Structured Computer-based Training in the Interpretation of Neuroradiological Images

Sharples, M.[1], Jeffery, N.P. [3], du Boulay, B. [2], Teather B.A. [3],

Teather, D. [3] & du Boulay, G.H. [3,4]

1. School of Electronic and Electrical Engineering, University of Birmingham, UK
2. School of Cognitive and Computing Sciences, University of Sussex, UK
3. Department of Medical Statistics, De Montfort University, Leicester, UK
4. Institute of Neurology, London, UK

Address for correspondence:
Professor Mike Sharples
Educational Technology Research Group
School of Electronic and Electrical Engineering
University of Birmingham
Edgbaston
Birmingham
B15 2TT
UK
Fax: +44 121 414 4291
Email: m.sharples@bham.ac.uk

Keywords: Medical imaging, computer-based training, statistical diagnosis, image description

Summary

Computer-based systems may be able to address a recognised need throughout the medical profession for a more structured approach to training. We describe a combined training system for neuroradiology, the MR Tutor, that differs from previous approaches to computer-assisted training in radiology in that it provides case-based tuition whereby the system and user communicate in terms of a well-founded Image Description Language. The system implements a novel method of visualisation and interaction with a library of fully described cases utilising statistical models of similarity, typicality and disease categorization of cases. We describe the rationale, knowledge representation and design of the system, and provide a formative evaluation of its usability and effectiveness.

Introduction

The specialty of radiology, in common with most of medicine, is gradually developing a more systematic approach to training, replacing the traditional mixture of ad hoc apprenticeship and formal lectures with a combination of structured tuition and case-based experiential learning. This is intended to meet a long-recognised need for clinicians to encapsulate general medical knowledge within the development of skills through diagnostic practice [1]. A structured approach to training can have the additional benefit of equipping learners with a coherent 'conceptual framework': an appropriately defined and organised notation that enables them to externalise, reflect on and share diagnostic knowledge.

Computer-enhanced reporting, whereby radiologists describe images by means of a structured notation for abnormal image features supplemented by a computer-generated diagnosis, has been shown (for mammography) to improve the diagnostic accuracy of general radiologists to equal that of specialists [2]. In Magnetic Resonance Imaging (MRI), the predictive power of data generated from structured reports has been demonstrated for two key clinical problems [3, 4].

Computer-based Training in Radiology

Computer-based systems to assist in radiology training include videodisk applications [5, 6], hypermedia-based programs (such as FACT/FILE [7], RADMAC [8] and CT-The Game [9]) and teaching files available on the World Wide Web. They offer a valuable backup to human or textbook teaching, particularly if the teaching is linked explicitly to the specific reference material, but they are limited by the shallowness of relations between the images and the textual material. The descriptive labels and pre-prepared texts do not form a systematic representation of knowledge that could be interpreted by the computer to provide active teaching, or to answer complex queries.

Few active knowledge-based tutors have been developed that can call on structured representations of domain knowledge to generate sequences of teaching actions, provide comparisons between cases and diagnose learner misconceptions. The CT Brain Tutor developed at the Medical College of Georgia [10] trains radiology residents in the diagnosis of brain tumours from CT and MRI scans. It presents images from a library of 120 cases indexed by case history and radiological features. The indicative features are disease based ('tumour', 'oedema', 'calcification') and do not form a comprehensive image description language.

The Radiology Tutor [11] generated Socratic tutorials about the appearance of chest X-rays. It was able to conduct a viable dialogue with a trainee in natural language but was constrained by its limited natural language understanding, only being able to parse a limited subset of English. RUI [12] is a generalisation of the Radiology Tutor into a general-purpose authoring shell. Demonstration tutors have been implemented in RUI for a limited subset of diagnostic problems in MR, CT, Ultrasound and X-ray. RUI and the Radiology Tutor take a conventional Intelligent Tutoring Systems approach to training, with the system acting as a simulated human tutor.

RadTutor is a prototype system to train radiologists in diagnosing mammograms exhibiting breast diseases [13]. A cognitive account of diagnostic reasoning in mammography underlies its instructional principles (multiplicity, activeness, accommodation and adaptation, and authenticity) and methods (including modelling, coaching, fading of assistance, structured problem solving, and situated learning).

Comparison between the MR Tutor and Other Systems

In this paper we describe a system, the MR Tutor, that teaches a structured approach to lesion description. It is based on a constructivist approach to learning, whereby the trainee is helped to acquire a well-structured approach to describing abnormal features of images by engaging in an active process of case-based reporting. The MR Tutor differs from other computer-based training systems for radiology in three fundamental ways:

– Its teaching and diagnostic support is based on a consistent and diagnostically powerful Image Description Language (IDL).

– It implements a novel method of visualisation, the overview plot, to display the distribution of cases across diseases.

– It utilises statistical-based models of similarity, typicality and disease categorization estimated from an archive of expert case descriptions and diagnoses. It enables the trainee to explore the archive, by viewing the cases according to similarity and typicality on the plot, by inspecting the MR images for selected cases, and by comparing their structured case descriptions. For any case, the system offers a teaching session in which the trainee is guided towards a full and correct description using the terminology of the IDL.

This paper provides an overview of the system requirements, structure, functions and interface of the MR Tutor and then describes the system in terms of the distinguishing aspects of its design. It concludes with the findings of a formative evaluation of the system and plans for future development.

System Requirements

The specific requirements for the MR Tutor were derived through socio-cognitive engineering: a process of describing and analysing the complex interactions between people and computer-based technology so as to inform the design of socio-technical systems that support human work, learning, communication and culture [14]. The investigations have included an analysis of the literature on cognitive and perceptual processes of medical image interpretation and the task of radiological reporting; detailed observations of reporting and radiological training in two medical institutions (a general teaching hospital and a specialist imaging centre); and interviews with trainee radiologists, consultants and lecturers to elicit differing conceptions of training practice, issues of concern and opportunities for introducing new technology in support of training [15]. The main findings of these studies are summarised here.

Radiological expertise is based on two kinds of skill: the swift and accurate processing of normal appearance, and the ability to distinguish disease from normal variation in appearance [16]. Thus, skill development in radiology requires exposure to, and

reporting of, a large range of images, so that recognition of varied normal anatomy becomes automatised and cognitive resources can be devoted to the process of describing abnormal appearance.

Lesgold et al. [17] propose that radiologists carry out a multi-stage process of interpretation. On first seeing a film they automatically invