAS database

The segmentor which was used as the microcalcification detector was S^* with $t=1$. This returns all regions more than eight pixels in size identified by the Woods segmentor and which overlap with a region detected by either the Outlier or Boundary detectors. The results show that this segmentor gave the best performance, but still generated an enormous number of false positives. The images used in the evaluation described in Chapter Eight were however of better quality than those provided in the MIAS database and it was anticipated that acceptable results could be achieved.

7.2.2 The classification of calcifications

The following measures of calcifications were identified in the last chapter as having been successfully used in existing systems for the classification of microcalcifications or as having an intuitive interpretation:

- compactness of cluster shape
- mean grey level
- standard deviation of grey level
- mean separation of microcalcifications
- number
- compactness of shape
- standard deviation of compactness
- eccentricity
- standard deviation of eccentricity
- mean area
- standard deviation of area

The first draft of the knowledge base contained a total of 241 statements about the properties of calcifications and their clusters, relevant to the task of differential diagnosis. Of that total 124 were covered by the above properties. The bulk of the remainder related either to the ‘form’ of the microcalcifications – that they are punctate or casting, for example – or refer to shape information not captured in the above properties, such as the so-called ‘popcorn’ shape.

The implementation of the measures is straightforward. The area of a microcalcification is calculated by counting the number of pixels; the perimeter by counting the number of outside edges of the edge pixels. Separation is calculated as the Euclidian distance between centres. Eccentricity and Compactness are defined in the Appendices. The area and perimeter of a cluster are computed from the Delaunay triangulation¹ of the centre points of calcifications in a cluster.

In order for these measures to be used in the decision support tool, it was necessary to establish the range of values corresponding to each of the different symbolic terms employed in the knowledge base. For example, high compactness values are an indication of more rounded shape; in order to provide a mapping between the image processing measure ‘compactness’ and the symbolic representation ‘round shape’, a decision has to be made about precisely which values of the compactness measure count as ‘high’. This was done with reference to a database of images. The MIAS database used in the development of the microcalcification operator did not contain sufficient images for this purpose so another database was obtained: the USF/LNL database which consists of 200 mammograms digitised at a spatial resolution of 35 microns per pixel and with 12 bits per pixel. All the mammograms are of patients with calcifications, each case has been confirmed histologically and the position and outline of clusters and individual calcifications has been indicated by a radiologist.

The image-processing measures were applied to the outlines of the clusters and calcifications as indicated by the radiologists. The resulting database of measurements was used to set thresholds for the application of terms in the knowledge base, as described in Section 8.3.2.

7.3 The Representation of Medical Knowledge

The decision support tool must include a knowledge base which provides the information required by the extended decision procedure described in Chapter Four. This model represents decision-making as an activity involving the suggestion of candidates and the construction of arguments for and against them. The knowledge

1. The Delaunay triangulation is the dual of the Voronoi diagram. A Voronoi diagram divides a plane containing sites of interest into regions, such that all the points in a region are closer to the site within the region than to any other. The Delaunay triangulation consists of line segments which connect sites sharing a Voronoi edge [Fortune1987]. I define the perimeter of a cluster as the set of Delaunay edges which only participate in one triangle. .

base, therefore, must contain information which allows the candidate solutions for a particular problem or focus to be identified, and then enables the appropriate arguments to be constructed. For each decision there must be a ‘decision specification’: task-specific knowledge that identifies the elements in the knowledge base which are going to play the roles of foci, candidates and arguments. These are summarised in Table 19.

Decision	Focus	Candidates	Arguments
Feature Classification	Radiological Signs	Causes	Positive Signs
Selection of Next Action	Action	Successors	Indications
Assessment of Tissue Density	Screening Mammogram	Difficulty Classes	Difficulty Factors

Table 19: the decision specifications for the mammography workstation

These specifications provide a model for the knowledge base. The following entities must be represented: Radiological Signs, Causes, Signs, Plans and Actions, Images, Difficulty Classes and Factors. The following relationships between relevant entities must be represented: Causes, Positive Signs, Successors, Indications, Difficulty Classes and Difficulty Factors.

Two significant decisions were made in adopting the above model. First that information about the relative incidence of different diseases and the specificity of their various signs was ignored: no indication was given of the strength of different arguments. This issue is discussed in the conclusion. The other decision was that the classification of image features is a decision akin to diagnosis, but made in isolation from other clinical data: it is an assessment of the significance of the information on the image in terms of the probable cause of the feature. A number of authors propose that microcalcifications and breast masses be classified on the basis of the probability of malignancy. Le Gal proposes a five-way classification; the American College of

Radiology (ACR) system is based on a three-way classification; Heilbrun [Heilbrun1994] suggests that four classes are required. In view of this controversy, the approach taken was to consider the underlying reasoning behind these classifications, and give arguments for and against the different possible causes of microcalcifications. An assessment of the likelihood of malignancy could then be made on the basis of an understanding the most probable explanation for the calcifications. The consequences of this are discussed in Section 10.7.1.

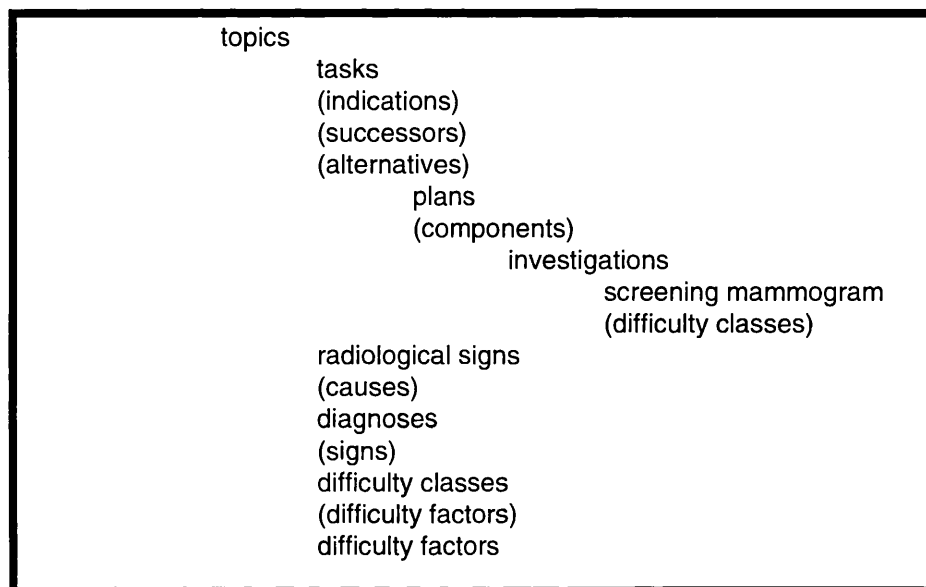


Figure 28: the knowledge model. This diagram shows part of the hierarchical structure of the knowledge base. The relationships associated with entities represented at or below each of the nodes are shown in brackets, only the highest node associated with each relationship is shown. The relationship 'alternatives' is not used in the decision procedure but serves, in the representation of the protocol, to identify points where there is a choice between successors.

In order to represent the knowledge required to support the characterisation of calcifications it is necessary to extract the information contained in the review articles summarised in the previous section and to cast it in consistent and appropriate terminology. The benign and malignant causes of calcifications must be enumerated and properties of calcifications which may serve as signs must be identified.

The construction of the knowledge base was carried out in three phases: first I identified the information required to support the differential diagnosis, next this information was organised into a consistent form and finally a draft knowledge base was evaluated by a panel of radiologists. This evaluation is described in Section 8.2.

The following texts on the differential diagnosis of calcifications were used in the construction of the knowledge base: [Tabar1983, Sickles1986, Lanyi1987, Bassett1992, Reeder1992]. Seventy-nine facts were obtained from Tabar and Dean, 34 from Sickles,¹ 106 from Lanyi, 33 from Bassett and 69 from Reeder. Three articles reviewed in Section 6.3.2 were not used, two because they did not attempt to describe all kinds of calcifications in terms of their cause [Le Gal1973, Dilhuydy1994], and one because it was not available until after the knowledge base was constructed [Monsees1995]. The total of 321 facts then had to be organised into a consistent knowledge base.

In order to remove duplicates and synonyms, it was necessary to settle on a set of unique terms. The most appropriate basis for this was the standardised reporting system proposed by the ACR [Kopans1992, D’Orsi1993]. This is the only published attempt to provide a structured terminology for mammography. According to this system, a radiologist’s report of a set of mammograms should consist of a pertinent clinical history, a description of the type of study performed, an indication of whether a comparison with previous studies was carried out, a summary of the distribution of tissue in the breast, a description of findings and an assessment. The features that should be used to describe breast masses and calcifications are also set out, as well as the set of possible assessments: recall from screening; normal; benign findings;

1. The article by Sickles was only located after the first draft of the knowledge base had been constructed, the facts obtained from it were added to the knowledge base as part of the revision process described in Chapter 8.

probably benign; suspicious or malignant. The ACR also publishes numerical codes for a set of radiological diagnoses [ACR1992]. There are nine different topics:

- normal, technique, anomaly
- inflammation
- neoplasm, neoplastic-like condition
- effect of trauma
- metabolic, endocrine, toxic
- other generalised systemic disorder
- duct disease, mammary dysplasia, cutaneous lesion of breast, lymphadenopathy
- miscellaneous
- other

The terms employed by the ACR were used in the draft knowledge base, with some additions to allow the inclusion of facts relating to findings not mentioned by the ACR. Some terms were changed to reflect common British usage, for example the term ‘intraductal carcinoma’ was replaced by ‘intraductal carcinoma’ or by ‘ductal carcinoma in situ’ on an assessment of whether the author intended to indicate ductal origin or the absence of invasion into the surrounding tissue. The term ‘scirrhous carcinoma’ was replaced by the more clearly defined term ‘invasive ductal carcinoma’.

Other changes reflect the difference between a standardised reporting scheme and a knowledge base. The term ‘calcification’ appears in the ACR scheme as a sub-kind of ‘effect of biopsy’ and also under ‘miscellaneous’. Kinds of ‘calcification’ include ‘benign’ and ‘malignant’, but there are also several other types, so that some benign calcifications appear in the classification as sub-types of ‘benign calcifications’, but others appear under a different heading. This may be satisfactory in a reporting

scheme, where the term 'benign' is used only to describe one particular category of characteristically benign calcifications, but it is not appropriate in a knowledge base where a clear structure is required since the position of a term in the hierarchy has semantic import. It is also necessary to be clear about whether an entity is a diagnosis or a description of a sign. For example, 'vascular calcifications' serves in the reporting scheme as a diagnosis, since it is an adequate description of a characteristic form of benign calcification. This, however, blurs the distinction between description and diagnosis. I use the term 'vascular calcification' as a diagnosis, a benign form of mammary dysplasia which can cause calcifications.

The facts were each represented as instances of the relationship 'positive sign' (information about negative signs could be included but in practice descriptions tended to concentrate on the characteristic features of microcalcifications associated with a cause):

'positive signs' of X include Y

The reverse of this relationship was also used in the knowledge base:

causes of Y include X

In each case the X variable was replaced by a term from the hierarchy of diagnostic terms and the Y variable was a structure of the form:

calcifications with value V for property P

The 12 terms used to describe properties and the 73 terms used to describe the values of properties were again a compromise between the desire to build on the standard terms in the ACR lexicon and the need to express all the information contained in the various review articles and monographs. Once the various synonyms, duplications and idiosyncrasies had been eliminated, and signs common to different sub-types were recorded only at the level of the super-type, the draft knowledge base consisted of 241 different facts about signs. This draft knowledge base was then greatly

simplified following the evaluation described in Section 8.2. The final version of the knowledge base is included in the Appendices.

There are two clear weaknesses in the way knowledge was represented. First, it was not possible to represent conjunctions of properties. So if ‘large eggshell calcifications’ are a sign of fibroadenoma then either ‘large calcifications’ and ‘eggshell calcifications’ are recorded as separate signs or a new term – ‘large eggshell’ – must be invented. Secondly no information about the relative value of signs was recorded. The overwhelming advantage of the approach is its simplicity. Users can view the hierarchy of diagnoses and see the signs associated with each; they also readily understand what is meant by presenting a simple statement about a sign as an argument in favour of a diagnosis.

7.4 The Representation of the Protocol

Computerised support for protocol-based care is currently one of the most fruitful areas of work in medical decision support. One of the aims of this thesis has been to explore the potential for the computerisation of imaging protocols. I identified a number of different protocols in the area of mammography which differed not only in content but also in their status within the institutions in which they were applied.

Two of the protocols were described in the literature reviewed in Section 6.3.2 [Dilhuydy1994, Monsees1995]. They took the form of simple flow charts (see Figure 26 and Figure 27) describing how patients with different categories of suspicious microcalcifications should be handled. The third was a guideline published for the benefit of Dutch general practitioners and summarised guidelines for the referral of patients with clinical symptoms. One protocol was a document, written by one of the surgeons at a breast cancer clinic (the Parapet Clinic of the King Edward VII Hospital,

Windsor) describing how patients there were handled. The final protocol, that followed at Guy's Hospital, was set down for my benefit. The protocols from the Parapet and Guy's were represented using sets of diagrams which correspond to a formal notation being developed at the ICRF for the representation of protocols [Fox1997]. The Guy's protocol was converted into a knowledge base and used in the prototype. The diagrams and the knowledge base describing the protocol are included in the Appendices.

Two issues needed to be addressed in relation to the representation of the protocol: first the terms and the relationships that needed to be represented and secondly the formalism to be used. The representation was to be used as the basis for decision support and therefore did not need to include every action which was to be performed, nor a complete description of every significant action. The protocol was represented at a level which allowed a convenient summary of the patient's current state and the available future options to be presented to a busy clinician. Thus, for example, the representation of the screening protocol records that a screening mammogram is to be taken, and specifies that this should consist of the mediolateral oblique view, but it does not state that left and right breasts are both to be imaged, nor the details of how the images are to be taken and how they are to be processed or to whom they must be sent. An alternative policy would have been to consider how many of the activities required by the protocol could be supported, or indeed performed, by computer and to represent them. This would have meant specifying, for each action, all the information which was to be recorded and how it was to be handled as well as providing an interface to software concerned with handling of appointments and worklists at imaging facilities and so forth. This did not seem appropriate for a prototype intended solely to test an approach to the provision of decision support.

The description of actions and their relations dealt only in the detail required to enable the desired decision support to be provided. Information about the sequence in which actions are to be performed and the circumstances in which an action may not be performed, or in which another action would be preferred, was recorded. The represen-

tation used was similar to that in the medical knowledge base and it consisted of triples expressing a relationship between two entities:

successors of X include Y
indications of X include Y
abortion conditions for X include Y

An additional specification which named a plan was included for some facts. This fourth term represents the scope of the fact and indicates that the fact is only to be used, only to be considered true, while the named plan is being executed.

7.5 Overview of the Prototype

I have described the basis of the decision support, the generic architecture and the construction of the image processing and knowledge base for the mammography workstation. The aim in this short section is to convey a sense of the complete application, constructed from these various parts.

The user interface provides a set of buttons for accessing the various display facilities and a set of sliders for altering the display characteristics of the image. The buttons provide access to menus which allow the display of:

- summary patient records for all the patients in the system (the term patient as used here includes the clients of the screening programme)
- a browser of the knowledge base of medical facts
- a browser of the representation of the protocol
- a set of reference images, illustrating typical presentations of microcalcifications and examples of histology slides
- a set of text files, providing access to abstracts of papers about mammography in the on-line version of the *British Medical Journal*

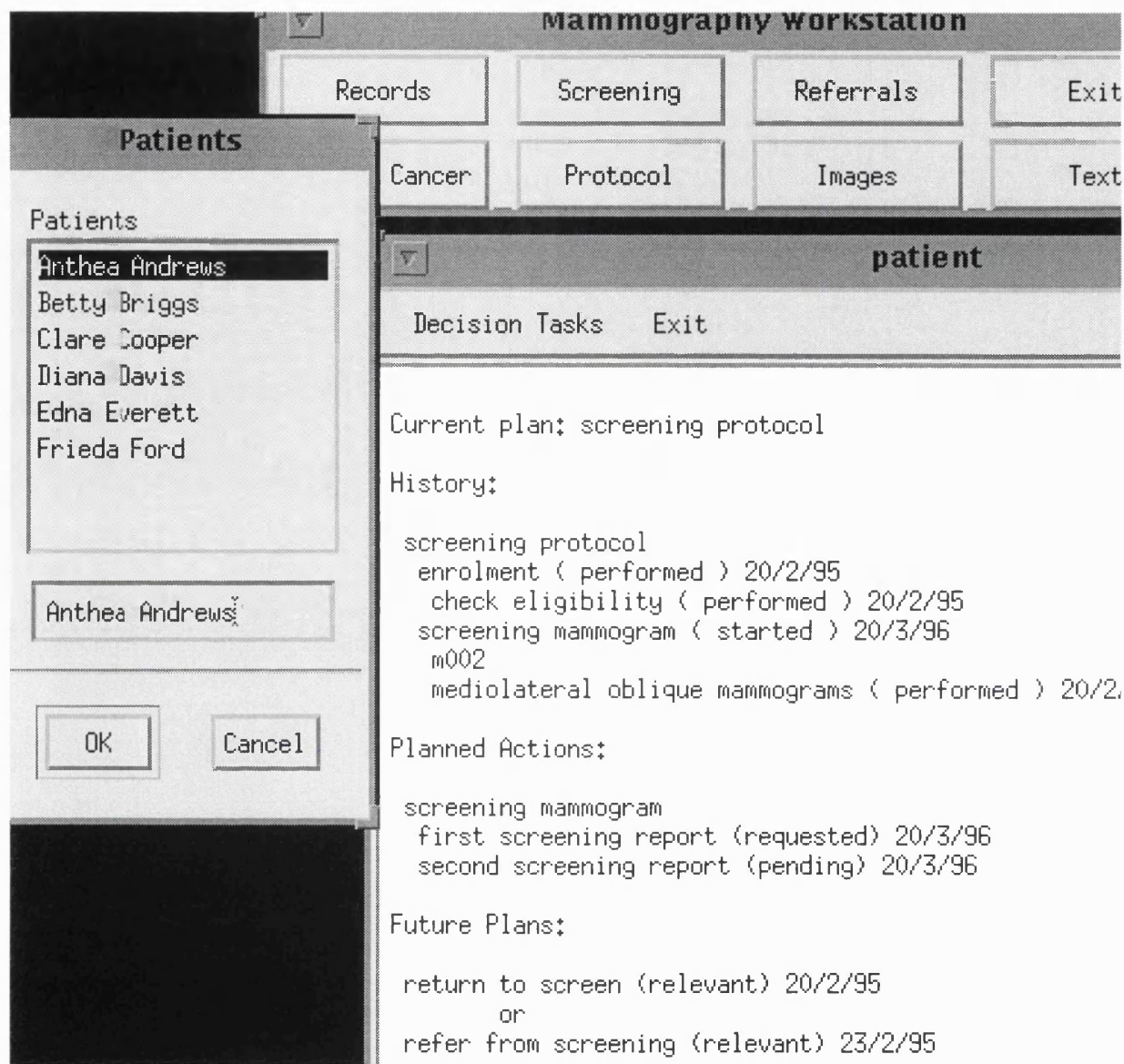


Figure 29: the interface of the mammography workstation with a summary of a patient's record

When the user selects a patient, the relevant set of images is displayed together with a summary of the protocol being followed. The images are displayed at a resolution of 900 by 1,500 pixels (140 microns per pixel) but images with a resolution of 3,600 by 6,000 pixels are also loaded into memory and sections of these can be displayed superimposed on the reduced image, using a device which employs the

the arguments in favour of possible candidates which are supported by measurements on the image.

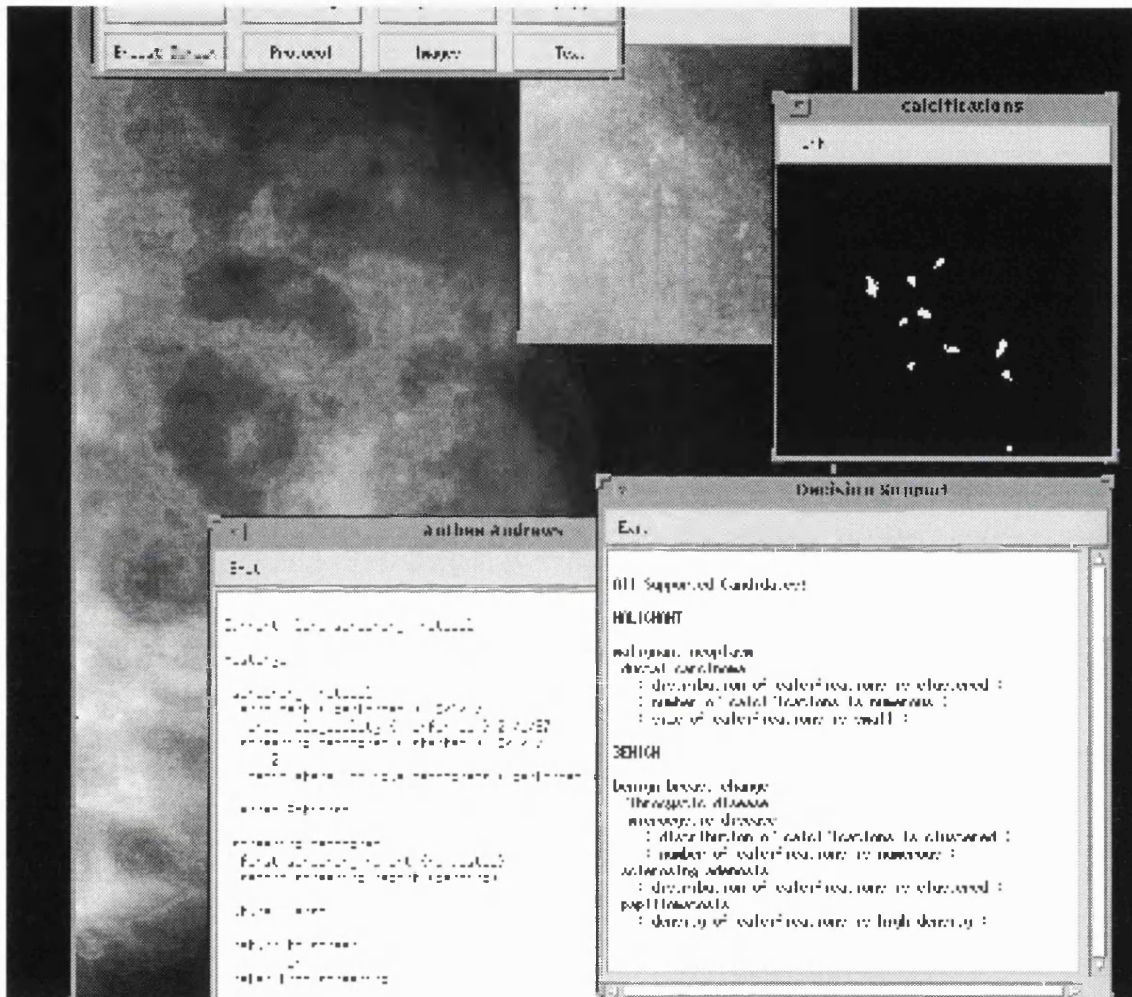


Figure 31: the display of decision support. The detected calcifications are shown in an additional window. The arguments in favour of the supported candidates are displayed according to the hierarchy of diagnoses.

The system does not take decisions; its role is to present information relating to the decisions which must be made about a particular patient. However, once a decision has been taken, and recorded in the system, this information is used by the task manager to update the information held on the patient record about the protocol being followed.

7.6 Conclusions

The implemented prototype is not a piece of software which could be used in a clinical environment without some further development. It has not been engineered with sufficient regard to the safety-critical nature of such use and some aspects of the user interface might be difficult for clinicians unfamiliar with this kind of technology, there is no facility for undoing changes when information is entered, for example. The purpose of developing this prototype was to establish the credibility of the approach and that a prototype was implemented using the generic architecture and incorporating the extended decision procedure is a valuable demonstration of the practicality of these ideas. The evaluation described in detail in the next chapter attempts to assess whether the information provided does assist in the classification of calcifications.

CHAPTER EIGHT

EVALUATION

8.1 Introduction

The aim of this research is to consider how image-processing techniques could be combined with knowledge-based reasoning to assist radiologists making decisions about images. The prototype which was most fully developed in the course of the research focused on the differential diagnosis of calcifications, providing assistance for a radiologist attempting to decide if calcifications detected on an image are likely to be benign or malignant. The constraints imposed by the need to develop the prototype as a piece of research meant that it would have been impractical to carry out a clinical trial of the system. I therefore performed a number of separate evaluations of this system: to assess the accuracy of the knowledge base, to assess the appropriateness of the image-processing measures and to assess the value of the tool as an aid to decision making.

8.2 Accuracy of the Knowledge Base

The first draft of the knowledge base included 241 statements of the form:
positive signs of PROCESS include calcifications with VALUE for PROPERTY

for example:

positive signs *of* plasma cell mastitis *include* calcifications *with* large size

These statements were obtained, as described in Section 7.3, by transcribing information from a number of standard texts on the differential diagnosis of calcifications. This generated a total of 321 facts which were then rewritten in a uniform terminology based on the ACR guidelines for standardised reporting. Once this had been done and duplicates removed, 241 facts remained.

The first stage of the evaluation was intended to establish that the information in the knowledge base was accurate, helpful and complete. A questionnaire was used to solicit assessments from radiologists. First, a short explanation of the context of the study was provided in order to ensure that, for example, the notion of a ‘sign’, as used in the system, was understood. The radiologists were then asked if they agreed that the categories of diseases, other processes, signs and properties are appropriate before being asked to rate each of the 241 facts on a scale from one to five, to record the extent to which they agreed or disagreed with it. The facts were grouped according to diagnosis. The radiologists were also invited to propose new properties. Twelve expert radiologists agreed to participate, two later withdrew. The reasons they gave for withdrawing were shared to a lesser extent by some of those who did participate, and it is worth recording their reactions. One said that she did not distinguish benign from malignant calcifications on the basis of conscious reasoning about properties and would therefore be unable to give a clear opinion of the accuracy of statements associating properties of calcifications with diagnoses. The other felt that the terms used to describe the properties of calcifications were not adequately defined, and therefore the questions could not be answered with confidence. Three radiologists failed to return the questionnaires in time to be included in the analysis, which was therefore based on seven sets of responses. The responding radiologists had an average of eight years

experience in mammography and were currently viewing an average of 12,000 mammograms a year.

In order to assess the extent to which radiologists agreed on their assessments, a statistic known as weighted kappa was computed for each of the 21 comparisons between pairs of radiologists. Kappa is a measure of inter-rater agreement. If p_o is the proportion of cases on which two raters agree, and p_e is the proportion of agreements expected by chance, then kappa is given as:

$$K = \frac{p_o - p_e}{1 - p_e}$$

Kappa is a measure of the difference between the observed proportional agreement and that expected by chance, expressed as a proportion of the possible scope for doing better than chance. Weighted kappa is a refinement which takes into account the degree of disagreement. So, a zero value for weighted kappa indicates agreement no better than chance while perfect agreement will have a weighted kappa value of 1. The mean value of weighted kappa for the 21 comparisons between radiologists was 0.24, indicating relatively low inter-rater agreement (see Altman [Altman1991] for a description of how to assess values for Kappa).

Four factors help to explain this.

- Ambiguities were introduced by the removal of statements from their original context. For example some statements had appeared in articles as possible but not typical explanations of calcifications. Some respondents disagreed strongly with such statements, which appeared in the questionnaire without the caveats that had appeared in the original articles, others agreed but noted that they would be “perverse” diagnoses.
- There were disagreements about terminology – one respondent marked as

“meaningless” a diagnosis for which other radiologists found statements with which they agreed.

- There were genuine disagreements about the medical knowledge; for example the statement that calcifications caused by lobular carcinoma in situ had a clustered distribution was given ratings of 5, 4, 2, 1, 1 and left blank by the other two respondents, one of whom noted that it was a “controversial statement”.
- Radiologists had very different response styles, one respondent only gave scores of 1, 3 or 5.

It is worth noting that there were relatively few overlaps in the various review articles and monographs from which facts in the knowledge base were extracted. Five different authors describing the same domain produced a total of 321 facts, containing only 80 duplications. This suggests that there is considerable variation between authors. This not largely a matter of terminological differences, nor does it seem to be a matter of contradictory positions; it seems mainly to be due to differences in approach with different authors focusing on different properties (such as cluster shape or location). Authors also vary in that some describe a prototypical appearance which can unambiguously be associated with a particular diagnosis, while others will describe a broader range of properties which *may* be associated with a diagnosis.

The results of the questionnaire were used to revise the knowledge base in a number of ways. The comments on terminology were used to remove a large number of possible causes of calcifications which were considered to be obscure, unhelpful or meaningless. Some of the terms used were changed at the suggestion of radiologists. A number of other facts were removed: if either three or more radiologists objected to the terminology or for some other reason failed to enter a response on the questionnaire, or if the average of rating given was not greater than three, then a statement was removed. After this 93 statements remained. These are listed in the Appendices.

Many of the facts which were removed relate to diagnoses which these seven experienced radiologists could not remember ever having encountered. Their inclusion in review articles must be due in part to the desire of the authors to be absolutely comprehensive, and must in part reflect epidemiological differences (the American College of Radiology includes gunshot wounds and ruptured breast implants as possible diagnoses, but the incidence of these conditions in 50-64 year old women in South East England cannot be expected to be significant).

Table 20 shows the mean scores on the draft knowledge-base, and on the facts from the draft which survived into the final version.

	A	B	C	D	E	F	G
1st draft	3.5	3.2	2.6	2.9	3.3	3.1	3.0
final draft	4.2	4.5	3.9	3.9	4.1	3.6	3.6

Table 20: the mean scores for the radiologists' ratings on the initial and final drafts of the knowledge base

8.3 Association between Knowledge and Image Processing

In the prototype decision aid, image processing is used to provide information which will assist in the differential diagnosis of calcifications. In order to establish which measures might usefully be employed, I reviewed, in Section 6.5, the literature on the use of image processing in systems designed automatically to classify calcifications as benign or malignant. Several authors have considered this problem and an analysis of the measures they found useful identified the following:

- number

- mean area
- standard deviation of area
- mean grey level
- standard deviation of grey level
- mean contrast
- compactness of shape
- standard deviation of compactness
- eccentricity of shape
- standard deviation of eccentricity of shape
- compactness of cluster shape
- mean separation of calcifications

The suitability of these measures was tested in two ways, described in the following two sub-sections. First the measures were applied to a selection of images from the USF database to see if they performed roughly as expected. After this, informal and subjective evaluation the measurements were compared with ratings obtained from a panel of radiologists.

8.3.1 An informal evaluation of the image-processing measures

In the first evaluation the measures were applied to 44 image sections selected from images in the USF database. The sections were selected to include all the available instances of malignancy and the broadest possible range of benign presentations. The association between the first three measures and the properties ‘number’, ‘size’ and ‘variation in size’ used in the knowledge base was assumed not to be problematic.

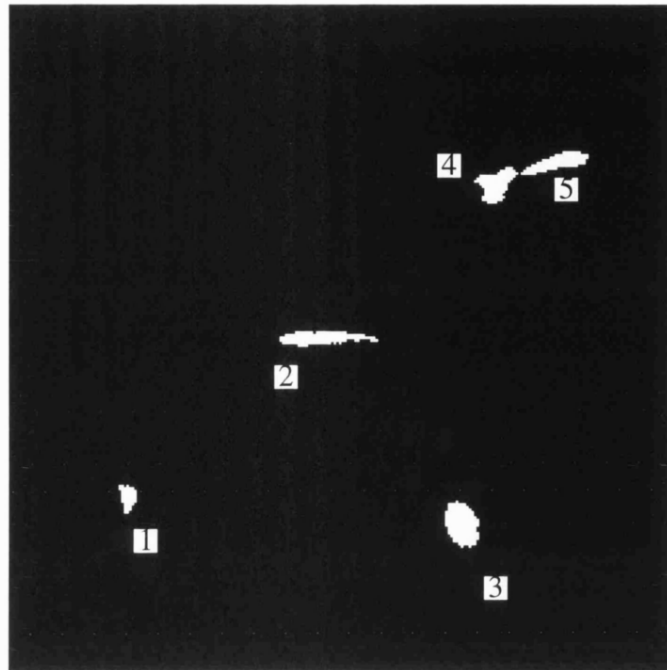


Figure 32: the segmentation of a cluster of calcifications described as pleomorphic in the USF database

Number	Area	Perimeter	Compactness
1	51	38	0.44
2	143	98	0.18
3	171	66	0.49
4	125	66	0.36
5	116	70	0.30

Table 21: values of area, perimeter and compactness for the five calcifications depicted in Figure 32

The terms used to describe calcifications tend to stress a distinction between linear calcifications of ductal or vascular origin, and round calcifications of lobular origin. Compactness (the ratio of area to perimeter), which is commonly used as a measure of shape, might seem an appropriate measure here since round objects have a

relatively short perimeter for their area. The data in Table 21, however, show that calcifications (such as number 4 in Figure 32) which have a slightly concave perimeter have compactness values which are not far removed from those – such as number 5 – which seem much longer and thinner. A further problem is that faint calcifications against a background of dense tissue tend to appear – if the segmentation is erroneous – as having convoluted perimeters.

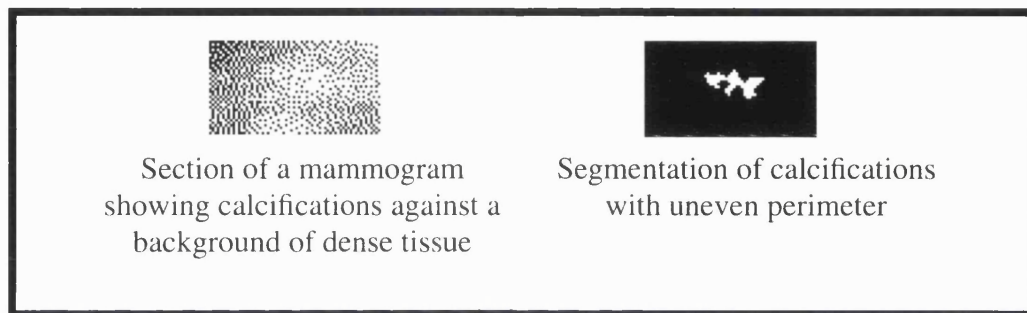


Figure 33: the segmentation of calcifications leads to errors which particularly affect measures based on the perimeter of the calcification

Eccentricity seemed a better way of distinguishing linear from rounded calcifications, since it is based on the ratio of the major to the minor axis. The equations used to calculate eccentricity, as defined in the Khoros image-processing package, are given in the Appendices. They are based on the moments about the centroid. These moments are not invariant to rotation, nor is the calculation of eccentricity based on them. Figure 34 shows two different orientations of calcification, and the dramatically different calculations of eccentricity produced. I therefore decided to consider a measure based on Hu's moments, which are invariant to rotation. The measure chosen was that of elongation, defined by Dunn and Brown [Dunn1986] and is defined in the Appendices.

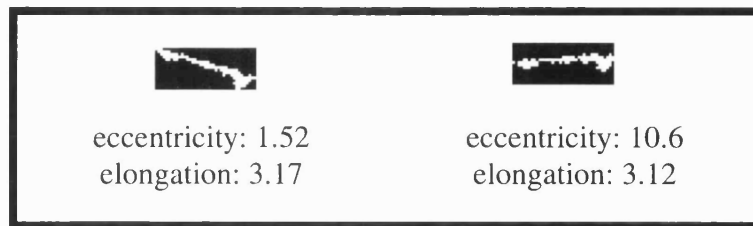


Figure 34: values of eccentricity and elongation for the same calcification at two different rotations

The most interesting questions arise from the use of compactness as a measure of cluster shape. Malignant clusters are described in the knowledge base as linear, branching, propeller-shaped, triangular or as having wavy contours. One radiologist suggested the term ‘cat’s paw’. These shapes reflect the ductal origin of the cancer associated with the calcifications. In contrast, benign clusters are all associated with a round cluster shape, reflecting their lobular origin. It is not surprising that three investigators who have used the compactness of the cluster shape to distinguish benign from malignant calcifications, since the shapes associated with malignant disease will have a lower area to perimeter ratio, and hence will be less compact. Two of these investigators specify that they used the compactness *of the convex hull*. This is interesting because the convex hull of a cluster with a branching shape – in other words, one which has spread out along ducts in different directions – will approximate to a triangle, square or pentagon: shapes which appear relatively compact.

I measured the compactness of the cluster shape on a sample of images from the USF/LNL database (see Section 7.2.2 for a description of this database), and it appeared to be a relatively powerful indicator of malignancy, but with the malignant clusters being the most compact. An analysis of the variance in cluster compactness for the 34 samples in which three or more calcifications were detected showed that the difference between the mean compactness of the benign clusters and the mean compactness of the malignant clusters was significant at the 5% level. This difference disappeared when attention was restricted to clusters of five or more calcifications. This

suggested that the compactness of the convex hull was actually operating as a measure of the extent to which calcifications formed a cluster rather than as a measure of the precise shape: the ‘clusters’ which appeared to have low compactness were actually chance arrangements of three or four benign calcifications.

8.3.2 Agreement between radiologists’ ratings and image-processing measures

It was necessary to verify that the intuitive link between the properties mentioned by authors of texts on calcifications and the measures used in the analysis of images does in fact exist. This was done by comparing values for the above measures with radiologists’ ratings of a representative sample of calcifications from the database.

Three radiologists were asked to rate a sample of calcification clusters on a scale of 1 to 5 for:

- roundness of cluster shape
- uniformity of density
- variation in size
- degree of polymorphism
- degree of clustering
- roundness of calcification shape
- radiopaque density
- definition of outline
- size
- number

The agreement between these ratings and the measurement obtained by the image processing was calculated using the weighted kappa statistic described in Section 8.2. Note that the test is not for correlation – the object is not to test whether two distinct properties are related – but for agreement, essentially to see if the two methods of measuring properties produced similar results.

The aim here was to test the hypothesis (or set of hypotheses) that the image-processing is measuring the same qualities referred to by radiologists when using the terms. It was not anticipated that the measures would agree perfectly with the radiologists; the test could show only if the measures, or some of the measures appeared weak.

A separate issue was where to place the threshold on, for example, the measure of “roundness” which determines whether or not a calcification is considered as round in the decision procedure. I originally proposed to set the thresholds so that e.g. only the 25% most round calcifications would be classed as round, this figure was revised after the radiologists’ ratings were assessed.

The three radiologists viewed 38 calcification clusters from the USF database. The agreement between the radiologists was rather low. On certain measures the value of weighted kappa was extremely low: the value for variation in size is little better than that which would be expected by chance. Low values were also obtained for agreement on outline and the other measures based on variation. The levels of agreement were much better on the assessment of calcification shape, attenuation and number.

Measure	Radiologists 1 and 2	Radiologists 2 and 3	Radiologists 1 and 3
size	0.60	0.24	0.35
variation in size	0.028	-0.019	0.20
roundness of cluster shape	0.49	0.38	0.38
roundness of shape	0.57	0.57	0.54
variation in shape	0.25	0.21	0.22
attenuation	0.67	0.54	0.52
variation in attenuation	0.082	0.15	0.28
number	0.50	0.42	0.55
separation	0.29	0.50	0.18
outline	0.28	0.20	0.26

Table 22: weighted kappa values for agreement between the radiologists on the ratings of calcifications

This task is an unfamiliar one for the radiologists and one which they seemed to find rather difficult. Disagreement could arise for three reasons: about the contents of the images, about the calibration of the scale or about the meaning of the terms used to describe the properties. Differences in what the radiologists were able to detect in the images were particularly apparent – one of the radiologists asserted that there were no visible calcifications on one particular image. Differences of what might be termed ‘calibration’ were obvious in judgements of size: since benign calcifications can grow to considerable sizes, the radiologists weren’t sure exactly what would constitute a ‘large’ calcification. Differences in the interpretation of the terms used to describe the properties of calcifications were less easy to detect from the observations of the radiologists but may well have been present. The questions raised by these findings are discussed further in Section 10.7.1.

My intention was to use the radiologists' ratings to establish that the proposed image-processing measures were actually measuring the properties that radiologists associated with terms in the knowledge base. The presence or absence of an association between the properties as perceived by the radiologists and as measured by image processing would be seen by plotting the measures (normalised to range from 1 to 5) against the average of the radiologists' ratings. The plots of average rating against measurement produced inconclusive results, largely because the average of the radiologists' ratings was not a terribly meaningful figure. These plots are shown in Figure 35 and Figure 36. The plots for size and elongation might be considered as evidence of association, but even there it is weak. It is clear that, if these findings are to be accepted, and there are other studies which support them (see Section 10.7.1 for a complete discussion), then a different approach should be taken to the development of a knowledge base to describe microcalcifications, one which doesn't assume a consensus on the meaning of terms.

In the light of these findings, it would have been desirable to run a further study, giving the radiologists a clearer set of instructions. Such a study would not attempt to identify a consensus about the meaning of terms as applied to calcifications but would impose a fairly rigid set of meanings for the terms (for example, one might build a knowledge base around a classification scheme proposed by a single author and require the radiologists to learn the author's definition of the terms used in the scheme) so that the radiologists' responses could be to assess the extent to which the image processing served to measure those terms. Unfortunately this was not possible within the constraints of this project and so decisions about which measures to use and where to set the thresholds had to be made on the basis of clearly unsatisfactory data.

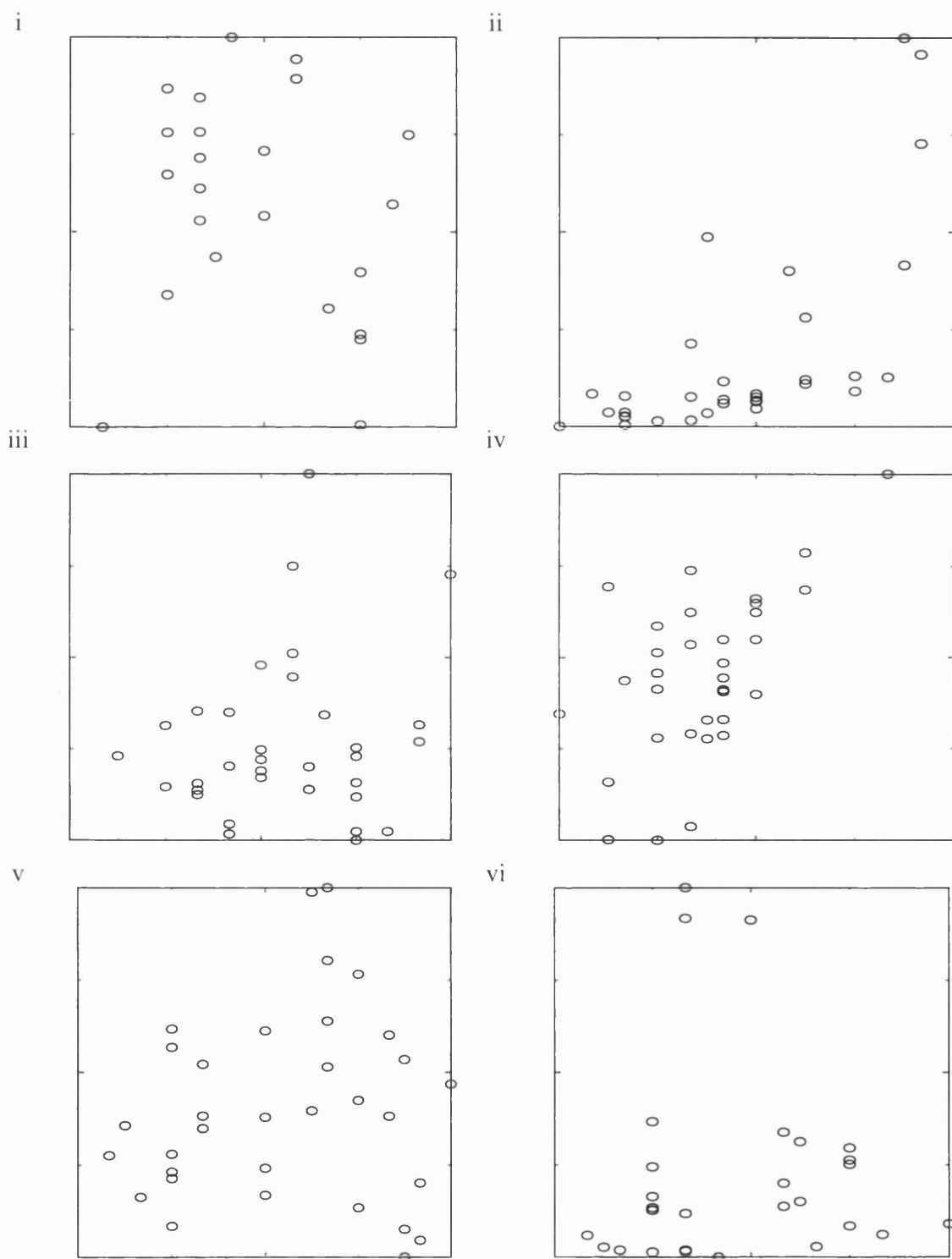


Figure 35: plots of mean radiologists' rating (x) against image-processing measurement (y) for (i) cluster compactness (ii) size (iii) outline (iv) elongation (v) attenuation and (vi) distribution

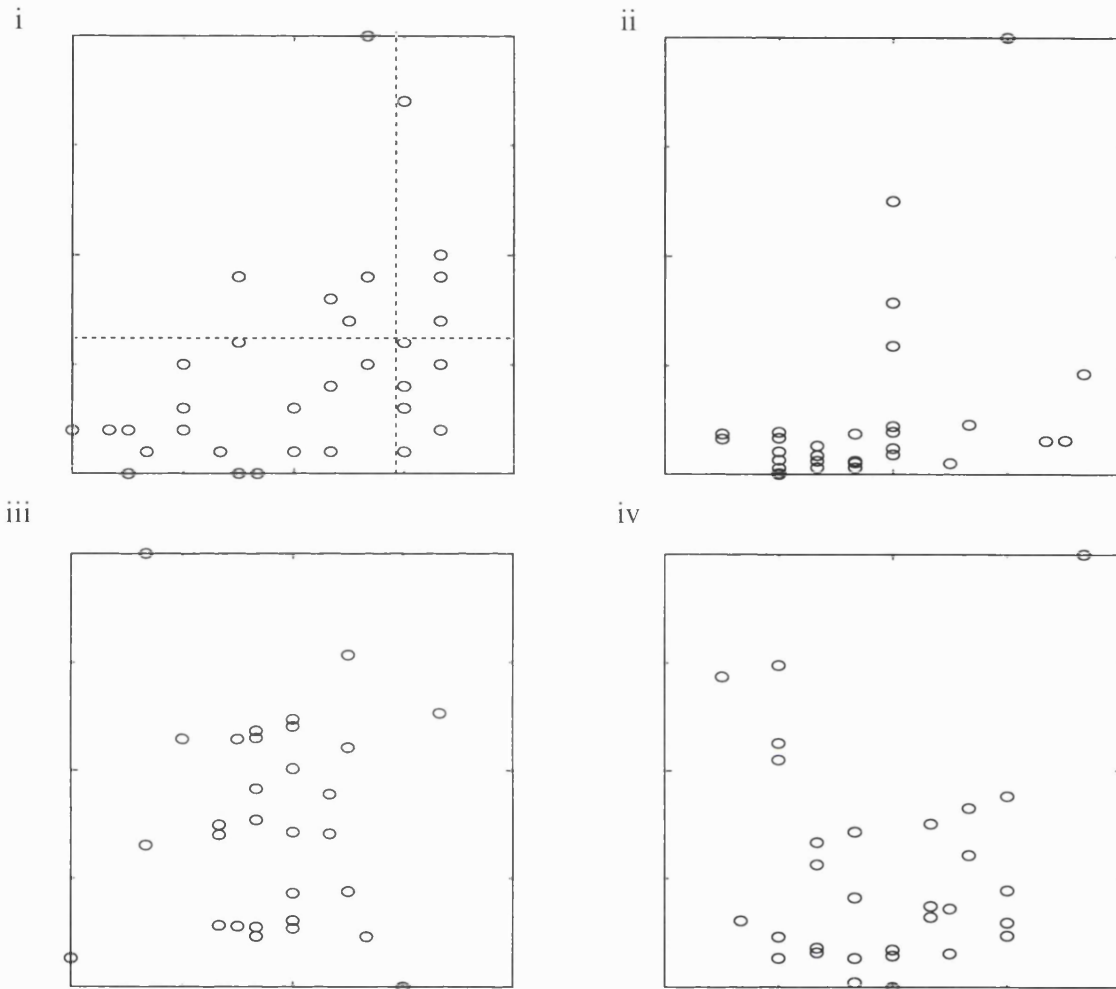


Figure 36: plots of mean radiologists' rating (x) against image-processing measurement (y) for (i) number (ii) variation in size (iii) variation in elongation and (iv) variation in grey-level. The threshold on the property number was set so as to classify a significant proportion of the images given an average rating of 4 or more – those to the right of the vertical dotted line in (i) – as numerous without including too many of those falling below the dotted line. The threshold was therefore set at a level corresponding to the horizontal dotted line.

It was decided to exclude the following measures, for which very low levels of agreement were found: attenuation, variation in attenuation, variation in size and

definition of outline. It was decided to include size and elongation which seemed relatively successful. It was also decided to include cluster compactness, distribution, number and variation in elongation despite their apparently poor performance, because they figured strongly in the literature on differential diagnosis.

Thresholds for the measures, above or below which calcifications would be classified as 'numerous' or 'elongated' for example – were set on the basis of the radiologists' ratings. Owing to the lack of a close association between the measures and the ratings, they had to be set at values which apparently classified incorrectly certain groups of calcifications. Consider the example of the plot obtained in Figure 36 (i) above for the property 'number'. A threshold was required which would allow a reasonable proportion of clusters with an average radiologists' rating of 4 and above to be classified as having 'numerous calcifications'.

8.4 Value of the Tool as an Aid to Decision Making

The ultimate aim of this research was to demonstrate that the approach can be used to improve decision making in image interpretation. It was therefore necessary to compare the performance of film readers using the system and film readers not using the system. The scope for detecting an improvement in performance will clearly be greater with less experienced film readers. A research project at the Jarvis Screening Centre in Guildford, England is attempting to demonstrate that experienced radiographers can be trained to read mammograms with the same accuracy as radiologists. This work is important since the volume of images generated by the UK's national screening programme threatens to create a demand for film readers which is not met by the supply of trained radiologists willing to work for the programme. A group of four radiographers being trained in film reading at the Jarvis Screening Centre agreed to take part in this study.

The study was performed at the Jarvis Screening Centre during the film reading course, after training sessions on the interpretation of images containing calcifications. The radiographers were shown the 44 image sections from the USF database. Half the images were seen with decision support, half without. Two parameters were varied: which images were seen with advice and which without, and whether the images with advice were shown first or second. There were therefore four conditions. Each image was viewed only once by each radiographer: no one saw the same image with and without advice.

Each radiographer was given a brief explanation of the purpose of the study and the operation of the system. Where images were shown without advice the radiographers saw the section of the image, displayed at the original resolution of 35 microns per pixel. To assist in the interpretation, the entire image was also displayed, at a resolution of 140 microns per pixel with the region corresponding to the section clearly outlined. Images were displayed on a standard 800x1,100 pixel 8 bit colour monitor. Where images were shown with advice, the arguments in favour of all the supported candidates were shown. Since the advice was sometimes based on an unreliable segmentation, the segmentation produced by the calcification detector was also shown. The advice was displayed in a single text window, with the supported candidates shown under two headings – BENIGN and MALIGNANT – and indentation was used to show the hierarchical classification of diagnoses. The arguments in favour of each diagnosis were shown in brackets. The diagnosis with the most supporting arguments was shown first. Where there was an equal number of arguments, the order was determined randomly but the proportions were set to reflect the proportions of benign and malignant calcifications in the sample.

All Supported Candidates:
MALIGNANT
malignant neoplasm
ductal carcinoma
(number of calcifications is numerous)
(size of calcifications is small)
BENIGN
benign breast change
fibrocystic disease
cyst
(number of calcifications is numerous)
microcystic disease
(number of calcifications is numerous)

Figure 37: an example of the decision support advice used in the evaluation

Radiographers were asked to rate each image section on a scale of 1 to 5 according to the risk of malignancy associated with the calcifications it contained. They were told that every section would contain some calcifications. It was also stressed that the advice generated by the system was based on an imperfect analysis of the images and that sometimes it would be inappropriate.

The ratings obtained from the radiographers were used to plot ROC curves. Each point on the rating scale (1 to 5) is used as a threshold to classify image sections as benign or malignant. At each threshold the proportion of cancers correctly identified as cancers (true positive rate) and the proportion of benign calcifications classified as cancers (false positive rate) are calculated. Plotting the true positive against the false positive rate for the different thresholds gives a threshold-independent indication of the accuracy of classification. The bottom left corner of the graph corresponds to a decision threshold in which everything is classified benign, at the top right everything is

classified as cancer. A random classifier will connect these two corners with a straight line, a perfect classifier will pass through the top left corner.

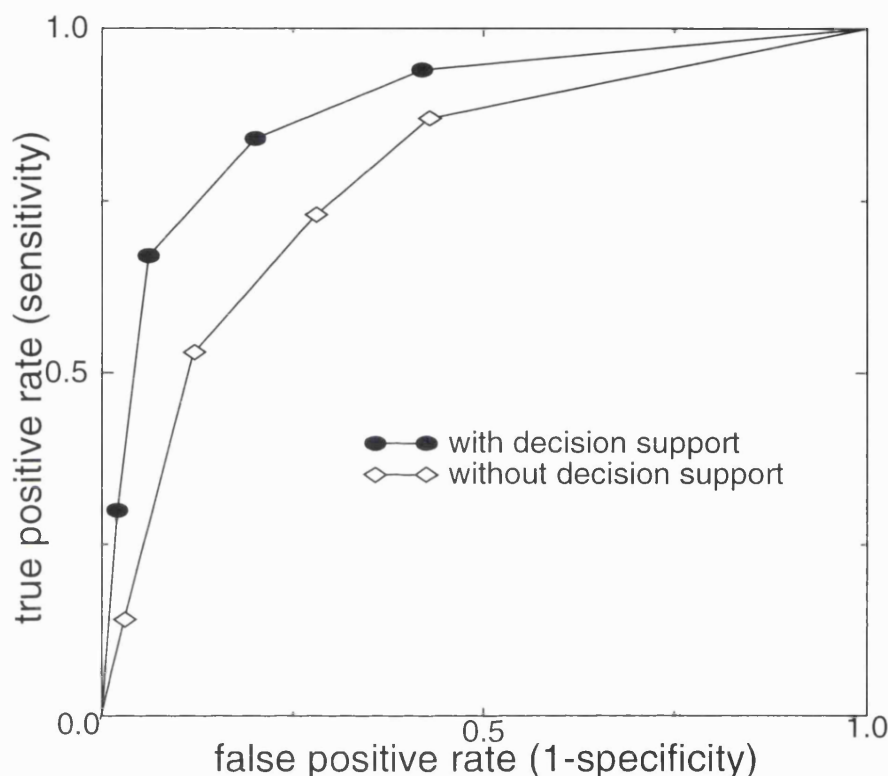


Figure 38: pooled ROC curves for four trained radiographers using the prototype decision aid with and without decision support

The areas under the two ROC curves were measured using ‘ROC analyzer’ software. Assuming no correlation between the images used in the two conditions, the probability that the curve with decision support is not greater than the curve for without decision support was calculated as 0.048. Hanley and McNeil [Hanley1983] provide a calculation for use in this kind of comparison study, where the same patients are used in the two conditions. In these cases it is inappropriate to calculate the standard error as:

$$SE (Area_1 - Area_2) = \sqrt{SE^2 (Area_1) + SE^2 (Area_2)}$$

since $Area_1$ and $Area_2$ are likely to be correlated. They propose using instead:

$$SE (Area_1 - Area_2) = \sqrt{\frac{SE^2 (Area_1) + SE^2 (Area_2)}{-2rSE (Area_1) SE (Area_2)}}$$

where r is a quantity representing the correlation introduced between the two areas by studying the same sample of patients. Hanley and McNeil provide tables which allow r to be determined from r_N and r_A , the correlation coefficient for ratings given to images of non-disease patients and the correlation coefficient for ratings given to images of disease patients. Kendall's tau [Ferguson1981] was used as the measure of correlation. Using this approach, a figure of 0.037 was found for the probability that the curve with decision support was not greater than the curve for without decision support. It can therefore be concluded that the difference between the two curves is statistically significant, and therefore that radiographers performed better using the decision support.

If image sections which radiographers rated 4 or 5 are considered to have been classified as malignant, then the sensitivity (proportion of cancers detected) of the radiographers is 0.68 with advice and 0.5 without; their specificity (proportion of non-cancers correctly classified) is 0.94 with advice and 0.88 without.

	with advice				without advice			
	TP	FP	TN	FN	TP	FP	TN	FN
Reader A	3	0	15	4	2	0	18	2
Reader B	7	2	13	0	3	4	14	1
Reader C	2	0	18	2	3	1	14	4
Reader D	3	2	16	1	3	3	12	4
Total	15	4	62	7	11	8	58	11

Table 23: true positive, false positive, true negative and false negative scores, with and without decision support for the 4 radiographers, considering ratings of 4 or 5 as a classification of malignant

The system failed to detect any calcifications on one image. On two others calcifications were detected but no advice could be generated. On the remaining 41 images, out of ten malignant cases, in six the most supported hypothesis was cancer, in one it was a benign disease and in three cases there was a tie. In the 31 benign cases, on six the most supported diagnosis was cancer, on 11 it was a benign disease and on 14 there was a tie. If only images for which more arguments were found in favour of cancer were classified as cancers and all others said to be benign, then the sensitivity of the system would be 0.55, and the specificity 0.82. This suggests that the radiographers using the system perform better than the system does by itself. The system is of course designed to present information about possible arguments to a decision-maker, not to actually take a decision itself. It is quite clear that simply counting the arguments constructed by the existing prototype would not represent an adequate classifier.

This test of the value of the prototype system as a decision aid provided a clear result supporting the hypothesis that trained radiographers perform better when using the system. Trained radiographers are relatively likely to be poor readers compared with radiologists and therefore it is easier to demonstrate improvement using a decision aid with this. It should be noted that the task they were invited to perform, classifying a lesion on the basis of a single image viewed on a standard monitor, is a difficult one and their low sensitivity will be due in part to the experimental conditions. Future work could attempt to show that radiographers using this system perform better than those using standard film-reading techniques, and to compare their performance to that of radiologists.

CHAPTER NINE

DECISION SUPPORT FOR ABDOMINAL CT

9.1 Introduction

In this chapter I describe a second prototype system to provide decision support for the interpretation of X-ray CT scans in the management of neuroblastoma. First the medical domain and the decisions are discussed, a detailed description is then given of the development of the image-processing operators, the symbolic knowledge base and the representation of the protocol. Finally a brief overview is given of the final prototype.

9.2 The Medical Context

Neuroblastoma is the most common extracranial malignant tumour in childhood, accounting for 8% of childhood cancers and 15% of cancer deaths in children. It originates in the neural crest cells of the sympathetic nervous system, but only one-fifth of children present with local or regional disease: the majority have lymph node involvement or distant metastases. The majority of patients with neuro-

blastoma present with an abdominal mass, often palpated in the flank which is characteristically hard, smooth and non-tender [Behrman1994].

9.3 Decision Making in Radiological Investigations of Neuroblastoma

Radiography, ultrasound, CT, MRI and scintigraphy are used in the diagnosis and management of neuroblastoma. In the next sections, the potential for computerised decision support is considered in relation to the problems of diagnosing, staging and assessing response to treatment.

9.3.1 Diagnosis

Various imaging modalities are useful in the diagnosis of neuroblastoma. MRI, CT and scintigraphic studies are the most important but the first imaging investigation for a child with an abdominal mass is usually ultrasound. Intravenous urography may be also be recommended. A plain chest X-ray is sensitive for the diagnosis of thoracic neuroblastoma.

Where an abdominal tumour is identified the principal diagnostic problem is to differentiate between neuroblastoma and Wilms' tumour. Imaging characteristics important in establishing the diagnosis include the site of the tumour and calcification . Wilms' tumour arises in the kidney and can be difficult to distinguish from a neuroblastoma which is invading a kidney. However, neuroblastoma more frequently displaces the aorta and Wilms' tumour often displaces and distorts the urinary collection system. Wilms' tumour may have a pseudo-capsule and is often lobulated, whereas the outline of the neuroblastoma is much harder to identify. Up to 85% of neuroblastomas are calcified, compared to 10% of Wilms' tumours. Rim calcification is specific for neuroblastoma.

- Ultrasound can often differentiate neuroblastoma from Wilms' tumour because of the extra-renal location and heterogeneous echogenicity of the former. Ultrasound is highly operator-dependent and studies have shown it to be less accurate than CT in detecting local spread.
- Radiography is sensitive for the detection of thoracic but not abdominal neuroblastoma.
- CT has been found to be the single most sensitive test for the detection of neuroblastoma. CT will detect calcification in 85% of all tumours. Contrast-enhanced CT may demonstrate the intraspinal extension of a tumour.
- MRI provides good anatomical detail, without the injection of an intravenous contrast medium. It does not show calcification however, and is less useful than CT if the diagnosis is in doubt [Berdon1992].

According to the International Neuroblastoma Staging System (INSS) [Brodeur1993], a diagnosis of neuroblastoma is established if:

- 1) An unequivocal pathological diagnosis is made from tumour tissue by light microscopy (with or without immunohistology, electron microscopy, increased urine or serum catecholamines or metabolites);*
- 2) Bone marrow aspirate or trephine biopsy contains unequivocal tumour cells (e.g., syncytia or immunocytoologically positive clumps of cells) and increased urine or serum catecholamine or metabolites.*

9.3.2 Staging

The International Neuroblastoma Staging System [Brodeur1993] defines six stages, listed in Table 24.

Stage 1:	Localised tumour with complete gross excision, with or without microscopic residual disease; representative ipsilateral lymph nodes negative for tumour microscopically (nodes attached to and removed with the tumour may be positive).
Stage 2A:	Localised tumour with incomplete gross excision; representative ipsilateral non-adherent lymph nodes negative for tumour microscopically.
Stage 2B:	Localised tumour with or without complete gross excision, with ipsilateral non-adherent lymph nodes positive for tumour. Enlarged contralateral lymph nodes must be negative microscopically.
Stage 3:	Unresectable unilateral tumour infiltrating across the midline with or without regional lymph node involvement; or, localised unilateral tumour with contralateral regional lymph node involvement; or, midline tumour with bilateral extension by infiltration (unresectable) or by lymph node involvement.
Stage 4:	Any primary tumour with dissemination to distant lymph nodes, bone, bone marrow, liver, skin and or other organs (except as defined in stage 4S).
Stage 4S:	Localised primary tumour (as defined for stage 1, 2A or 2B), with dissemination limited to liver, skin and/or bone marrow (limited to infants < 1 year of age).

Table 24: the INSS criteria for the staging of neuroblastoma

These definitions appear to be reasonably succinct, however their interpretation requires additional information. For example, “resectability” is defined in the text, as removal without a radical surgical procedure, which means without removing vital organs, compromising major vessels or disfiguring the patient. An especially problematic area is the definition of the midline. The 1993 revisions to the criteria state:

... it was decided to consider the vertebral column as the midline and only tumours that extended by contiguous invasion to or beyond the opposite side of the vertebral bodies would be classified as stage 3. Although there will still be ambiguous cases, these can be considered to have stage 3 because they would be stage 3 unequivocally by the old definition. A grossly resectable tumour arising in the midline from pelvic ganglia or the organ of Zuckerkandl would be considered stage 1. A midline tumour that extended beyond one side of the vertebral column and was unresectable would be considered stage 2A. Ipsilateral lymph node involvement would make it stage 2B, whereas bilateral lymph node involvement would make it stage 3. A midline primary tumour with bilateral infiltration that was resectable would be considered stage 3. A tumour of any size with malignant ascites or peritoneal implants would be stage 3 (but a thoracic tumour with malignant pleural effusion unilaterally would be stage 2B).

This is clearly an area where confusion can arise.

9.3.3 Response to Treatment

Six different categories of response to treatment are defined [Brodeur1993].

Response	Primary	Metastases
Complete Response	No tumour	No tumour; catecholamines normal
Very Good Partial Response	Decreased by 90-99%	No tumour; catecholamines normal; residual ^{99m} Tc changes allowed
Partial Response	Decreased by 50%-90%	All measurable sites decreased by > 50%. Number of positive bone sites decreased by > 50%; no more than 1 positive bone marrow site allowed
Mixed Response	No new lesions; > 50% reduction of any measurable lesion (primary or metastases) with < 50% in any other; < 25% increase in any existing lesion	
No Response	No new lesions; < 50% reduction but < 25% increase in any existing lesion	
Progressive Disease	Any new lesion; increase of any measurable lesion by > 25%; previous negative marrow positive for tumour	

Table 25: the INSS criteria for the classification of response to treatment

The same tests used to determine the extent of disease should be used to determine response. In responding patients it is not necessary to repeat tests which were negative on initial evaluation. Patients who relapse should be assessed as if they were new patients. An assessment of overall response requires an evaluation of primary tumour and metastases, a patient with complete response in metastatic sites and partial response in the primary is classified partial response. Evaluations for response in newly diagnosed patients are recommended at the end of induction (3-4 months) at the end of

maintenance (8 to 12 months), before and after surgery, before marrow transplantation and if indicated clinically. Note that if surgery is performed it is important to report the response before and after in order to allow the relative contributions of chemotherapy and surgery to be assessed.

The response criteria are based on a decrease in size, which is based, by convention, on the product of the two greatest measurable diameters, not on volume.

It is not possible to give quantitative assessments of response on the basis of scintigraphy, besides scintigraphic changes lag behind histologic changes by weeks or months. Scintigraphy is used to determine if a lesion has appeared, progressed, improved or disappeared.

9.3.4 Conclusions

In assessing the potential for a decision-support tool based on image processing and symbolic reasoning, the following criteria must be considered:

- the difficulty of the decision
- the importance of the decision
- the role to be played by image processing
- the role to be played by symbolic reasoning

There are problems in the diagnosis of neuroblastoma, but these are not to do with the difficulty of interpreting images. The two decisions which provide the greatest scope for a decision aid of the kind considered in this thesis are the staging of the disease and the assessment of response to treatment, since these decisions are crucially dependent on the use made of information obtained from X-ray CT. It is apparent from

the explanations of the revisions to the staging and assessment criteria that the authors suspect that there is confusion in the application of these guidelines, for example in the definition of the midline. Another difficulty is that of obtaining accurate measurements of tumour volume. Eggli et al. found that the standard method of estimating volume from 2D measurements overestimated volume by 50% in 43 out of 45 CT scans of paediatric patients with thoracic or abdominal tumours [Eggli1995].

The staging and assessment decisions are both important since the lethal nature of the disease and the traumatic nature of the available treatments place stringent demands on the planning of therapy. Some stage 2 patients only require surgery [Philip1992] whereas stage 4 neuroblastoma may require 'megatherapy', a chemotherapy which is so damaging that patients require a bone marrow transplant afterwards [Pritchard1995]. An accurate assessment of the tumour is therefore crucial. A further demand on the management of abdominal tumours is the need to minimise the use of ionising radiation, and hence to ensure that CT scans are only performed when required and that the maximum use is made of the information obtained.

The two pieces of image processing which are required to support the use of CT in the staging and assessment of response to treatment are the location of the tumour with respect to the midline and the volume of the tumour. The consensus in the medical community over the definitions of classes greatly simplifies the business of constructing a knowledge base, since the published criteria serve as a single authoritative source. The amount of knowledge that can be put into the knowledge base is limited and the rules for staging and assessment are fairly concise. The value of a decision is perhaps that it can assist in the promulgation of the guidelines for staging and help to ensure consistent application.

9.4 Image Processing for the Management of Neuroblastoma

Two pieces of image processing were required to support staging and assessment of response: to detect the midline and to measure the volume of the tumour.

9.4.1 Detection of the Midline

The midline is defined as the vertebral column. The detection of the midline therefore requires the detection of the vertebral column. This is readily identifiable to the human eye as a large bright structure with a distinctive shape and location. The distinctive appearance of the vertebral column seemed the most promising basis for detection. I obtained five CT data sets of patients with neuroblastoma from the X-ray department at the Royal Marsden.



Figure 39: an example of a CT scan for a patient with neuroblastoma. The vertebral column is the large bright object with a hole in the middle at the base.

The simplest strategy for developing a model of the shape of the vertebral column seemed to be to select, by eye, a representative set of samples and merge them into a single image which would provide a template to be matched against the test images. This was done by selecting a total of 16 160x160 patches from the five data sets, adding together the pixel values at each point and dividing by 16. This simple approach, with no attempt to normalise or provide a more accurate registration of the different samples than that obtained by picking regions by hand, resulted in an image which, subjectively, seemed an appropriate template. The cross-correlation of this template with the images in the data set was used to identify the position of the vertebral column, as shown in Figure 40. This procedure was tested on the total number of slices (including those used to build the template) and correctly identified the centre in 90 out of 95 slices.

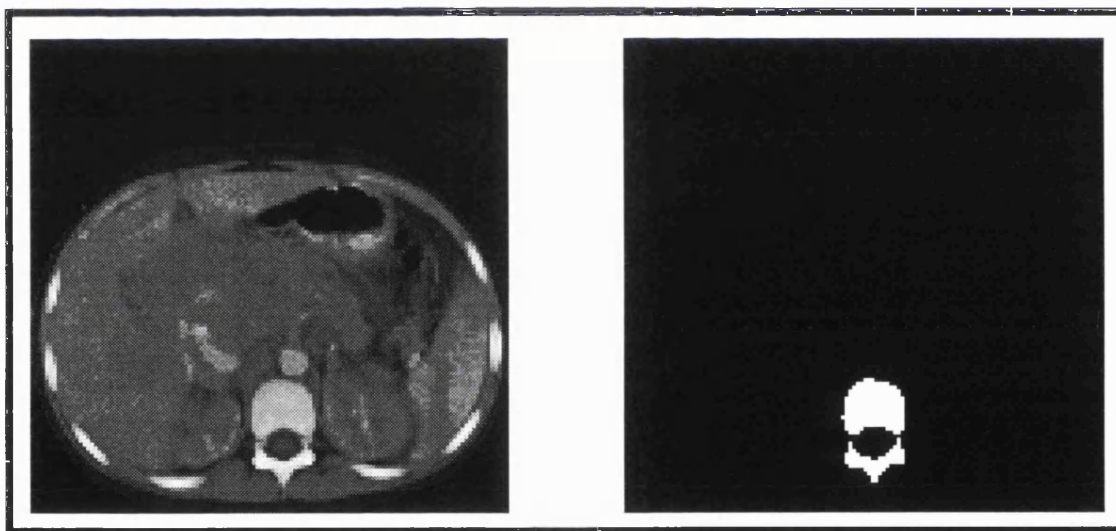


Figure 40: the output from the spinefinder program

Once the position of the column is identified, it is necessary to find its boundaries. This was done by thresholding. The position can be used to define a relatively small region of interest (160x160 pixels) within which the boundaries of the column can be guaranteed to lie. A histogram of the Hounsfield numbers within this region shows a distribution with a large peak covering the range of values corresponding to

soft tissue, another peak corresponding to air and a peak corresponding to the values for bone. The histogram was smoothed until there was only one peak above the largest peak and the threshold was set at the base of this peak. This procedure produced a sufficiently accurate segmentation of the vertebral column to allow the midline to be identified in 77 out of the 95 slices.

9.4.2 Measurement of tumour volume

In order to measure tumour volume accurately it is necessary to identify the boundaries between the tumour and adjacent organs. The task of decomposing an image into anatomically significant regions is called segmentation and is one of the fundamental problems in computer vision. Techniques for segmentation fall into three classes: thresholding, edge-based and region-based methods – and it is not uncommon to amalgamate them.

In threshold segmentation, a threshold, or a set of thresholds, is determined and connected pixels with brightness values above the threshold are assumed to represent a region distinct from those below it. The threshold may be determined by calculations over the whole image, or by local calculations within a region. It may be calculated from peaks or troughs in the grey-level histogram or the co-occurrence matrix (a histogram of the spatial co-occurrences of pixel values). Edge-based segmentation methods identify regions on the basis of discontinuities or steps in the image. This is done by matching a mask, or a template, of an appropriate edge with each point or region in the image, and at every orientation. Alternatively, the gradient, or the derivative of the gradient can be measured. A common technique is to convolve the image with the Laplacian (or second derivative) of a two-dimensional Gaussian, also known as the Marr-Hildreth [Marr1982] operator. This operation is computationally efficient because it is not specific to edges in a particular orientation.

Pavlidis [Pavlidis1992] wrote of work on edge-detection that comparisons of techniques which concentrated on the differences between the performance of various edge-detectors on the same image served, inadvertently, to conceal the differences in performance of the same edge-detector over different images. People often responded to this discovery by saying that improved edge-detectors were required for the more difficult images. Pavlidis suggests a different way of interpreting the results is to remember that: “what we were doing was not evaluating how well the edge-detector found steps (which is what they were actually designed to do) but evaluating how well they did in outlining objects”. That this is a very different task is easily understood when one thinks of the patterns of light that fall on an object in a room with many different light sources and reflecting surfaces. The same perceptual processes that allow humans to interpret such scenes are used when radiologists interpret medical images and may explain how they are able to identify the boundaries of objects on X-rays in the absence of a complete sequence of detectable discontinuities and despite the existence of detectable but irrelevant discontinuities away from object boundaries.

Region-based segmentation methods operate by partitioning an image into connected sub-regions so that every pixel is assigned to one region and every region is defined in such a way that all the pixels in it share some property. This may be done by selecting a pre-determined number of seed points and testing neighbouring points on some set of criteria to decide if they should be aggregated with the seed. The criteria may be based on the intensity and the spatial properties of the pixel (its proximity or connectivity to the seed), or the shape and textural properties of the region. Another technique of region-based segmentation involves setting criteria for the splitting and merging of regions and iteratively revising the segmentation until these constraints are satisfied. As with edge-based methods, region-based methods assume a coherence in the surfaces of objects which does not seem to be required for the successful identification of regions by the human eye. There is no segmentation method which is successful in the general case, nor is there a clear ‘best available method’ of image segmentation, nor any set of principles for determining which method will be the best

for a particular problem. Nor are there agreed criteria for comparing the adequacy of different segmentation techniques [Pal1993].

Although there is an enormous literature on the segmentation of medical images, there have not been many attempts to apply segmentation techniques to abdominal CT images. Karssemeijer et al. [Karssemeijer1988] describe a knowledge-based approach predicated on a simple model which records the spatial relationships of abdominal structures, and indicates which are most likely to be successfully identified. This model controls the invocation of thresholding and boundary smoothing operations, which are applied to the different abdominal structures in turn in order to identify the appropriate search area for the organ of interest. A similar approach is taken by Kobashi et al. [Kobashi1995] who again use thresholding methods to segment abdominal CT images. The segmentation operation is controlled by a system which uses knowledge of the expected positions, areas, relative grey-level, boundary strength and shape constraints to determine when an object has been adequately identified. Bae et al. developed a system to measure liver volume [Bae1993]. This relied on a thresholding operation carried out in the region of the abdomen where the liver is likely to be found and employing constraints of area, location and shape to select the appropriate region from among the candidates identified by the segmentation.

These systems have all been applied to the problem of segmenting images of normal anatomy. Identifying the boundary of a tumour, for which the concepts of normal shape, size or location do not apply, and which may contain tissue of highly variable grey-level (due to the presence of calcium), is a much more difficult operation, and one which is less easily guided by prior knowledge. It was therefore decided to provide a 'semi-automated segmentor', thus greatly reducing the manual effort required to identify the tumour boundary. Such a system is described by Sivewright and Elliot [Sivewright1994]: the user selects a voxel to serve as a seed point and the neighbours of the seed point are added to an empty boundary list. The members of the list are tested against an inclusion criterion and the neighbours of those that pass are appended

to the list (the ordering of the list is thus determined by how the list is built up and not by any property of the data). Two alternative stopping rules are permitted: one in which the maximum size of the volume is set by the user, and one in which the region grows until no more voxels can be found to pass the inclusion criterion. The inclusion criterion found to be most effective by Sivewright and Elliot was a threshold distance in brightness value from the original seed point.

Dellepiane and Fontana describe a technique, also based on growing a region from a seed point, which provides a simple basis for implementing a more sophisticated volume-growing algorithm. The technique is called 'intensity connectedness' and is based on earlier work by Rosenfeld [Rosenfeld1983]. The 'fuzzy connectedness' of a voxel with a seed point is defined as the minimum connectedness value on the optimal path between a voxel and the seed point, where the optimal path is that along which the minimum connectedness between neighbouring voxels is highest. The algorithm starts, as the Sivewright and Elliot algorithm does, with the neighbours of the seed point. These are placed on the boundary list, ordered according to their connectivity value. Each voxel is tested in turn, and if the connectivity exceeds a certain threshold its neighbours are placed on the boundary list in order of their connectivity values. Connectedness information is propagated from the seed point to every point in the image. Since connectedness always decreases, if the best connected candidate is considered first at every step, the algorithm is non-iterative: each point need only be considered once. The connectedness of a voxel with its neighbours is calculated as the difference in grey-level.

Both algorithms require a seed point and both construct a candidate list. They use different principles for ordering the candidate list and have different inclusion criteria. The same stopping rule can be used in both. I implemented both algorithms and compared them under two conditions with manual segmentations produced by an expert radiologist on three CT data sets. In the first condition, 'By Volume', one seed point was provided and the algorithms were used to implement a volume-growing

segmentation algorithm. In the second, 'By Slice', a seed point was placed in each slice of the data and the volume was computed as the sum of a set of region-growing operations. The results are presented in Table 26.

By Volume						
Data Set	Connect			Threshold		
	TP	FN	FP	TP	FN	FP
Set 1	41	59	152	25	75	49
Set 2	52	48	32	40	60	34
Set 6	52	48	28	59	41	8
Mean	48	52	71	41	59	30

By Slice						
Data Set	Connect			Threshold		
	TP	FN	FP	TP	FN	FP
Set 1	51	49	28	44	56	22
Set 2	61	39	26	53	47	17
Set 6	79	21	11	80	20	11
Mean	64	36	22	59	41	17

Table 26: results for the connectivity (Dellepiane and Fontana) and threshold (Sivewright and Elliot) algorithms tested on three CT data sets, under two conditions. In the first, 'By Volume', one seed point was provided and the algorithms were used to implement a volume-growing segmentation algorithm. In the second, 'By Slice', a seed point was placed in each slice of the data and the volume was computed as the result of a set of region-growing operations. The figures given are for the true positive (the region identified by both the algorithm and the radiologist), false positive (the region identified by the algorithm but not the radiologist) and false negative (the region identified by the radiologist but not the algorithm) fractions expressed as a percentage of the actual volume of the tumour.

The 'intensity connectedness' algorithm did slightly better, but neither algorithm gave a performance which one would expect a radiologist to consider impressive. Two points are worth noting. The first is that the radiologist was not terribly confident of her identification of the tumour boundaries. This suggests first that the gold standard against which performance was judged may not have been entirely reliable, and secondly that the task is not a simple one. The performance of the algorithms could have been improved by including some information about shape. As was noted earlier, tumours vary and there is no such thing as normal or expected shape. There are however, shapes which are inherently unlikely and further development of the more promising of the algorithms might produce a worthwhile tool.

9.5 The Representation of Medical Knowledge

As with the mammography system described in the Chapter Seven, the decision aid must include a knowledge base founded on the model provided by the specifications of the decisions it is to support. The decision aid described here was developed to support decisions of diagnosis, staging and assessment of response to treatment. The specifications of the decisions are given in Table 27.

The knowledge base must represent the following classes of entity: Diseases, Clinical signs, Radiological signs, Pathological signs, Symptoms, Classifications, Plans and Actions, and the following relationships between them: Causes, Clinical signs, Symptoms, Criteria, Criterion, Excluding criterion, Components, Successors and Indications.

Decision	Focus	Candidates	Arguments
Diagnosis	Abdominal mass	Causes	Clinical signs Symptoms
Staging	Neuroblastoma	Stages	Criteria Criterion Excluding criterion
Assessment of response	Neuroblastoma	Responses	Criteria Criterion Excluding criterion
Selection of next action	Current action	Successors	Indications

Table 27: the decision specifications for the neuroblastoma workstation

The information represented in the knowledge bases was taken from the document describing the INSS classifications [Brodeur1993] which constitutes a single authoritative source, thus avoiding some of the problems associated with the development of the mammography system. A diagnosis would not generally be made on the basis of the radiological interpretation alone, but information on the diagnostic significance of radiological information has been published in a small number of review articles [Berdon1992, Caty1993], and it was decided to include it for demonstration purposes.

The classifications defined by the INSS for the staging and classification of neuroblastoma are represented in the knowledge base as sets of necessary 'criteria' all of which must be true, or as sets of statements each defining a sufficient 'criterion', only one of which need be true for a tumour to be assigned to a given stage. The representation also includes a relation 'excluding criterion' which is sufficient to rule out a given stage or class of response. In contrast to the mammography knowledge base, the complexity of the definitions used in the staging and assessment classifications required that some concepts be defined in terms of other concepts. In effect a theorem

prover was required which could consider arguments composed using the logical operators ‘and’, ‘or’ and ‘not’. This was implemented as a layer between the Symbolic Decision Procedure described in Section 4.2 and the model of image interpretation. Complex arguments are constructed by potentially recursive calls to the representation of the staging and assessment classifications, and then analysed by a predicate which recurses through the construction, establishing the evidence for each atomic element within it and then generating a structure which represents the evidence for the complete argument, taking account of the logical operators used to compose it. The evidence is associated with a value, which in the implemented prototype is either true or false, either the evidence required was found or it was not; for example either the tumour is infiltrating across the midline or it is not.

```
/* Definition of stage 3 neuroblastoma
('abdominal mass'::resectability=unresectable AND 'abdominal mass'::infiltration='across the midline')
OR
('abdominal mass'::infiltration='localised tumour' AND 'contralateral lymph nodes'::involvement=involved)
OR
('abdominal mass'::site=midline AND 'abdominal mass'::infiltration=bilateral)
OR
('abdominal mass'::site=midline AND 'regional lymph nodes'::involvement=involved)
*/

criterion, 'stage 3 neuroblastoma', 'stage 3 neuroblastoma with infiltration'
criterion, 'stage 3 neuroblastoma', 'stage 3 neuroblastoma with lymph node involvement'
criterion, 'stage 3 neuroblastoma', 'midline stage 3 neuroblastoma with infiltration'
criterion, 'stage 3 neuroblastoma', 'midline stage 3 neuroblastoma with lymph node involvement'

criteria, 'stage 3 neuroblastoma with infiltration', 'abdominal mass'::resectability=unresectable
criteria, 'stage 3 neuroblastoma with infiltration', 'abdominal mass'::infiltration='across the midline'

criteria, 'stage 3 neuroblastoma with lymph node involvement', 'abdominal mass'::infiltration='localised tumour'
criteria, 'stage 3 neuroblastoma with lymph node involvement', 'contralateral lymph nodes'::involvement=involved

criteria, 'midline stage 3 neuroblastoma with infiltration', 'abdominal mass'::site=midline
criteria, 'midline stage 3 neuroblastoma with infiltration', 'abdominal mass'::infiltration=bilateral

criteria, 'midline stage 3 neuroblastoma with lymph node involvement', 'abdominal mass'::site=midline
criteria, 'midline stage 3 neuroblastoma with lymph node involvement', 'regional lymph nodes'::involvement=involved
```

Table 28: stage 3 neuroblastoma is defined by four instances of the criterion relationship, each of which identify a term defined by two instances of the criteria relationship. The logical form of the definition is therefore a disjunction of four conjunctions.

The staging and assessment classifications therefore have definitions which are composed of atomic elements linked by logical operators. These atomic elements refer to radiological signs, pathological signs, markers or to the age of the patient. The

majority of radiological signs concern the properties of the abdominal mass. The other signs concern metastatic sites and include the results of nuclear medicine studies. Ultimately one would want a radiologist's workstation to be able to integrate findings from all imaging modalities. In the prototype developed here it was assumed that information regarding the abdominal mass would be obtained from the application of image-processing operators to the CT images, but that other findings would have to be entered by the user. Two properties of the abdominal mass mentioned in the staging and assessment classifications concern whether or not the tumour is resectable and whether or not it has been completely excised. It was not expected that these properties would be defined in terms of image-processing operators. An extension to this work could attempt to provide a definition of, for example, 'resectability' in terms of properties which could be measured using image processing.

One property required for the assessment decision concerns the change in size of the abdominal mass. In order for this to be measured it is necessary to compare two values for the volume of the mass. The knowledge base includes statements about the change in volume of the abdominal mass; the business of identifying the appropriate images and measurements and calculating the difference is handled at the level of the interface with image processing.

9.6 The Representation of the Protocol

The protocol represented in this prototype is that used at Great Ormond Street Hospital in the management of neuroblastoma. This protocol is not written down in a single document but is rather a representation of the usual procedures followed by the clinicians concerned. The basis for the protocol is the set of tests required by the INSS to confirm a diagnosis, assign a stage and assess the response to treatment. In order to obtain a detailed specification of the protocol, I first represented the actions implied by the INSS criteria and the given treatment options. This draft representation consisted of

a set of diagrams using the formalism described in Section 7.4. These were then used as the basis for knowledge elicitation sessions with the collaborating clinicians, a senior oncologist and a senior radiologist at Great Ormond Street Hospital. As a result of these interviews a number of changes were made to reflect the reality of clinical practice and a more natural terminology.

One problematic aspect was the need for a flexible representation of the testing required by the protocol. The protocol is an attempt to represent the procedure followed after a patient is diagnosed as having neuroblastoma. Obviously, a variety of tests will be performed in the course of that diagnosis, that is, before the protocol comes into play. The first phase of the protocol, however, is a set of tests which must be performed, if they haven't already been carried out in establishing the diagnosis. This 'confirmation of diagnosis' phase of the protocol could be represented as a set of sub-plans which each consist of a choice between doing a test or not doing a test and the fact that a test had already been done could be used as an argument against performing it. This seems clumsy, every instruction to carry out an investigation might implicitly include an instruction to check if the test had already been performed, if there had been an intervention since it was performed, and only to carry it out if required. This was the approach taken: the checks were not represented in the protocol but carried out automatically by the task manager when updating the patient record.

As with the mammography prototype, the level of detail included in the protocol was determined by the desire to support the radiologist's role in managing the disease. The details of examinations were not thought to be required and nor was information about the treatment options (dosage, side-effects, contra-indications etc.) included.

9.7 Overview of the Prototype

The prototype was implemented to demonstrate the generic character of the architecture described in Chapter Five and particularly of the extended decision model described in Chapter Four. The design and appearance of the prototype was therefore very similar to that of the prototype described in Chapter Seven.

The user interface provides a set of buttons for accessing the various display facilities and a set of controls for altering the characteristics of the display. The buttons provide access to menus which allow the display of relevant lists of patients known to the system and to browsers of the medical knowledge and the protocol.

When the user selects a patient from one of the menus, the relevant images are displayed at a resolution of 256x256 pixels, together with a summary of the protocol being followed. The user can request the computation of the midline, or use the interactive segmentation tool to identify the tumour volume. The user can also request decision support for diagnosis, staging, assessment or the selection of the next action. In the staging and assessment decisions, requesting decision support has the effect of initiating any image processing required to obtain evidence from the images.

The decision-support display is analogous to that described in Chapter Seven. Once the displayed information has been used to make a decision, and that decision recorded in the system by the user, the task manager updates the representation of the patient record.

Unlike the mammography system, the image processing required here by the decision support involves an element of user interaction: either the provision of design support has to pause to allow the user to place seed points and execute the region

growing procedure, or the user has to have performed this operation before requesting support. An appropriate use of the system might be to support an oncologist making decisions on the basis of a radiologist's segmentation, in which case the latter mode of operation would be realistic.

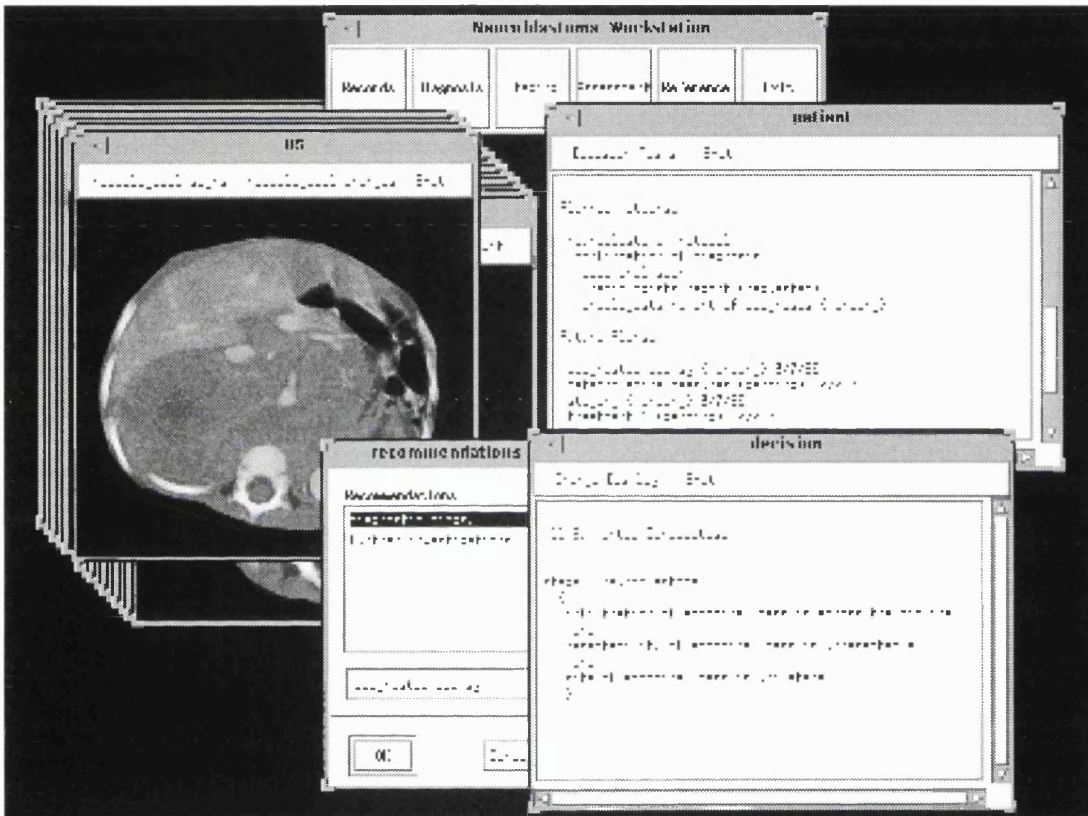


Figure 41: a screendump of the neuroblastoma application. The stack of images making up the CT data set are displayed on the left. On the right a brief summary of the patient record is shown. The decision support window (bottom right) contains the arguments confirming that the tumour is a stage 3 neuroblastoma.

9.8 Conclusion

The prototype described in this chapter was not as fully researched as the mammography system described in Chapter Seven and has not been exposed to the

same kind of evaluation. It was implemented to demonstrate that the decision model described in Chapter Four and the architecture described in Chapter Five could form the basis of a range of decision-support applications and the successful implementation of the prototype serves to support the claim that the architecture is generic.

There are, however, some differences between the two applications. The decision supported in the mammography application – the differential diagnosis of calcifications – involves a number of uncertainties: the knowledge employed in the decision is not hard and fast; the evidence used is ambiguous and prone to error. The decisions supported in this application are different: the available candidates are clearly defined, non-overlapping categories and the evidence obtained from the images is all-or-nothing. Although the information provided by the decision-support tool is still presented in the form of arguments for and against candidates, it is anticipated that there will always be exactly one candidate for which an entirely conclusive argument has been found.

The system has a number of obvious weaknesses. The first is that it was not possible within the constraints of this project to identify a segmentation algorithm capable of providing an accurate assessment of tumour volume. The method used, that of ‘intensity connectedness’, seems promising, however, and could form the basis of an adequate method. The second is that although the decision is an important one it is not complex enough by itself to warrant a radiologist learning to use a new software tool. Therefore, if the goal of ensuring that the guidelines are applied consistently is to be met through the use of a computer system, that system must be provided as a component of one which the radiologist can also use for other tasks.

CHAPTER TEN

DISCUSSION

10.1 Introduction

In the opening chapter I listed the four aims of this work. The first two – to see how the range of systems used to assist in the interpretation images of compare when viewed as decision support systems and to consider how it might be possible to combine different kinds of information to provide decision support – were addressed by the literature reviews in Chapters Two and Three. The third, to provide a general framework for building decision support systems which combine symbolic knowledge and signal data, was addressed in Chapters Four and Five. The final aim, to build prototype systems demonstrating the practicality of the ideas, was addressed in Chapters Six to Nine. The purpose of this chapter is to consider the extent to which the aims set out at the start have been met, highlight issues brought out by the work and suggest directions for future research.

10.2 The Potential for Decision Support in Radiology

Computers are increasingly viewed as tools for obtaining information. There are now a variety of computer applications which doctors can consult, from CDROM

databases of the medical literature to electronic patient records. In radiology the penetration of information technology is perhaps deeper than in other areas of medicine, driven in part by the demands of modern digital imaging procedures. There is therefore both an understanding of the potential of computerised information sources and the technology to enable the provision of decision support. The promise of this situation should not be exaggerated, however. Most of the radiologists I encountered were not confident users of computer technology.

In Chapter Two, I described a number of different kinds of system which have been developed to assist in radiological decision making. These included systems based on databases of radiological images, on symbolic knowledge bases, on stored probabilities and on image-processing operations. There are clear differences in the way that the research in each of these areas relates to this thesis. A great deal of work has been carried out in image processing, but although much of this has been applied to radiology relatively little has resulted in systems which are used by radiologists to help in making decisions about images. Knowledge-based decision aids and Bayesian systems have been successfully developed for a number of clinical applications, but relatively few attempts have been made to apply them to radiology. Research in image databases has largely been concerned with the development of systems which allow images to be stored and made available through networks. There is also current work looking at how clinicians might use such systems as information sources.

Focusing on the needs of the radiologist as a decision-maker reveals the shortcomings of these different systems. Decision aids based on symbolic reasoning or Bayesian calculations can provide accurate information about diagnoses but only on the basis of a detailed description of the image contents, a description which must be provided by the radiologist – a rather irksome task. It is also worth noting that radiology, by and large, is not rich in the kind of declarative knowledge which can be represented symbolically. The Keravnou et al. [Keravnou1993] system, described in Section 2.3.3, deals with a domain in which there are thousands of syndromes to be

considered when making a diagnosis, so it seems plausible that mastering the knowledge required to make a diagnosis is a difficulty for radiologists. Such domains are relatively rare in radiology. It may also be relevant here to recall that the findings of the evaluation in Section 8.2 suggest that much of the knowledge that is written down about radiological decisions is not used by radiologists.

Image-processing algorithms have been developed which perform adequately as detectors or classifiers of certain kinds of feature or abnormality on certain kinds of image but which are also not accepted by radiologists as everyday tools. This may be because they can be affected by apparently minor changes in the image acquisition process and may seem unduly fragile to radiologists, or it may be that the information they provide appears unrelated to the decisions the radiologists are trying to make. Image databases provide large stores of example images which contain information useful to radiologists, but the problem of how to arrange the retrieval of an appropriate image to help in a particular case is as yet unsolved.

Systems based on symbolic representations, on the one hand, are problematic because it is difficult for radiologists to put information about images into a symbolic form. Systems which deal in signal information, on the other hand, have difficulties because it is hard to extract the information from the medium used for representation. The approach taken in this thesis has been to develop a decision aid which combines the two forms of information: image processing to provide the descriptions which knowledge-based systems require and symbolic representations to relate image data to the decisions radiologists take.

Four types of information source were mentioned in the opening chapters of the thesis but I concentrated on only two: knowledge-based systems and image processing. Future development of the model of decision making as the basis for decision aids in image interpretation could incorporate the other two kinds of information source:

image databases and numerical decision aids. One of the key findings in the evaluation of the knowledge base was that radiologists disagree strongly about the terms which could be used to describe an abnormality and about which properties of the abnormality are characteristic of malignancy. It might therefore be interesting to see if the decision model could be used to present arguments for and against a malignant interpretation using example images from a database. The use of quantitative data (either in the form of subjective estimates of probability or as objective probabilities obtained through experimental studies) would be another possible extension of the decision model. This is discussed in more detail in Section 10.4.

If computer aids are to be developed to improve radiological decision making - and given the increasing computerisation of radiology and the need to make the best possible use of imaging investigations and radiological expertise this seems a highly desirable goal - they must allow the combination of different kinds of technology. A decision-support system based on such an approach should meet the requirements identified by Shortliffe: clinicians will want to be confident that they understand the basis for the advice being offered and will be more likely to use a system which is integrated with those storing the appropriate clinical data, the patient record and so on [Shortliffe1991]. The challenge is therefore to set out an approach within which different kinds of information, stored in different systems, can be used to provide information in a way which is transparent to the user.

10.3 The Signal-Symbol Question

One aim of this research has been to consider how information obtained from the processing of images can be combined with symbolic knowledge to provide decision support. The problem of combining symbolic knowledge and signal data is a general problem of great significance for artificial intelligence and computer vision. Signal data consist of large volumes of information stored as a consequence of the

signal-formation process: to be used the information must first be extracted. If signal data is to be mapped to symbolic knowledge there must be a mechanism for extracting the information from the data and there must be an association between the result of this process and elements in the symbolic representation.

There are three related questions to be considered: what kind of representation is appropriate for the results of image processing; what information is passed across the signal-symbol boundary and how is the process controlled?

The problem of constructing a general framework to represent the results of image-processing operations is beyond the scope of this thesis, but a number of approaches which have been used in the past were compared in the review in Chapter Three. These include the use of image data (in the form of templates and reference images which can be matched with the images being interpreted), and of region-based, boundary-based and axial descriptions which can be derived from image data. At the moment it does not seem possible to say that any one of these approaches is likely to prove pre-eminent. These systems are vulnerable in different ways. Just as the different approaches to image processing perform with varying success on different classes of image, the variety of characteristic errors which result differ in the consequences they have for different representational schemes.

It was not necessary for my purposes here to settle on a single approach to image processing, however. It is possible to envisage a symbolic representation of an image in which each symbol is associated with the appropriate signal-derived description. There is no need to adopt a single approach to image processing or to represent its results in a consistent form, so long as the symbolic representation is able to construct queries which it can map to the appropriate image-processing operations.

The second question concerns the kind of information which is obtained from image processing. The approach adopted here is a comparatively simple one, in that image processing is used only to establish whether statements generated by the symbolic reasoning are true or false. Noise, error and ambiguity mean that the results of image processing are inevitably uncertain. No indication of the uncertainty associated with a particular result is used in the symbolic reasoning.

The third question concerns the control of the interaction between symbolic reasoning and image processing. The image data which form the basis of perception are inherently ambiguous and it is therefore widely assumed that prior knowledge must play some role in image interpretation. The phrase 'knowledge-based image interpretation' is frequently used to describe computer systems which employ some form of prior knowledge. In some systems this knowledge is represented as signal data, which play an analogous role to the prior probabilities in a system of Bayesian updating. In others the knowledge is represented symbolically and used to control the application of image-processing operations, that is, to implement some form of 'intelligent' strategy for image interpretation. The approach taken in this thesis, as discussed in the next section, is to use the symbolic knowledge to generate the goals of the image processing and to determine which image-processing operators are chosen. A great deal more could however be done to use knowledge about anatomy within certain image-processing operations. Section 10.8.1 includes some discusses this in the context of the neuroblastoma application.

10.4 The Extended Decision Procedure

The approach taken in this thesis to the development of a decision aid employing both signal data and symbolic knowledge has been to develop a framework within which information derived from signal data can be used in symbolic reasoning. The framework is based on a model of decision making extended to include a model of

three generic tasks involved in image interpretation. The symbolic reasoning is used to propose candidate solutions and to construct arguments, or lines of reasoning, for and against candidates. The rules used to propose candidates and arguments draw on information represented in a symbolic knowledge base. The rules used to garner the evidence required by the arguments draw on symbolically represented knowledge about the findings which may be detected in the image data and control the application of image-processing operations to obtain the required information from the image.

This model has been used in two prototypes, one in the domain of mammography to assist in the differential diagnosis of calcifications and one in the domain of abdominal CT to assist in decisions made in the management of childhood abdominal tumours. The approach seems to be a useful one. The symbolic decision procedure allows the construction and presentation of lines of reasoning which assist in deciding between different candidate solutions. The extensions to the decision procedure allow the system to use image processing in interpretation tasks involving the detection of elements in an image and in the classification and measurement of detected image features. The model has been applied successfully to different kinds of decision in two quite different clinical areas.

Two aspects of the extended decision procedure seem particularly valuable, both of which relate to its role in controlling image-processing operations. The first concerns the notion of a 'model' for the classification and measurement tasks. The model lists the structures which must be located before such a task can be attempted. Neither the model nor the resulting interpretation needs to be complete, they need only represent what is required for a particular task. The interesting point here is that the model is part of the description of an interpretation task: it is a declarative statement of part of an interpretation strategy.

The second valuable aspect of the extended decision procedure is that the declarative representation of the depictions of abnormalities on particular classes of images and of the available detectors provides the basis for a knowledge-based image interpretation system. Since the choice of image-processing operators is determined by matching the description of the operators with a description of the way findings are depicted in images, the system represents knowledge about the suitability of different image-processing operators for different tasks and provide the same kind of functionality as systems which provide intelligent interfaces to image-processing libraries.

Two possible extensions to the extended decision procedure were suggested in the conclusion to Chapter Four and are worth restating here. The first is that although I used this model to provide information to radiologists making decisions about the interpretation of images, it would be equally possible to apply the model in systems which automatically interpreted images in order to provide information to clinicians making decisions about patients. Indeed, in Chapter Four an example was given of a simple, relatively low-risk decision which could be made entirely automatically. If a mammography screening programme were designed so that easy-to-read images were to be interpreted by trained radiographers, it would be necessary to decide which mammograms were easy to read. This kind of decision could be made automatically by a computer on the basis of a simple measurement of the amount of dense tissue in the image, since this can be reliably measured and seems to be the principal cause of difficulty in interpretation.

The other extension concerns the nature of the information provided to the user. In the implemented prototypes information is provided in the form of arguments for and against a particular candidate. No information is given as to the prior probability of the candidate, the force of the argument or the uncertainty associated with the evidence on which it is based. Uncertainty arises at many points in the imaging process: through errors introduced into the image data, the fallibility of any image processing employed, the probabilistic nature of much radiological knowledge and the imprecise terms used

to express that knowledge. Any attempt to present the user with information about the uncertainties associated with an argument must therefore be based on attempts to model these different kinds of uncertainty or incompleteness. There are a number of formalisms available for handling different kinds of uncertainty (including Bayesian probabilities, certainty factors, belief functions, possibility theory, non-monotonic logic and qualitative approaches, like that used in the argumentation-based model of decision making adopted here).

Sucar et al. [Sucar1991] present an approach to modelling uncertainty in knowledge-based vision, based on a formalism known as a Bayesian network. A Bayesian probability is a number which expresses either the subjective (estimated) or objective (measured) likelihood of a proposition or a relationship. It can be used to represent the influence a piece of information has on our conviction that another piece of information is true. However the mathematical basis for calculating this influence allows pieces of information to be combined in updating a belief only if they are 'independent', that is to say, only if knowing one to be true does not affect our belief in the other. Bayesian networks are used to represent the 'conditional independence' assumptions between propositions.

Sucar et al.'s system employs a hierarchical probabilistic network in which the nodes at the lower levels represent image features and the higher level nodes represent hypotheses about objects. The development of the network has two phases. First the topology of the network has to be determined, which requires an understanding of the process of object recognition, and the way it depends on the recognition of features; subjective probabilities are obtained from heuristics about relevance. The initial topology is revised in a learning process based on statistical tests for correlation and redundancy. In the second phase, once a topology has been obtained which accurately represents the structure of the problem, objective probabilities are determined by testing the performance of the feature recognition software.

It would be interesting to investigate whether this approach to modelling the uncertainties involved in obtaining information from images (due to noise and error in the image formation process and the unreliability of image processing) could be combined with an approach based on the representation of arguments. It might be that Bayesian probabilities provide a suitable approach for modelling all the different kinds of uncertainties involved, or it might be that an eclectic approach has some benefits and that argumentation – since it allows the terms indicating the strength of an argument to be drawn from different dictionaries – provides a suitable framework for it. An investigation which explored not just the computational tractability of the different approaches but also their suitability as aids to decision making would be potentially very useful. It should be noted that the question concerns not just which is the most appropriate or accurate way of capturing information about uncertainty, but also what is the most helpful way to present information about uncertainty to a clinical user.

10.5 The Generic Architecture

One of the key requirements for a usable decision-support system is that it must be integrated into the computing tools which provide the everyday working environment for the user [Greenes1991, Shortliffe1991]. A great deal of the discussion about computer support for medical imaging has centred on the topic of the integrated clinical workstation. The concept is discussed by Meinzer, who argues that what clinicians require of such a workstation is not access to more advanced image-processing tools but rather to the full range of relevant clinical data, including protocols [Meinzer1996]. In Chapter Five I set out an architecture for a class of decision-support systems based on the extended decision procedure described in Chapter Four. The aim was to provide a general design for a system which combined access to the different kinds of information system described above with a representation of an imaging protocol and a patient record system. It is envisaged that such a system would exist

within a network of inter-operable systems and so functions for the administration of patient records, the archiving of images, the scheduling of investigations and so on would be provided by an electronic patient record system, a PACS system and a hospital information system.

There are numerous problems associated with the integration of these different systems, some of which are beyond the scope of this thesis. The most obvious one concerns the lack of a standard representation for the information which must be shared between the different applications. For a clinical workstation of the kind described here to become a practical reality the following are required:

- a formal language for the representation of clinical protocols
- a model for the representation of the knowledge required to provide decision support
- the definition of an interface allowing the protocol management and decision-support functions to issue queries to other systems – such as an electronic patient record system – in which information may be stored about a patient

The work described in this thesis has contributed to and builds on early work carried out at the Imperial Cancer Research Fund on the representation of clinical protocols. Developments of that work have now led to the development of a formal language for the representation of protocols which is now being piloted and on which future developments of the work described here could build [Fox1997].

The knowledge bases required for the applications described are relatively small and the symbolic decision procedure provides a clear specification of the knowledge required to support decision making. This means that designing and populating a knowledge base which is clear, complete and internally consistent is not a

major difficulty, or at least not for the domains tackled here. In other areas, where a larger and more heterogeneous body of knowledge must be represented, the authorship and evaluation of the knowledge base would have to be given more thought. Much work has been done in these areas [Coiera1995] and the problems seem tractable for limited domains.

Perhaps the greatest challenge facing such systems is to ensure that a general query language can be defined allowing information to be obtained from other information sources. This area is of considerable interest at present, now that the potential for co-operating or inter-operable systems has become apparent. Future work in this area would seek to capitalise on research into 'query servers' which provide a model for the type and form of requests which can be passed across the interface between co-operating systems as well as a representation of the clinical knowledge required to map from one representation to another [Kalra1994].

The design and implementation of a robust clinical workstation completely integrated with existing clinical systems would be a major undertaking, and one which is incompatible with the demands of a PhD thesis. However the work described here provides a suitable framework for the development of such a system. Three principles which should guide the design of such a system were identified.

- The design must support functions such as decision-support, image reporting, protocol management and information retrieval. In order to do this it must provide a set of modules which can be connected to other clinical information systems.
- For each knowledge source incorporated within the architecture, there must be a definition both of how information is to be displayed to the user and of the form of queries supported by the knowledge source.
- The design of user interface must reflect the tasks radiologists perform.

A generic architecture based on these three principles was used successfully in the implementation of two prototype systems.

10.6 The Choice of Clinical Problem

In Chapter Two I listed a number of requirements for an effective decision support system. It must provide an adequate information source and an effective means for retrieving information from that source; the information retrieved must be relevant and helpful. Developing an effective decision aid is only possible in a project where the available resources allow these requirements to be met. There are therefore strong constraints on the kind of problem that can be tackled. First, there must be clinical support for the project: it must, in other words, tackle a decision where there is a recognised need for this kind of decision aid. This means that the problem must be sufficiently difficult for the clinician to require support and common or serious enough to warrant attention. Secondly, information allowing the development of an adequate knowledge base must be available and the image-processing problems must be tractable.

The two clinical examples tackled in this thesis are the differential diagnosis of calcifications on mammograms and the use of CT in the management of childhood abdominal tumours. In the UK there are about 32,500 new cases of breast cancer every year and about 14,500 deaths, making the disease the most common cancer in women. In the UK screening programme in 1994/1995, 1,207,316 women were screened, 63,925 of whom were recalled for further mammography, 6,334 were given biopsies, 2,000 of which were found to be benign and 6,500 cancers were detected [Patnick1996]. It isn't easy to assess how many cancers are missed in screening. One study which has attempted to quantify this is that of Peeters et al. [Peeters1989] who

reviewed the films for screened patients who presented with a palpable lesion between screening visits, an average of 1.63 per 1,000 women screened. In 153 interval cancers 40 were classified as 'missed', 24 as radiographically occult and 89 as being 'true' interval cancers. Of the missed cancers, in two cases the image had been taken incorrectly, in 12 cases the tumour was outside the normal field of view, in 14 cases observers were deemed to have been in error and in ten cases the observed signs would not normally have led to a diagnosis of malignancy.

A number of studies have shown that having mammograms read by two readers or even by the same reader on two different occasions can significantly improve the detection rate. Unfortunately, however, there is a shortage of radiologists willing to work in the breast screening programme and it is therefore worth considering whether non-radiologists can be trained to act as second readers. Pauli et al. [Pauli1996] investigated exactly this possibility and showed that radiographers once they gain sufficient experience, are able to perform with an accuracy which approaches that of experienced radiologists. In the light of this, it seemed appropriate to explore the idea that image processing could be used in a decision-support tool designed for use by radiographers being trained as film readers.

The available literature on the differential diagnosis of calcifications seemed to suggest that a knowledge-based decision tool, incorporating image processing, would be appropriate for this application. It is important to be clear about what is meant by 'differential diagnosis of calcifications' – the aim of a radiologist (or a trained radiographer) is not really to ascertain whether the calcifications on a mammogram have a benign or a malignant interpretation, but rather to assess whether the risk of malignancy warrants further investigation. The difficulty of the decision (and therefore the possible scope for a decision-support tool) is a function of the degree of risk which is thought to warrant investigation, and that, in turn, is determined by, in part, by the nature of the investigation envisaged. Some authors argue that radiologists should identify a category of 'probably benign' calcifications which should only be monitored

[Sickles1995]. If, however, the policy is that any calcification which is not obviously benign should be biopsied, the need for decision support is much reduced.

The other application tackled in this thesis was the staging and assessment of neuroblastoma. Here the motivation was slightly different. The decisions involved are not complicated: the number of categories is small and the definitions are reasonably succinct. However, the internationally accepted classification is only one of number of schemes which have been proposed and has itself been the subject of revisions [Brodeur1993]. Some of its definitions involve terms which are open to misinterpretation, for example, the definition of 'infiltration across the midline'. It would therefore seem potentially useful to encapsulate these definitions within a decision-support tool capable of providing decision support based on an interpretation of the image. Another problem in staging and assessment is the measure used of tumour volume. Conventionally volume is estimated by measuring the cross-sectional area and length of the tumour. This, effectively means making assumptions about the tumour's shape which are often extremely inaccurate. Eggli et al. [Eggli1995] have shown that in 43 out of 45 cases the tumour volume, measured using the traditional formula, differed by over 50% from the true 3-D volume calculations. A decision aid for staging which incorporated a tool allowing the accurate measurement of 3-D volume would therefore enable a greatly improved assessment. Research into the treatment of neuroblastoma is continuing and it is possible that the advent of more accurate and consistent staging and assessment will allow the development of more refined criteria for treatment selection.

10.7 The Mammography System

This section reviews the work carried out in the development and evaluation of the mammography prototype. The first two sub-sections describe the implementation of the knowledge base and the image-processing operators. The third section discusses the evaluation of the prototype.

10.7.1 The construction of the knowledge base

The knowledge base was constructed in two phases. A database of facts was abstracted from textbooks and review articles. These facts were statements of a relationship between the processes which cause calcification and the properties of calcifications and listed the characteristic properties of calcifications resulting from a particular process. Seven radiologists were then consulted and their opinions used to revise the knowledge base. The resulting knowledge base was largely a selected subset of the original facts, those selected as being the ones with which the radiologists agreed. Once these two phases had established a useful subset of diagnoses and properties, it would ideally have been appropriate to carry out an exhaustive check for omissions; by asking the radiologists to assess all possible associations between properties and diagnoses. It seems likely that the existing knowledge base contains some omissions since properties which are useful in the diagnosis of some potential causes of calcifications but are also true of others, will be associated in the knowledge base with the former and not the latter. Two findings of my research, however, suggest that a quite different approach should be taken to the problem of building a knowledge base.

The first is to do with the fact that the knowledge base was designed with reference to a decision which is not the one that film readers actually make. When a radiologist or a suitably trained radiographer looks at a film, they have to assess the risk of malignancy in the context of the available management options. Since there is no consensus which management options are most suitable, I decided to structure the knowledge base around the knowledge which must underlie an assessment of the risk of malignancy: that is, the properties associated with the different possible causes of calcifications. So, in effect, the system supports a differential diagnosis decision although users are making a decision on a course of action. It became apparent that the radiologists involved in the construction of the knowledge base found this confusing.

Since the knowledge base had to take into account all the possible causes of calcifications, radiologists felt that they were being asked to comment on much more precise diagnoses than can actually be made on the basis of radiological appearance.

The most dramatic finding in the evaluation of the knowledge base was the extremely low level of agreement between radiologists both in their assessments of which statements about calcifications were actually true and in their ratings of the properties of calcifications. Other authors have found similar results. Simpson et al. [Simpson1996] describe a study which identified a set of 17 terms used to characterise microcalcifications. The hope was that this set could be used as the basis for a standardised reporting scheme that would be of value for clinical audit. Eight radiologists viewed 80 sets of films containing microcalcifications and for each selected the terms from the set that they felt were appropriate. The study found low levels of agreement between the radiologists, with values of kappa between 0.43 and 0.05. This finding is consistent with the low levels of agreement reported in Chapter Eight. Simpson et al. conclude:

Skilled radiologists learn to recognise the appearance of cancer on a mammogram, analysis of individual features may form a part of this but even this is usually based on previous experience of seeing similar features, not textbook description.

It seems that radiologists can agree about whether or not an abnormality is indicative of cancer, but not about the physical characteristics on which it might be thought they are basing their judgement.

The system was designed to encapsulate the knowledge contained in the review articles and, as far as possible, to provide image processing which allowed that knowledge to be used in reasoning without the user having to enter a detailed description of the calcifications. A different design philosophy can be envisaged, one which would start by establishing which image-processing measures are most effective

in distinguishing benign from malignant calcifications and would then attempt to build a knowledge base which would accommodate them. In effect the knowledge base would no longer represent the knowledge that radiologists say they use to distinguish benign from malignant calcifications but would describe in radiological terms those measures which a computer can demonstrably use to make the distinction. However, establishing which image-processing measures can reliably make this distinction is not entirely straightforward. Section 6.4 shows that there is very little consensus on which measures are appropriate. Perhaps the best basis for constructing a knowledge base would be to take a single scheme for the classification of calcifications, such as that employed by Monshees (and described in Figure 27) to distinguish between definitely benign, equivocal and suspicious calcifications. Simpson et al. suggest that it may be possible to get radiologists to agree on the definition of the terms used in such a scheme if they are illustrated in drawings and example images, and it would be interesting to explore this idea by extending the decision model to allow information from an image database to be used in decision support.

The approach suggested in the above paragraph is profoundly different from the one taken in this thesis, and from that taken in the development of most decision support systems. In proposing it I am arguing, in effect, that a decision support system for radiology may not be able to depend on terms which radiologists can be assumed to understand but may have to impose on radiologists a set of rigid definitions which they will have to learn. This decision-support system would not be a formalisation of existing medical practice (although it might build on a set of definitions used in certain centres) but would require users to adopt a more rigid approach to their work.

As was discussed in Section 10.4, in the existing prototype there is no way of conveying the prior likelihood of a hypothesis, the force of an argument or the strength of evidence supporting it. This meant that information about some of the more obscure conditions had to be removed from the knowledge base and that it was not possible to represent the fact that a particular shape or configuration of microcalcifications was

possible but atypical. It is debatable whether it is best to provide information about all the possible presentations of all the possible conditions or whether it is more effective just to provide information which is likely to be relevant. It might be thought that users would require information about unusual presentations from a system of this kind, but the feedback I received from radiologists suggested that statements about presentations which were unusual or obscure were not thought to be helpful.

10.7.2 Image processing

Perhaps the most obvious weakness of the system is the limitation of the image processing. The primary focus of the work was not the detection and segmentation of calcifications, so it was important not to devote too much attention to this difficult problem. An adequate solution had to be found however and I decided to implement a simple technique, based on an approach which had been successfully applied in the literature. It is not clear that this solution was adequate, especially given the vulnerability of the shape description measures to inaccurate assessments of the calcification boundary. Work in the detection and classification of calcifications is continuing, however and progress is made not just in devising and applying new techniques but also in the quality of the available data.

The image-processing measures used in the system were selected on the grounds that they had been successfully applied in previous systems for classifying calcifications. Taking measures to feed into a classification procedure is, however, a rather different task from taking measures to support explicit reasoning about properties. In the former case the classification rule can be flexible in the use it makes of a measure. I found the compactness of the convex hull of a cluster to be an extremely strong predictor of malignancy when applied to the data set as a whole but to have no predictive value when applied to clusters of five or more calcifications.

Many of the measures used were too simple. For example, the only measure of shape was a measure of elongation. This was used primarily because it was a property used to distinguish benign from malignant calcifications on which radiologists might reasonably be expected to agree and for which a number of plausible measures exist in the literature. But it did not distinguish well between the 'grain of salt' type calcifications identified by Le Gal [Le Gal1973] and the more rounded ones. Polymorphism and cluster shape are two signs to which radiologists ascribe a great deal of significance, but neither, as measured in this system proved to be of great value. It is perhaps the case that, to be accepted by clinical users, a decision aid for the differential diagnosis of calcifications will require more sophisticated measures of shape than it was possible to develop in this thesis. Future work could take as a starting-point the scheme discussed above for the three-way classification of calcifications and attempt to design measures of shape specifically for the descriptions it employs.

10.7.3 The evaluation of the mammography system

The aim of the evaluation was to identify a setting in which the prototype could be used and to demonstrate that its use brought about an improvement in decision making. The prototype was not given a clinical trial – an evaluation of that kind is not really possible given the current state of the art in digital mammography – and the results of the evaluation are therefore not definitive. Given the shortcomings of the system, it was decided to test the system on a group of users who were less skilled than experienced screening mammographers. The radiographers being trained at the Jarvis Screening Centre seemed a suitable group.

The evaluation demonstrated a significant difference between radiographers using the system and radiographers not using the system, even only a very limited amount of data could be collected – only a very small number of radiographers are currently trained to read films and there are a limited number of occasions on which they can be brought together. Pauli et al. [Pauli1996] studied radiographers for a period

of 12 months after their initial training and found their performance was comparable to that achieved by radiologists. They also found that this level of performance was maintained over time, but that, contrary to what one might have expected, the radiographers did not improve with practice. This might be because they were only reading the minimum number of films considered necessary to maintain film-reading skills. It therefore seems likely that a system of this kind could be of value for such users.

The implemented prototype has a number of limitations which would have to be overcome. First, and most obviously, the display of the images would have to be of a sufficiently high spatial and contrast resolution to compare with conventional film. Secondly, the image processing and the knowledge base would have to be improved. Thirdly, it would have to be provided with a convenient and robust interface and backed up by the processing power required to support the rapid reading of large volumes of images. Although these problems could not be solved in this project, they are not insoluble in principle.

10.8 The Neuroblastoma System

The neuroblastoma prototype was implemented to demonstrate that the approach to decision support set out in the description of the extended decision procedure and the generic architecture could be used in a range of applications. The prototype was the subject of a much less extensive programme of research and implementation than the mammography system. Nevertheless, there are some important points to be made. The decisions supported by this system were much simpler than those in the mammography prototype: they are selections between clearly defined alternatives; the arguments which are to be constructed are always either confirming or excluding; they are based on definitions set out in a document and there is no real uncertainty involved. The arguments are complex, however, in that they involve conjunctions and disjunctions, necessitating an additional phase of 'argument

construction' to be built into the model. The two principal weaknesses of the prototype concern first the inaccuracy of the segmentation operation required to measure tumour volume and secondly the question of how, and by whom, the system would be used.

10.8.1 The measurement of tumour volume

Two segmentation algorithms were compared. Both were basically region-growing algorithms, one in which each voxel (or pixel) is added to the region on the basis of a comparison of its brightness value with that of a seed point, the other in which the 'intensity connectedness' of the two points is measured.

Looking at CT images of neuroblastoma it quickly becomes apparent that the tumour is not distinguished from the surrounding tissue by its distinctive grey-level, nor is there a difference in texture which is obvious to the eye. Watching a radiologist attempt to trace the boundary suggests that the information in the image does not actually allow the boundary to be identified with certainty. On occasions the radiologist is forced to make a guess as to whether a portion of the image is or is not to be included within the boundary. It is therefore not surprising that these segmentation algorithms performed poorly on this particular task. The simpler of the two seems unlikely to provide a promising basis for future work, if only because of the widely differing grey-levels in a neuroblastoma.

One obvious difference between the segmentations produced by the algorithm and those drawn by the radiologist is that the radiologist is guided by an idea of the kind of shape the tumour is likely to have, or perhaps more accurately, of the kind of shape the tumour is unlikely to have. Both the segmentation algorithms might be considerably improved by building into the process of region-growing a cost function which inhibited the development of shapes with long complicated boundaries. The work of Kobashi et al. [Kobashi1995] in which segmentations are revised on the basis

of knowledge about what kinds of shape are unlikely or impossible might be relevant here. There would, however, be problems with such an approach: neuroblastoma tumours are lobulated and can calcify, both of these attributes can introduce strong edges within the tumour boundary. Building a measure of shape into the algorithm could, therefore, cause the region-growing process to stop at what would be, in effect, a local maximum.

10.8.2 The need for decision support for staging and assessment decisions.

One of the difficulties here concerns the collaborative nature of the decision-making process. In order for staging and assessment decisions to be taken, radiologists must first provide an interpretation of the appropriate images, and the prototype developed here is intended to assist in this. The final decision, however, is taken by the clinician responsible for the management of the disease. A workstation which could be used in clinical practice would therefore have to support the collaboration between the two. In effect, it would be necessary to design two inter-operable systems, both of which were also integrated with other systems used by the two clinicians. Computer support for teams of collaborating healthcare workers is a goal of current research, but there are many problems to be solved before a system of this kind could be installed where users might realistically be likely to take advantage of it.

10.9 Conclusions

In the opening chapter I listed the four aims of the thesis:

- to consider the range of systems which could potentially improve medical decision making
- to consider how different kinds of system could be integrated into a single decision aid

- to propose a generic model for decision support tools which allows users to access relevant information
- to develop prototype systems demonstrating the practicality of these ideas

There are now a variety of systems which can be used to make available different kinds of information which can assist in decision making. Each has advantages and disadvantages, and so it seemed desirable to consider how a system could be developed which could combine information from different sources. Specifically, I wanted to use image processing to provide the descriptions which knowledge-based systems require and symbolic representations to relate image data to the decisions radiologists take.

The most important element of this thesis is the logical model which allows image-processing operators to be used to obtain evidence to support reasoning about decisions made in the interpretation of medical images. The model is based on a symbolic decision procedure which constructs lines of reasoning for and against possible solutions: this is an effective mechanism for generating the display of information to be used in making decisions. The extension of this model to include a representation of three image interpretation tasks allows these displays to include arguments based on the detection, classification and measurement of image features.

Three principles which should guide the design of such a system were identified.

- The design must support functions such as decision support image reporting, protocol management and information retrieval. In order to do this it must provide a set of modules which can be connected to other clinical information systems.
- For each knowledge source incorporated within the architecture, there must be a definition both of how information is to be displayed to the user and of

the form of queries which are supported by the knowledge source.

- The design of the user interface must reflect the tasks radiologists perform.

A generic architecture based on these three principles was elaborated.

The practicality of these ideas was demonstrated by the implementation of two prototype systems which addressed decision problems in two quite different clinical areas: the differential diagnosis of microcalcifications on mammograms and the use of CT in the assessment of neuroblastoma. An evaluation of the mammography system demonstrated that it significantly improved the decision making performance of trained film-readers interpreting mammograms.

10.10 Future Work

The work described here could be taken forward in a number of ways. Two obvious paths for further research concern the two kinds of information source which have received the least attention in this thesis: image databases and numerical decision aids. It would be interesting to see if example images from a database could be used to provide supporting arguments in certain kinds of decisions. It might also be fruitful to explore the use of quantitative measures of uncertainty in the symbolic decision procedure.

These two paths are both associated with developments of the extended decision procedure. From a practical point of view, the effectiveness of different ways of presenting information to the user seems to warrant early investigation. From a technical point of view, the fact that all information about uncertainty is disregarded in the symbolic reasoning seems unsatisfactory.

The overall aim of this thesis has been to provide systems to assist radiologists in making decisions about images. In practice such systems will have to fit within the doctors' working environment. The increasing computerisation of radiology departments suggests that there will be a great deal of useful work to be done in this area. One of the key aims of this research must be the development of decision support systems which can obtain information from other systems, such as electronic healthcare records and hospital information systems. The development of the generic architecture could be carried out as part of collaborative project tackling development of electronic patient records and their integration with imaging systems and image databases.

In addition, further work could be done in the two application areas considered in this thesis. In the discussion of the mammography system it was suggested that it might be more appropriate to organise the knowledge base around a simple three-way classification of microcalcifications and to develop image-processing operations that would provide accurate measures of the terms used in the classification. The two evaluations which measured agreement between radiologists both produced quite striking results. These may be worthy of further investigation as part of project looking at the construction of a symbolic knowledge base but may also have implications for understanding how radiologists acquire their skills and for assessing the consistency of radiological diagnoses. The chief problem in the neuroblastoma system is the segmentation of the tumour. Knowledge about which shapes are atypical could be incorporated into an algorithm to provide a more accurate segmentation.

There are, therefore, a number of promising avenues for future research. The theoretical basis for the work could be clarified and extended and the prototype applications further developed. Advancing these ideas could provide real benefits for patient care and, if the clinical setting allows the decision support system to be integrated with conventional information systems, then there is every possibility of practical success.

APPENDICES

11.1 Image Processing Calculations

This appendix provides details of those of the image processing operators described in Section 6.4. which were employed in, or considered for, the mammography prototype described in Chapter Seven.

11.1.1 Compactness

Compactness is a measure of the extent to which an object has a circular outline, it is defined, for example by Shen et al. [7], as:

$$\frac{4\pi a}{p^2}$$

where a is the area and p the perimeter of the object.

11.1.2 Hu's Invariant Moments

A number of different measures, including eccentricity and elongation, make use of the invariant moments first developed by Hu.

Considering an image as a 2-D continuous function $f(x,y)$, the moment of order $(p+q)$ is defined as:

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x,y) dx dy$$

for $p, q = 1, 2, 3, \dots$

So, for a binary image m_{00} is simply the number of foreground pixels. Assuming the moment is computed for an image fragment depicting a single object, it corresponds to the size of the object in the image. The moments m_{10} and m_{01} represent the sum of the x and y co-ordinate values for all the pixels in the object. These moments can be used to determine the x and y co-ordinates of the ‘centroid’ of an image:

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad \bar{y} = \frac{m_{01}}{m_{00}}$$

It is then possible to calculate the moments about the centroid:

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x,y) dx dy$$

This makes the moments of an object invariant to the position of the object in the image. These ‘central moments’ are often normalised, by dividing by the zeroth moment, raised to the appropriate power. This makes them invariant to the size of the object.

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma}$$

where: $\gamma = \frac{p+q}{2} + 1$ for $p + q = 2, 3, \dots$

These moments can be used to calculate Hu's moments, which are invariant to rotation. Hu described 7 moments, the first two of which are used in Dunn and Brown's

$$\begin{aligned}\phi_1 &= \eta_{20} + \eta_{02} \\ \phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2\end{aligned}$$

calculation of elongation, as described in Section 11.1.4

11.1.3 Eccentricity

The eccentricity of a shape is calculated using the moments about the centroid.

If θ is defined as:

$$\theta = \frac{1}{2} \left(\text{atan} \left(\frac{2 \times \mu_{11}}{\mu_{20} - \mu_{02}} \right) \right)$$

then the eccentricity is given by:

$$\sqrt{\frac{\mu_{02} \times \cos(\theta)^2 + \mu_{20} \times \cos(\theta)^2 - \mu_{11} \times \sin(2\theta)}{\mu_{02} \times \sin(\theta)^2 + \mu_{20} \times \cos(\theta)^2 + \mu_{11} \times \cos(2\theta)}}$$

11.1.4 Elongation

Elongation is a rotation invariant measure of how much a shape must be minimised along its long axis in order to maximise its resemblance to a circle [Dunn1986]. It is calculated from two rotational invariants, which are, in turn calculated from Hu's invariant moments, described in Section 11.1.2.

The two invariants are:

$$\lambda_1 = 2\pi \left(\phi_1 + \sqrt{\phi_2} \right)$$

$$\lambda_2 = 2\pi \left(\phi_1 - \sqrt{\phi_2} \right)$$

and the equation for elongation is:

$$elongation = \log_2 \left(\sqrt{(\lambda_1/\lambda_2)} \right)$$

11.2 Knowledge base used in the mammography system

This appendix describes the knowledge base used in the mammography system to support the differential diagnosis of calcifications. It is presented in three sections. First the hierarchy of possible diagnoses is presented, then the set of different properties and values used to describe calcifications and finally the statements associating the different properties with diagnoses are given.

11.2.1 Hierarchy of terms

The following is a list of all the terms used in the knowledge base of the mammography prototype. The indentation reflects the hierarchical structure of the database, with all the kinds and sub-kinds of a term appearing beneath it: so 'check eligibility' is a kind of action, and actions are kinds of tasks. The term 'topics' is used simply to provide a top-level element for the tree.

- topics
 - tasks
 - actions
 - check eligibility
 - mediolateral oblique mammograms
 - craniocaudal mammograms
 - lateral mammogram
 - insert needle
 - confirm needle location
 - enrol in lumpectomy and chemotherapy
 - enrol in lumpectomy radiotherapy and chemotherapy
 - enrol in high dose radiotherapy
 - enrol in mastectomy and chemotherapy
 - recall patient in a months time
 - recall patient in six months time
 - recall patient in a years time
 - recall patient in three years time
 - reports
 - first screening report
 - second screening report
 - radiologists report
 - report of patient history
 - oncologists report
 - report of assessment meeting
 - surgeons report
 - histologists report

- decisions
 - diagnosis
 - feature classification
 - assessment of difficulty
 - selection of next action
- plans
 - investigations
 - history taking
 - clinical examination
 - a posteriori clinical examination
 - screening mammogram
 - additional mammograms
 - further mammograms
 - pre-operative mammogram
 - ultrasound
 - fine needle aspiration
 - histological analysis
 - procedures
 - excision
 - assessment of tumour
 - assessment meeting
 - treatment plans
 - lumpectomy and chemotherapy
 - lumpectomy radiotherapy and chemotherapy
 - high dose radiotherapy
 - mastectomy and chemotherapy
 - surveillance plans
 - return in a months time
 - screening every six months
 - screening every year
 - screening every three years
 - protocols
 - screening protocol
 - enrolment
 - surveillance
 - refer from screening
 - return to screen
 - biopsy
 - mammography
 - treatment
 - further investigation
 - further investigations
- radiological signs
 - breast mass
 - circumscribed lesion
 - stellate lesion
 - calcifications
 - architectural distortion
 - asymmetry
 - density gradient
 - skin thickening
- diagnoses
 - neoplasm or neoplastic-like condition
 - benign neoplasm

- fibroadenoma
 - giant fibroadenoma
 - papilloma
 - lipoma
 - hemangioma
 - abscess
- malignant neoplasm
 - lobular carcinoma
 - invasive lobular carcinoma
 - lobular carcinoma in situ
 - ductal carcinoma
 - invasive ductal carcinoma
 - ductal carcinoma in situ
 - medullary carcinoma
 - papillary carcinoma
 - tubular carcinoma
 - intracystic carcinoma
 - sarcoma
 - angiosarcoma
 - liposarcoma
- metastatic malignant neoplasm
- lymphoma
- effect of trauma
 - effect of aspiration
 - foreign body
 - fat necrosis
 - haematoma
 - contusion
 - postsurgical scar
 - silicone globules
- duct disease
 - ductal ectasia
 - plasma cell mastitis
- benign breast change
 - fibrocystic disease
 - cyst
 - galactocele
 - oil cysts
 - microcystic disease
 - adenosis
 - blunt duct adenosis
 - sclerosing adenosis
 - papillomatosis
 - radial scar
 - vascular calcification
- cutaneous lesion of breast
- lymph nodes
 - lymph nodes in breast
 - lymph nodes in axila
- artifact
 - circumscribed parenchyma out of puberty
 - normal fat lobule
 - sternal injection of pectoralis muscle
 - summation shadow

- nipple out of profile
- artifacts on skin
- end-on vein
- difficulty classes
 - easy
 - difficult
- difficulty factors
 - predominantly dense tissue
 - predominantly fatty tissue

11.3 Properties of calcifications

The following is a list of all the properties which calcifications and clusters of calcifications may have. The possible values of each property are listed after the property, indented.

- calcification shape
 - crescent
 - cup of tea
 - linear
 - oval
 - round
- cluster shape
 - branching
 - butterfly
 - linear
 - propeller
 - round
 - triangular
- density
 - high density
 - radiolucent centre
 - uniform density
 - variable
- distribution
 - along ducts
 - bilateral
 - clustered
 - densely packed
 - intermittent
 - parallel
 - scattered
- form
 - amorphous
 - bizarre
 - branching
 - casting
 - coarse
 - eggshell

- flaky
- granular
- gravity dependent
- irregular
- monomorphic
- needle-like
- polymorphic
- popcorn
- punctate
- rod-like
- location
 - inside fibroadenoma
 - near skin
 - periphery of mass
- margins
 - sharp
 - smooth
 - smudged
- number
 - numerous
- orientation
 - towards nipple
- size
 - large
 - small
 - tiny
 - intermediate

11.4 Positive signs for the possible diagnoses

The following is a list of those terms in the diseases section of the ‘kinds’ tree in Section 11.2 which are associated with properties of calcifications. The properties of calcifications are listed in italics.

- neoplasm or neoplastic-like condition
 - benign neoplasm
 - fibroadenoma
 - calcification shape is round*
 - density is high density*
 - density is uniform density*
 - distribution is clustered*
 - form is bizarre*
 - form is coarse*
 - form is flaky*
 - form is irregular*
 - form is popcorn*
 - form is punctate*
 - location is inside fibroadenoma*

location is periphery of mass
margins is sharp
size is large
variation in size is variable size
papilloma
form is eggshell
form is polymorphic

malignant neoplasm
ductal carcinoma
number is numerous
size is tiny
calcification shape is linear
cluster shape is branching
cluster shape is butterfly
cluster shape is linear
cluster shape is propeller
cluster shape is triangular
distribution is clustered
distribution is scattered
form is branching
form is casting
form is granular
form is polymorphic
variation in size is variable size
form is bizarre
form is irregular
size is small

effect of trauma

fat necrosis
density is high density
form is eggshell
size is intermediate

postsurgical scar
form is eggshell
form is amorphous
size is large
silicone globules
form is eggshell

duct disease

ductal ectasia
density is uniform density
form is monomorphic
form is needle-like
form is rod-like
orientation is towards nipple
calcification shape is linear
calcification shape is oval
density is high density
density is radiolucent centre
distribution is along ducts
distribution is bilateral
distribution is scattered
form is branching

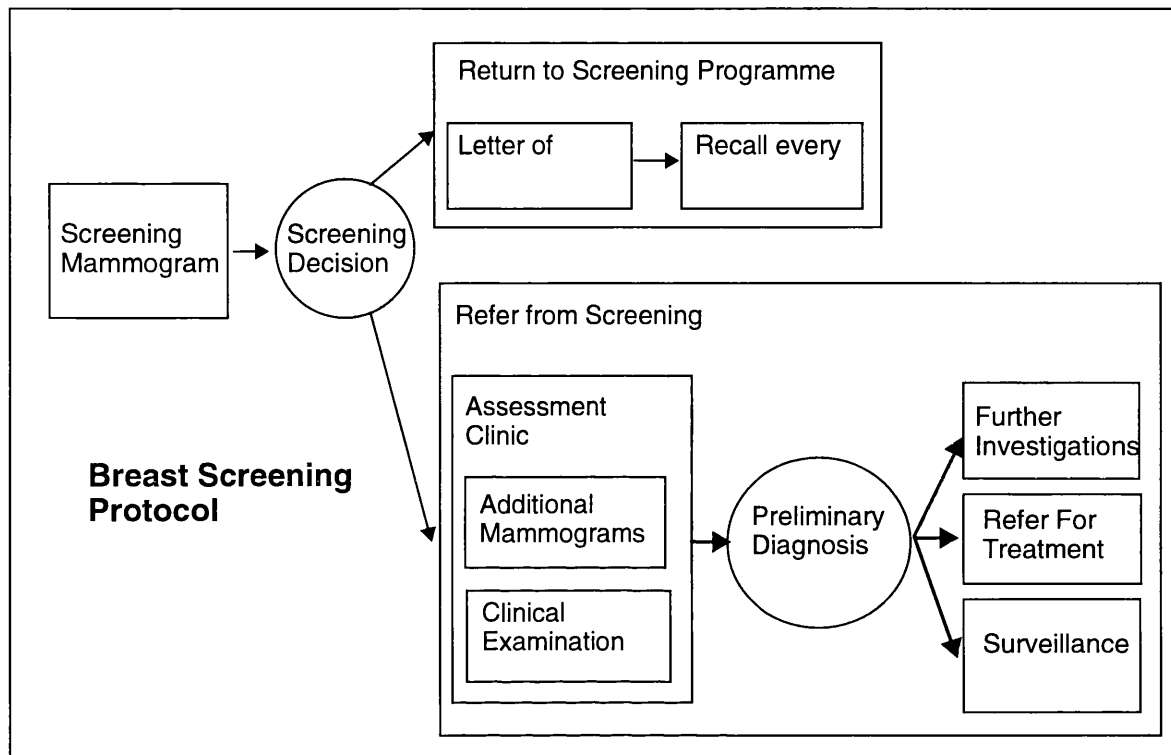
form is casting
size is large
 plasma cell mastitis
calcification shape is crescent
margins is smooth
size is large
calcification shape is linear
calcification shape is oval
calcification shape is round
density is high density
density is radiolucent centre
distribution is along ducts
distribution is bilateral
distribution is scattered
form is branching
size is large
 benign breast change
 fibrocystic disease
 cyst
 form is eggshell
 size is large
 variation in size is variable size
 oil cysts
 calcification shape is round
 density is radiolucent centre
 size is large
 microcystic disease
 calcification shape is cup of tea
 distribution is bilateral
 distribution is clustered
 distribution is scattered
 form is gravity dependent
 margins is smudged
 number is numerous
 variation in size is variable size
 adenosis
 distribution is densely packed
 calcification shape is round
 cluster shape is round
 distribution is clustered
 blunt duct adenosis
 density is variable
 form is monomorphic
 form is punctate
 distribution is scattered
 sclerosing adenosis
 form is polymorphic
 papillomatosis
 density is high density
 radial scar
 vascular calcification
 calcification shape is linear
 distribution is intermittent
 distribution is parallel

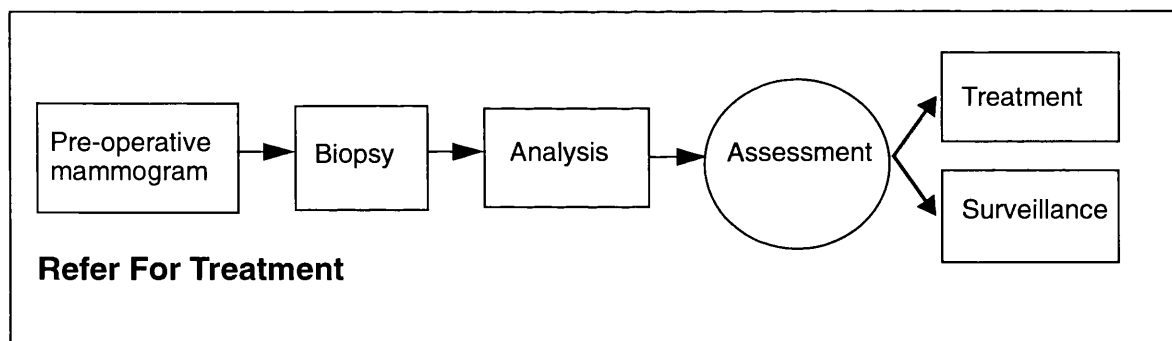
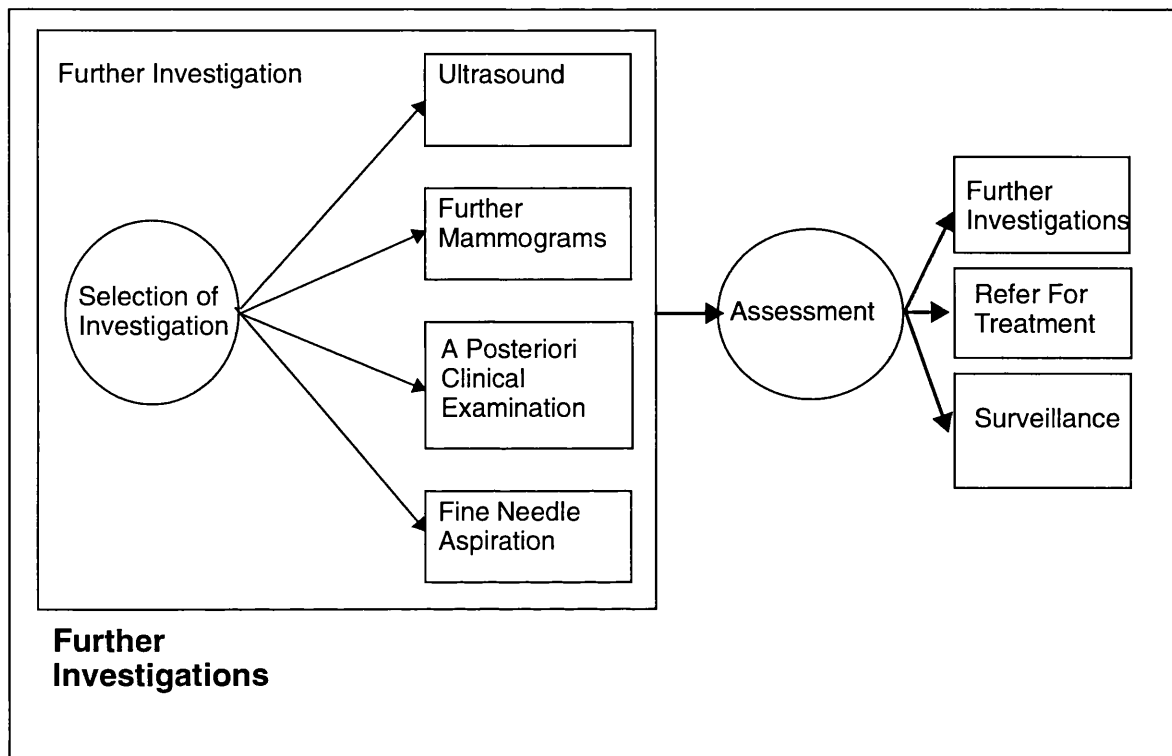
distribution is scattered
form is punctate
variation in size is variable size
cutaneous lesion of breast
location is near skin
calcification shape is round
density is radiolucent centre
size is large
artifact
 artifacts on skin
 distribution is scattered

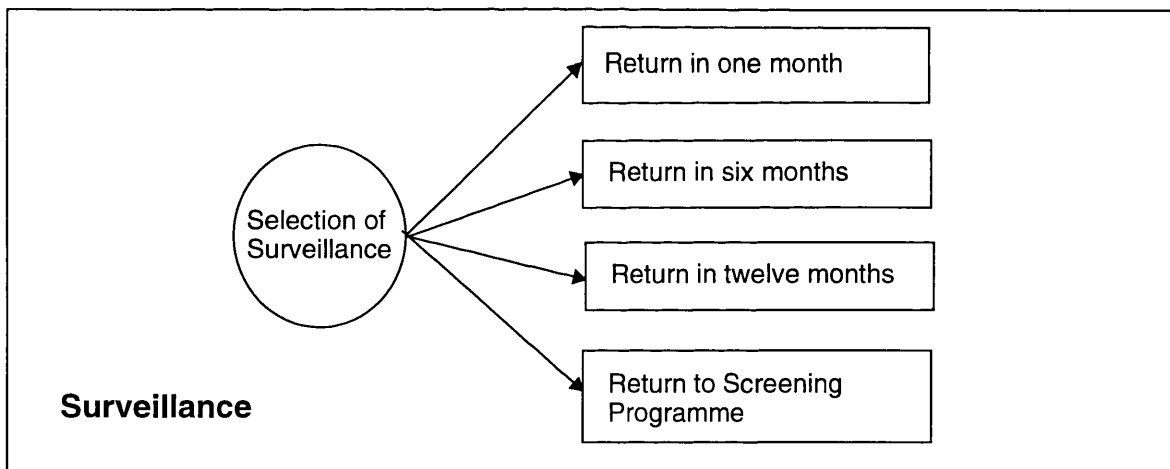
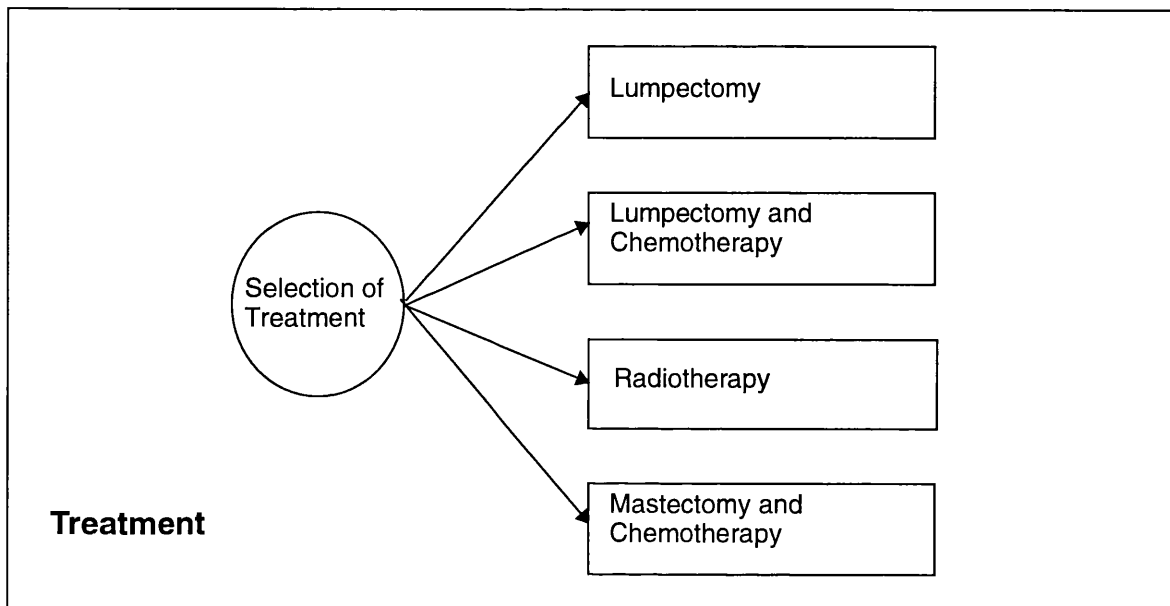
11.5 The Mammography Protocol

11.5.1 Graphical Representation

The following diagrams provide a graphical representation of the protocol used in the prototype, as described in Section 7.4. A box denotes a task or an action, a circle a decision. Temporal constraints are denoted by arrows (an arrow means that the task or action represented by the box pointed at cannot be done until the task or action represented by the box pointed from is complete).







11.5.2 Knowledge Base

The following is a list of the statements used to represent the protocol. The statements are facts of the form:

components of X include Y

indicating that X is a task which includes task or action Y

successors of X include Y

indicating that X must be done before Y

alternatives of X include Y

indicating that where X and Y are both successors of an action, it is necessary to choose between them

indications of X include Y

indicating that where X is an action which could be performed, Y would be an argument in its favour.

components of screening protocol *include* enrolment

components of screening protocol *include* screening mammogram

components of screening protocol *include* return to screen

components of screening protocol *include* refer from screening

successors of enrolment *include* screening mammogram

successors of screening mammogram *include* refer from screening

successors of screening mammogram *include* return to screen

alternatives of return to screen *include* refer from screening

indications of refer from screening *include* stellate lesion

indications of refer from screening *include* breast mass

indications of refer from screening *include* circumscribed lesion

indications of refer from screening *include* calcifications

indications of refer from screening *include* architectural distortion

indications of refer from screening *include* asymmetry

indications of refer from screening *include* density gradient

indications of refer from screening *include* skin thickening

components of enrolment *include* check eligibility

components of screening mammogram *include* mediolateral oblique mammograms

components of screening mammogram *include* first screening report

components of screening mammogram *include* second screening report

successors of mediolateral oblique mammograms *include* first screening report

successors of first screening report *include* second screening report

components of return to screen *include* screening every three years

components of refer from screening *include* history taking

components of refer from screening *include* clinical examination

components of refer from screening *include* additional mammograms

components of refer from screening *include* surveillance

components of refer from screening *include* biopsy

components of refer from screening *include* further investigations

successors of additional mammograms *include* surveillance

successors of additional mammograms *include* biopsy

successors of additional mammograms *include* further investigations

alternatives of biopsy *include* surveillance

alternatives of biopsy *include* further investigations

components of history taking *include* report of patient history

components of clinical examination *include* oncologists report

components of additional mammograms *include* craniocaudal mammograms

components of additional mammograms *include* radiologists report

components of additional mammograms *include* report of assessment meeting

successors of craniocaudal mammograms *include* radiologists report

successors of radiologists report *include* report of assessment meeting

components of assessment meeting *include* report of assessment meeting

indications of further investigations *include* round opacities

indications of biopsy *include* intraductal calcifications

indications of surveillance *include* lobular calcifications

components of further investigations *include* further investigation

components of further investigations *include* assessment meeting

successors of further investigation *include* assessment meeting

components of further investigation *include* further mammograms

components of further investigation *include* ultrasound

components of further investigation *include* fine needle aspiration

components of further investigation *include* a posteriori clinical examination

alternatives of further mammograms *include* fine needle aspiration

alternatives of further mammograms *include* ultrasound

alternatives of further mammograms *include* a posteriori clinical examination

components of further mammograms *include* radiologists report

components of ultrasound *include* radiologists report

components of fine needle aspiration *include* oncologists report

components of a posteriori clinical examination *include* oncologists report

components of assessment meeting *include* report of assessment meeting

components of biopsy *include* pre operative mammogram

components of biopsy *include* excision

components of biopsy *include* histological analysis

components of biopsy *include* assessment of tumour

components of biopsy *include* treatment

components of biopsy *include* surveillance

successors of pre operative mammogram *include* excision

successors of excision *include* histological analysis

successors of histological analysis *include* assessment of tumour

successors of assessment of tumour *include* treatment

successors of assessment of tumour *include* surveillance

components of pre operative mammogram *include* insert needle

components of pre operative mammogram *include* confirm needle location

successors of insert needle *include* confirm needle location

components of excision *include* surgeons report

components of histological analysis *include* histologists report

components of assessment of tumour *include* confirmation of diagnosis

components of assessment of tumour *include* staging

successors of confirmation of diagnosis *include* staging

successors of staging *include* oncologists report

components of treatment *include* lumpectomy and chemotherapy

components of treatment *include* lumpectomy

components of treatment *include* high dose radiotherapy

components of treatment *include* mastectomy and chemotherapy

alternatives of lumpectomy and chemotherapy *include* lumpectomy

alternatives of lumpectomy and chemotherapy *include* high dose radiotherapy

alternatives of lumpectomy and chemotherapy *include* mastectomy and chemotherapy

components *of* surveillance *include* return in a months time

components *of* surveillance *include* screening every six months

components *of* surveillance *include* screening every year

components *of* surveillance *include* screening every three years

alternatives *of* screening every three years *include* return in a months time

alternatives *of* screening every three years *include* screening every six months

alternatives *of* screening every three years *include* screening every year

components *of* lumpectomy and chemotherapy *include* enrol in lumpectomy and chemotherapy

components *of* lumpectomy *include* radiotherapy and chemotherapy

components *of* high dose radiotherapy *include* enrol in high dose radiotherapy

components *of* mastectomy and chemotherapy *include* enrol in mastectomy and chemotherapy

components *of* return in a months time *include* recall patient in a months time

components *of* screening every six months *include* recall patient in six months time

components *of* screening every year *include* recall patient in a years time

components *of* screening every three years *include* recall patient in three years time

References

- [ACR1992] American College of Radiology. *Index for Radiological Diagnoses*. American College of Radiology, 1992.
- [Altman1991] D. G. Altman. *Practical Statistics for Medical Research*. Chapman and Hall, 1991.
- [Astley1993] S. Astley, C. Taylor, and C. Boggis et al. Cue generation and combination for mammographic screening. In David Brogan, editor, *VISUAL SEARCH II*, chapter 15, pages 207–223. 1993.
- [Audet1990] A. M. Audet and S. Greenfield. Medical practice guidelines: Current activities and future directions. *Annals of Internal Medicine*, 113:709–714, 1990.
- [Bae1993] K. T. Bae, M. L. Giger, and C. T. Chen et al. Automatic segmentation of liver structure in CT images. *Medical Physics*, 20(1):71–78, 1993.
- [Bassett1992] L. W. Bassett. Mammographic analysis of calcifications. *Radiologic Clinics of North America*, 30(1):93–105, 1992.
- [Behrman1994] R. E. Behrman and R. M. Kliegman. *Nelson Essentials of Pediatrics*. W. B. Saunders, second edition, 1994.
- [Berdon1992] W. E. Berdon, C. Ruzal-Shapiro, S. J. Abramson, and J. Garvin. The diagnosis of abdominal neuroblastoma: Relative roles of ultrasonography, CT and MRI. *Urologic Radiology*, 14:252–262, 1992.
- [Blois1980] M. S. Blois. Clinical judgement and computers. *New England Journal of Medicine*, 303(4):192–197, 1980.
- [Bonadonna1992] F. Bonadonna, S. Gaglio, M. Giacomini, and C. Ruggiero. Visual knowledge processing in computer-assisted radiology: a consultation system. *Medical Informatics*, 17(1):11–19, 1992.

- [Brettle1994] D. S. Brettle, G. M. Mcleod, and R. J. Oddy et al. Automatic microcalcification localisation using matched fourier filtering. In A. Gale, S. Astley, D. Dance, and A. Cairns, editors, *Digital Mammography*, pages 21–30. Elsevier Science, Amsterdam, 1994.
- [Brodeur1993] G. M. Brodeur, J. P. Pritchard, and F. Berthold et al. Revisions of the international criteria for neuroblastoma diagnosis, staging and response to treatment. *Journal of Clinical Oncology*, 11(8):1466–1477, 1993.
- [Caty1993] M. G. Caty and R. C. Samburger. Abdominal tumors in infancy and childhood. *Pediatric Clinics of North America*, 40:1253–1271, 1993.
- [Chan1988] H.-P. Chan, K. Doi, and C. J. Vyborny et al. Computer-aided detection of microcalcifications in mammograms: Methodology and preliminary clinical study. *Investigative Radiology*, 23(9):664–671, 1988.
- [Chan1995a] H. P. Chan, D. T. Wei, and M. A. Helvie et al. Computer-aided classification of mammographic masses and normal tissue - linear discriminant-analysis in texture feature space. *Physics In Medicine and Biology*, 40:857–876, 1995.
- [Chan1995b] H. P. Chan, S.-C. B. Lo, and B. Sahiner et al. Computer-aided detection of mammographic microcalcifications: pattern recognition with an artificial neural network. *Medical Physics*, 22(10):1555–1567, 1995.
- [Chandrasekaran1988] B. Chandrasekaran. What is a deep expert system? *Knowledge Engineering Review*, 3:183–219, 1988.
- [Chang1988] S. K. Chang, C. W. Yan, D. C. Dimitroff, and T. Arndt. An intelligent image database system. *IEEE Transactions on Software Engineering*, 14(5):681–688, 1988.
- [Chang1990] P. Chang and G. W. Friedland. The role of imaging in screening for prostate cancer. a decision analysis perspective. *Investigative Radiology*, 25:591–605, 1990.
- [Chitre1994] Y. Chitre, A. P. Dhawan, and M. Moskowitz. Classification of mammographic microcalcifications using image structure and cluster features. In A. Gale, S. Astley, D. Dance, and A. Cairns, editors, *Digital Mammography*, pages 31–40. Elsevier Science, Amsterdam, 1994.
- [Ciatto1987] S. Ciatto, L. Cataliotti, and V. Distante. Non-palpable lesions detected with

- mammography: 512 consecutive cases. *Radiology*, 165:99–102, 1987.
- [Claridge1994] E. Claridge and J. H. Richter. Characterisation of mammographic lesions. In A. Gale, S. Astley, D. Dance, and A. Cairns, editors, *Digital Mammography*, pages 241–251. Elsevier Science, Amsterdam, 1994.
- [Clark1990] D. Clark, J. Fox, A. Glowinski, and M. O’Neil. Symbolic reasoning for decision making. In Borcharding K, O. I. Larichev, and D. M. Messick, editors, *Contemporary Issues in Decision Making*, pages 57–75. Elsevier Science Publishers, 1990.
- [Cohn1990] A. I. Cohn, P. L. Miller, and P. R. Fisher et al. Knowledge-based radiologic image retrieval using axes of clinical relevance. *Computers and Biomedical Research*, 23:199–221, 1990.
- [Coiera1995] E. Coiera. Medical informatics. *British Medical Journal*, 310:1381–1387, 1995.
- [Collins1995] D. L. Collins, A. C. Evans, C. Holmes, and T. M. Peters. Automatic 3D segmentation of neuro-anatomical structures from mri. In Y. Bizais, C. Barillot, and R Di Paola, editors, *Information Processing in Medical Imaging*, Computational Imaging and Vision, pages 139–152, Dordrecht, 1995. Kluwer Academic Publishers.
- [Cook1989] H. M. Cook and M. C. Fox. Application of expert systems to mammographic image analysis. *American Journal of Physiologic Imaging*, 4:16–22, 1989.
- [Cootes1994] T. F. Cootes, A. Hill, C. J. Taylor, and J. Haslam. The use of active shape models for locating structure in medical images. *Image and Vision Computing*, 12(6), 1994.
- [Davies1989] D. H. Davies, D. R. Dance, and C. H. Jones. Automatic detection of microcalcifications in digital mammograms. *British Journal of Radiology*, 62:10, 1989.
- [De Dombal1992] F. T. de Dombal, S. Barnes, and V. Dallos et al. How should computer-aided decision-support systems present their predications to the practising clinician. *Theor. Surg.*, 7:111–116, 1992.
- [DeLafontan1994] B. deLafontan, J. P. Daures, and B. Salicru et al. Isolated clustered microcalcifications: Diagnostic value of mammography. *Radiology*, 190:479–483, 1994.

- [De Rosis1992] F. de Rosis, A. Giancaspro, P. Pizzutilo, and G. Pasquariello. A consultation system for medical image analysis. *Expert Systems With Applications*, 5:35–42, 1992.
- [Dhawan1990] A. P. Dhawan and S. Juvvadi. Knowledge-based analysis and understanding of medical images. *Computer Methods and Programs in Biomedicine*, 33:221–239, 1990.
- [DICOM1993] *A draft standard for communications in radiology V3.0*. National Electrical Manufacturers Association, Washington DC, USA, 1993.
- [Dilhuydy1994] M.-H. Dilhuydy. *Anomalies mammographiques infracliniques: apport de l'imagerie dans la discussion des indications chirurgicales*. Editions Bergeret, Bordeaux, France, 1994.
- [D'Orsi1992] C. J. D'Orsi, D. J. Getty, and J. A. Swets et al. Reading and decision aids for improved accuracy and standardization of mammographic diagnosis. *Radiology*, 184:619–622, 1992.
- [D'Orsi1995] C. D'Orsi and M. D. Debor. Communication issues in breast imaging. *Radiologic Clinics of North America*, 33(3):1231–1245, 1995.
- [Dunn1986] G. A. Dunn and A. F. Brown. Alignment of fibroblasts on grooved surfaces described by a simple geometric transform. *Journal of Cell Science*, 83:313–340, 1986.
- [Egan1980] R. L. Egan, M. B. McSweeney, and C. S. Sewell. Intramammary calcifications without an associated mass. *Radiology*, 137:1–7, 1980.
- [Eggli1995] K. D. Eggli, P. Close, and P. W. Dillon et al. Three-dimensional quantitation of pediatric tumour bulk. *Pediatric Radiology*, 25:1–6, 1995.
- [Fam1988] B. W. Fam and S. L. Olson. The detection of calcification clusters in film-screen mammograms: A detailed algorithmic approach. In *SPIE Vol. 914 Medical Imaging II*, pages 620–634, 1988.
- [Feig1992] S. A. Feig. Breast masses: Mammographic and sonographic evaluation. *Radiologic Clinics of North America*, 30(1):67–92, 1992.
- [Feig1995] S. A. Feig and M. J. Yaffe. Digital Mammography, Computer-Aided Diagnosis and Telemammography. *Radiologic Clinics of North America*, 33(3):1205–1230, 1995.

- [Ferguson1981] G. A. Ferguson. *Statistical Analysis in Psychology and Education*. McGraw-Hill, 1981.
- [Fox1988] J. Fox and N. Walker. Knowledge based interpretation of medical images. In M A. Viergever and A. Todd-Pokropek, editors, *Mathematics and Computer Science in Medical Imaging*, pages 241–267. Springer-Verlag, Berlin, 1988.
- [Fox1990] J. Fox, A. Glowinski, and C. Gordon et al. Logic engineering for knowledge engineering: design and implementation of the ‘Oxford System of Medicine’. *Artificial Intelligence in Medicine*, 2:323–339, 1990.
- [Fox1995] J. Fox, S. K. Das, and D. Elsdon et al. Decision-making and planning by autonomous agents: a generic architecture for safety-critical applications. In *Proceedings of Expert Systems 95*, pages 159–169, 1995.
- [Fox1997] J. Fox, N. Johns, and A. Rahmanzadeh. Protocols for medical procedures and therapies: a provisional description of the PROforma language tools. In *Proceedings of Artificial Intelligence in Medicine Europe*, page submitted, 1997.
- [Fritsch1994] D. S. Fritsch, S. M. Pizer, and B. S. Morse et al. The multiscale medial axis and its applications in image registration. *Pattern Recognition Letters*, 15:445–452, 1994.
- [Gale1993a] A. Gale. Human response to visual stimuli. In W. R. Hendee and P. N. Temple, editors, *The Perception of Visual Information*, pages 115–133. Springer-Verlag, 1993.
- [Gale1993b] A. G. Gale, A. R. M. Wilson, and E. J. Roebuck. Mammographic screening: Radiological performance as a precursor to image processing. In *Proceedings of Biomedical Image Processing and Biomedical Visualization*, pages 458–464. SPIE, 1993.
- [Garfagni1994] H. Garfagni, B. Gibaud, and W. Stut et al. MIMOSA: a functional model of the management of medical images. *Medical Informatics*, 19:95-108, 1994.
- [Giger1994] M. L. Giger, C. J. Vyborny, and R. A. Schmidt. Computerized characterization of mammographic masses. *Cancer Letters*, 77:201–211, 1994.
- [Goldberg1992] V. Goldberg, A. Manduca, and D. L. Ewert et al. Improvement in specificity of ultrasonography for diagnosis of breast tumours by means of artificial

- intelligence. *Medical Physics*, 19(6):1475–1481, 1992.
- [Good1992] B. Good. *Effect of computer generated feedback on the performance of radiologists*. PhD thesis, University of Pittsburgh, 1992.
- [Greenes1989] R. A. Greenes. The radiologist as clinical activist: A time to focus outward. In S. K. Mun, M. Greberman, and W. R. Hendee, editors, *Proceedings of the First International Conference on Image Management and Communication in Patient Care*, pages 136–140. IEEE Computer Society Press, 1989.
- [Greenes1991] R. A. Greenes. A building block approach to application development for education and decision support in radiology. *Journal of Digital Imaging*, 4(4):213–225, 1991.
- [Grimshaw1993] J. M. Grimshaw and I. T Russell. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *The Lancet*, 342:1317–22, 1993.
- [Grosky1992] W. I. Grosky and R. Mehrota. Image database management. *Advances in Computers*, 35:237–292, 1992.
- [Guillemet1996] H. Guillemet, H. Benali, E. Kahn, and R. di Paola. Detection and characterisation of microcalcifications in digital mammography. In K. Doi, editor, *Third International Workshop on Digital Mammography*, page in press. Elsevier Science, Amsterdam, 1996.
- [Guillemet1997] H. Guillemet. Reconnaissance de microcalcifications en mammographie numérique. Approche hiérarchique. PhD thesis, Institut National des Télécommunications, Paris, 1997.
- [Gupta1995] R. Gupta and P. E. Undrill. The use of texture analysis to delineate suspicious masses in mammography. *Physics In Medicine and Biology*, 40:835–855, 1995.
- [Hall1988] F. M. Hall, J. M. Storella, and D. Z. Siverstone et al. Nonpalpable breast lesions: Recommendations for biopsy based on suspicion of carcinoma at mammography. *Radiology*, 167:353–358, 1988.
- [Hanley1983] J. A. Hanley and B. J. McNeil. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 148:839–843, 1983.
- [Heathfield1993] H. Heathfield and J. Wyatt. Philosophies for the design and development

- of clinical decision-support systems. *Methods of Information in Medicine*, 32:1–9, 1993.
- [Heilbrun1994] K. S. Heilbrun. The American College of Radiology’s mammography lexicon: Barking up the wrong tree. *The American Journal of Roentgenology*, 162:593–594, 1994.
- [Highnam1994] R. Highnam, M. Brady, and B. Shepstone. Computing the scatter component of mammographic images. In *Proceedings of IEEE Medical Imaging 1994*, pages 180–192, 1994.
- [Huang1993] J. Huang, J. Fox, C. Gordon, and A. Jackson-Smale. Symbolic decision support in medical care. *Artificial Intelligence in Medicine*, 5:415–430, 1993.
- [Huo1995] Z. Huo, M. L. Giger, and C. J. Vyborny et al. Analysis of spiculation in the computerized classification of mammographic masses. *Medical Physics*, 22:1569–1579, 1995.
- [Hutt1994] I. W. Hutt, S. M. Astley, and C. R. M. Boggis. Prompting as an aid to diagnosis in mammography. In A. Gale, S. Astley, D. Dance, and A. Cairns, editors, *Digital Mammography*, pages 389–388. Elsevier Science, Amsterdam, 1994.
- [Jackson1993] V. P. Jackson, R. E. Hendrick, S. A. Feig, and D. B. Kopans. Imaging of the radiographically dense breast. *Radiology*, 188:297–301, 1993.
- [Kegelmeyer1994] W. P. Kegelmeyer Jnr. Computer-aided mammographic screening for spiculated lesions. *Radiology*, 191(2):331–337, 1994.
- [Johnston1994] M .E. Johnston, K .B. Langton, B. Hayes, and A. Mathieu. Effects of computer-based clinical decision support systems on clinician performance and patient outcome. *Annals of Internal Medicine*, 120:135–142, 1994..
- [Karssemeijer1988] N. Karssemeijer, L. J. T. O. van Erning, and E. G. J. Eijkman. Recognition of organs in CT-image sequences: a model guided approach. *Computers and Biomedical Research*, 21(5):434–48, 1988.
- [Karssemeijer1992] N. Karssemeijer. Stochastic model for automated detection of calcifications in digital mammograms. *Image and Vision Computing*, 10(6):369–375, 1992
- [Keen1978] P. G. W. Keen and M. S. Scott-Morton. *Decision Support Systems: An Organizational Perspective*. Addison Wesley, Reading, Massachusetts, USA, 1978.

- [Kegelmeyer1994] W. P. Kegelmeyer and M. C. Allmen. Dense feature maps for the detection of microcalcifications. In A. Gale, S. Astley, D. Dance, and A. Cairns, editors, *Digital Mammography*, pages 3–13. Elsevier Science, Amsterdam, 1994.
- [Keravnou1993] E. T. Keravnou, F. Dams, and J. Washbrook et al. Modelling diagnostic skills. Technical report, Dept. of Computer Science, University College London, 1993.
- [Kippenham1992] J. S. Kippenham. Evaluation of a neural-network classifier for PET scans of normal and alzheimer’s disease subjects. *Journal of Nuclear Medicine*, 33:1459–67, 1992.
- [Kobashi1995] M. Kobashi and L. G. Shapiro. Knowledge-based organ identification from CT images. *Pattern Recognition*, 28:475–491, 1995.
- [Kofakis1992] P. Kofakis and S. Orphanoudakis. Image indexing by context. In M. Osteaux, A. R. Bakker, and D. Bell et al., editors, *A Second Generation PACS Concept*, pages 250–293. Springer-Verlag, 1992.
- [Krause1995] P. Krause, S. Ambler, Morten Elvang-Goransson, and J. Fox. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence*, 11(1):113–131, 1995.
- [Kuhns1989] L. R. Kuhns, J. R. Thornbury, and D. G. Fryback. *Decision Making in Imaging*. Year Book Medical Publishers, Chicago, 1989.
- [Kulikowski1995] C. A. Kulikowski, L. Gong, and R. S. Mezrich. Knowledge-based medical image analysis and representation for integrating content definition with the patient record. *Methods of Information in Medicine*, 34:371–381, 1994.
- [Kundel1972] H. L. Kundel and P. S. La Follette. Visual search patterns and experience with radiological images. *Radiology*, 103:523–528, 1972.
- [Kundel1990] H. L. Kundel, C. F. Nodine, and E. A. Krupinski. Computer-displayed eye position as a visual aid to pulmonary nodule interpretation. *Investigative Radiology*, 25:890–896, 1990.
- [Lai1986] S.-M. Lai, X. Li, and W. F. Bischof. On techniques for detecting circumscribed masses in mammograms. *IEEE Transactions on Medical Imaging*, 5(1):8–15, 1986.
- [Lanyi1987] M. Lanyi. *Diagnosis and Differential Diagnosis of Microcalcifications*.

Springer-Verlag, 1987.

- [Ledley1959] R. S. Ledley and L. B. Lusted. Reasoning foundations of medical diagnosis. *Science*, 130(3366):9–21, 1959.
- [Lefebvre1991] F. Lefebvre, R. Gilles, J. Masselot, and E. Kahn. Computer classification of clustered microcalcifications in mammograms. *Proceedings of Computer Assisted Radiology*, pages 582–587, 1991.
- [Le Gal1973] M. Le Gal, J.-C. Durand, M. Laurent, and D. Pellier. Conduite a tenir devant une mammographie revelatrice de microcalcifications groupées sans tumeur palpable. *La Nouvelle Presse Medicale*, 5:1623–1627, 1973.
- [Li1993] C. Li, D. Goldgof, and L. Hall. Knowledge-based classification and tissue labeling of MR images of human brain. *IEEE Transactions on Medical Imaging*, 12(4):740–750, 1993.
- [Li1994] H. Li, R. Deklerck, and J. Cornelis. Integration of multiple knowledge sources in a system for brain ct-scan interpretation based on the blackboard model. In *Second International Conference on the Practical Applications of Prolog*, pages 336–343, 1994.
- [Lo1995] J. Y. Lo, J. A. Baker, P. J. Kornguth, and C. E. Floyd. Computer-aided diagnosis of breast cancer: artificial neural-network approach for optimized merging of mammographic features. *Academic Radiology*, 2:841–850, 1995.
- [Lodwick1986] G. S. Lodwick. Computers in radiologic diagnosis. *Applied Radiology*, pages 61–65, 1986.
- [Magnin1989] I. Magnin, M. Alaoui, and A. Bremond. Automatic microcalcifications pattern recognition from x-ray mammograms. *SPIE vol 1137: Science and Engineering of Medical Imaging*, pages 170–175, 1991.
- [Marr1982] D. Marr. *Vision*. W. H. Freeman, San Fransisco, 1982.
- [Matsumoto1992] T. Matsumoto, H. Yoshimura, and K. Doi et al. Potential usefulness of computerized nodule detection in screening programmes for lung cancer. *Investigative Radiology*, 27:471–475, 1992.
- [Matsuyama1993] T. Matsuyama. Expert systems for image processing, analysis, and recognition for computer vision. *Advances in Electronics and Electron Physics*, 86:81–171, 1993.

- [Meinzer1996] H-P. Meinzer and U. Engelmann. Medical images in integrated health care workstations. In J. H van Bemmelen and A. T. McCray, editors, *Yearbook of Medical Informatics: Integration of Information for Patient Care*, pages 87–94. Schattauer Verlagsgesellschaft, Stuttgart-New York, 1996.
- [Mendelson1995] E. B. Mendelson. The development and meaning of appropriateness guidelines. *Radiologic Clinics of North America*, 33(3):1081–1084, 1995.
- [Miller1992] P. Miller and S. Astley. Classification of breast tissue by texture analysis. *Image and Vision Computing*, 10(5), 1992.
- [Miller1990] R. A. Miller and F. E. Masarie. The demise of the greek oracle model for medical diagnostic systems. *Methods of Information in Medicine*, 29:1–3, 1990.
- [Monsees1995] B. S. Monsees. Evaluation of breast microcalcifications. *Radiologic Clinics of North America*, 33(3):1109–1121, 1995.
- [Nazif1984] A. Nazif and M. D. Levine. Low level image segmentation: an expert system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:555–557, 1984.
- [Niemann1985a] H. Niemann. A homogeneous architecture for knowledge based image understanding systems. *Proc 2 Conf Artificial Intelligence Applications*, pages 88–93, 1985.
- [Niemann1985b] H. Niemann, H. Bunke, and I. Hofmann et al. A knowledge based system for analysis of gated blood pool studies. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7:246–259, 1985.
- [Niemann1990] H. Niemann, G. Sagerer, S. Schroder, and F. Kummert. ERNEST: A semantic network system for pattern understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:883–906, 1990.
- [Nishikawa1993] R. M. Nishikawa, Y. Jiang, and M. L. Giger, et al. Characterization of the mammographic appearance of microcalcifications. In *Image Processing 1993: SPIE vol. 1898*, pages 422–429. SPIE, 1991.
- [Nishikawa1994] R. M. Nishikawa, Y. Jiang, and M. L. Giger et al. Performance of automated CAD schemes for the detection and classification of clustered microcalcifications. In A. Gale, S. Astley, D. Dance, and A. Cairns, editors, *Digital Mammography*, pages 13–20. Elsevier Science, Amsterdam, 1994.

- [Orphanoudakis1994] S. C. Orphanoudakis, C. Chronaki, and S. Kostomanolakis. I²C: A system for the indexing, storage and retrieval of medical images by content. *Medical Informatics*, 19(2):109–123, 1994.
- [Oza1993] A. M. Oza and N. F. Boyd. Mammographic parenchymal patterns: A marker of breast cancer risk. *Epidemiologic Reviews*, 15(1):196–207, 1993.
- [Pal1993] N. R. Pal and S. K. Pal. A review of image segmentation techniques. *Pattern Recognition*, 26(9):1277–1294, 1993.
- [Parker1994] J. Parker, D. R. Dance, and J. Suckling. The hysteresis technique for detection of calcifications in digital mammograms. In A. Gale, S. Astley, D. Dance, and A. Cairns, editors, *Digital Mammography*, pages 41–51. Elsevier Science, Amsterdam, 1994.
- [Parker1995] J. Parker, D. R. Dance, and D. H. Davies et al. Classification of ductal carcinoma in-situ by image-analysis of calcifications from digital mammograms. *British Journal of Radiology*, 68:150–159, 1995.
- [Patnick1996] J. Patnick, editor. *NHS Breast Screening Programme Review 1996*. NHS Breast Screening Programme, Old Fulwood Rd, Sheffield, England, 1996.
- [Patrick1991] E. A. Patrick, M. Moskowitz, V. T. Mansukhani, and E. I. Gruenstein. Expert learning system network for diagnosis of breast calcifications. *Investigative Radiology*, 26:534–539, 1991.
- [Pauli1996] R. Pauli, S. Hammond, J. Cooke, and J. Ansell. Radiographers as film readers in screening mammography: an assessment of competence under test and and screening conditions. *British Journal of Radiology*, 69(817):10–14, 1996.
- [Pavlidis1992] T. Pavlidis. Why progress in machine vision is so slow. *Pattern Recognition Letters*, 13:221–225, 1992.
- [Peeters1989] P. H. M. Peeters, A. L. M. Verbeek, and J. H. C. L. Hendriks et al. The occurrence of interval cancers in the Nijmegen screening programme. *British Journal of Cancer*, 59:929–932, 1989.
- [Piraino1989] D. W. Piraino, B. J. Richmond, and M. Uetani et al. Problems in applying expert system technology to radiographic image. *Journal of Digital Imaging*, 2(1):21–26, 1989.
- [Pizer1990] S. M. Pizer and B. M. ter Haar Romeny. Fundamental properties of medical

- image perception. *Journal of Digital Imaging*, 4:1–20, 1990.
- [Powell1983] R. W. Powell, M. B. McSweeney, and C. E. Wilson. X-ray calcifications as the only basis for breast biopsy. *Annals of Surgery*, 197:555–559, 1983.
- [Pritchard1995] J. Pritchard. Megatherapy for advanced neuroblastoma-rationale and role. *European Journal of Cancer*, 31:134–136, 1995.
- [Qian1995] W. Qian, L. P. Clarke, and B. Zheng et al. Computer-assisted diagnosis for digital mammography. *IEEE Engineering In Medicine and Biology Magazine*, 14:561–569, 1995.
- [Reeder1992] M. Reeder and W. Bradley, editors. *Reeder and Felson's Gamuts in Radiology*. Springer-Verlag, third edition, 1992.
- [Renaud1994] J-L. Renaud-Salis. Distributed clinical management information systems. In *Knowledge and Decisions in Health Telematics*, pages 139–146. IOS Press, Amsterdam, 1994.
- [Robertson1989] C. L. Robertson and D. B. Kopans. Communication problems after mammographic screening. *Radiology*, 172:443–444, 1989.
- [Robinson1992] G. P. Robinson, A. C. F. Colchester, L. D. Griffin, and D. J. Hawkes. Integrated skeleton and boundary shape representation for medical image interpretation. *Lecture Notes in Computer Science*, 588:725–729, 1992.
- [Robinson1994] G. P. Robinson, A. C. F. Colchester, and L. D. Griffin. Model-based recognition of anatomical objects from medical images. *Image and Vision Computing*, 12(8):499–507, 1994.
- [Rogers1995] E. Rogers. VIA RAD: a blackboard system for diagnostic radiology. *Artificial Intelligence in Medicine*, 7:343–360, 1995.
- [Rosenfeld1983] A. Rosenfeld. On connectivity properties of grayscale pictures. *Pattern Recognition*, 16(1):44–50, 1983.
- [Seltzer1992] S. E. Seltzer, B. J. McNeil, and C. J. D Orsi et al. Combining evidence from multiple imaging modalities: a feature-analysis method. *Computerized Medical Imaging and Graphics*, 16(6):373–380, 1992.
- [Shen1993a] L. Shen, R. M. Rangayyan, and J. E. Leo Desautels. An automatic detection and classification system for calcifications in mammograms. In *Proceedings of Biomedical Image Processing and Biomedical Visualization*, pages 799–817.

SPIE, 1993.

- [Shen1993b] L. Shen, R. M. Rangayyan, and J. E. Leo Desautels. Detection and classification of mammographic calcifications. *International Journal of Pattern Recognition and Artificial Intelligence*, 7:1403–1416, 1993.
- [Shortliffe1991] E. Shortliffe. Clinical decision-support systems. In E. Shortliffe and L. E. Perreault, editors, *Medical Informatics*, chapter 15, pages 466–502. Addison-Wesley, 1991.
- [Sickles1986] E. Sickles. Breast calcifications: Mammographic interpretation. *Radiology*, 160:289–293, 1986.
- [Sickles1995] E. A. Sickles. Management of probably benign breast lesions. *Radiologic Clinics of North America*, 33(3):1123–1130, 1995.
- [Simpson1996] W. Simpson, F. Neilson, and P. J. Kelly. Descriptive terms for mammographic abnormalities: observer variation in application. *Clinical Radiology*, 51:709–713, 1996.
- [Sivewright1994] G. J. Sivewright and P. J. Elliot. Interactive region and volume segmenting volumes in MR and CT images. *Medical Informatics*, 19:71–80, 1994.
- [Smets1990] C. Smets, F. van de Werf, P. Suetens, and A. Oosterlinck. An expert system for the labeling and 3D reconstruction of the coronary arteries from two projections. *International Journal of Cardiac Imaging*, 5:145–54, 1990.
- [Sparacia1996] G. Sparacia, M. Tartamella, and M. Midiri et al. RADPA -a World-Wide Web radiology server on the Internet. In H. U. Lemke, M. W. Vannier, K. Inamura, and A. G. Farman, editors, *Computer Assisted Radiology 1996*, page 636 640, Amsterdam, 1996. Elsevier.
- [Stewart1994] B.D. Stewart, A. Y. Cairns, and I. W. Ricketts et al. The analysis of breast abnormalities by computer. In A. Gale, S. Astley, D. Dance, and A. Cairns, editors, *Digital Mammography*, pages 251–261. Elsevier Science, Amsterdam, 1994.
- [Strickland1994] R. N. Strickland and H. I. Hahn. Detection of microcalcifications using wavelets. In A. Gale, S. Astley, D. Dance, and A. Cairns, editors, *Digital Mammography*, pages 79–91. Elsevier Science, Amsterdam, 1994.
- [Sucar1991] L. E. Sucar, D. F. Gillies, and D. A. Gillies. Handling uncertainty in

- knowledge-based computer vision. *Lecture Notes In Computer Science*, 548:328–332, 1991.
- [Suckling1994] J. Suckling, J. Parker, and D. Dance et al. The mammographic image analysis society digital mammogram database. In A. Gale, S. Astley, D. Dance, and A. Cairns, editors, *Digital Mammography*, pages 375–378. Elsevier Science, Amsterdam, 1994.
- [Swett1987] H. A. Swett and P. L. Miller. ICON: A computer-based approach to differential diagnosis in radiology. *Radiology*, 163:555–558, 1987.
- [Swett1993] H. A. Swett, M. L. Giger, and K. Doi. Computer vision and decision support. In W. R. Hendee and P. N. Temple, editors, *The Perception of Visual Information*, pages 274–315. Springer-Verlag, 1993.
- [Tabar1983] L. Tabar and P. Dean. *Teaching Atlas of Mammography*. Thieme, Stuttgart, 1983.
- [Taylor1994] P. Taylor, S. Hajnal, M.-H. Dilhuydy, and B. Barreau. Measuring image texture to separate difficult from easy mammograms. *British Journal of Radiology*, 67:456–463, 1994.
- [Tourassi1993] G. D. Tourassi and C. E. Floyd. Artificial neural networks for single-photon emission computed tomography. *Investigative Radiology*, 28(8):671–678, 1993.
- [Valatx1994] L. Valatx, I. E. Magnin, and A. Bremond. Automatic microcalcifications and opacities detection using a multiscale approach. In A. Gale, S. Astley, D. Dance, and A. Cairns, editors, *Digital Mammography*, pages 51–59. Elsevier Science, Amsterdam, 1994.
- [Walker1990] N. Walker. *Biomedical Image Interpretation*. PhD thesis, Queen Mary and Westfield College, 1990.
- [Warner1992] E. Warner, G. Lockwood, and M. Math et al. The risk of breast cancer associated with mammographic parenchymal patterns: A meta-analysis of the published literature to examine the effect of method of classification. *Cancer Detection and Prevention*, 16(1):67–72, 1992.
- [Weierich1993] P. Weierich, D. Wetzel, and H. Niemann. AUDIGON - a medical expert system with new judgement functions for the fuzzy rulebase. In *Proceedings of*

First European Congress on Fuzzy and Intelligent Technologies, pages 472–477, 1993.

[Wiederhold1989] G. Wiederhold, J. Brinkley, R. Samadani, and C. R. Clauer. Model-driven image analysis to augment databases. In T. I. Kunii, editor, *Visual database systems: proceedings of the IFIP TC2/WG 2.6 working*, pages 159–180. Elsevier Science, Amsterdam, 1989.

[Winograd1986] T. Winograd and F. Flores. *Understanding Computers and Cognition*. Addison Wesley, Reading, Massachusetts, 1986.

[Wittgenstein1958] L. Wittgenstein. *Philosophical Investigations*. Basil Blackwell, Oxford, 1958.

[Wolfe1976a] J. N. Wolfe. Breast patterns as an index of risk for developing breast cancer. *American Journal of Roentgenology*, 126:1130–1139, 1976.

[Wu1993] Y. Wu, M. L. Giger, and K. Doi et al. Artificial neural networks in mammography: Application to decision making in the diagnosis of breast cancer. *Radiology*, 187:81–87, 1993.

[Wyatt1991] J. Wyatt and D. Spiegelhalter. Field trials of medical decision-aids: potential problems and solutions. In P. Clayton, editor, *Proceedings of Fifteenth Annual on Computer Applications in Medical Care*, pages 3–7. American Medical Informatics Association, 1991.

[Yin1994] F. F. Yin, M. L. Giger and K. Doi et al. Computerized detection of masses in digital mammograms. *Medical Physics*, 21:445–452, 1994.

[Zheng1995] B. Zheng, Y.H. Chang, and D. Gur. Computerized detection of masses in digitized mammograms using single-image segmentation and a multilayer topographic feature analysis. *Academic Radiology*, 2:959–966, 1995.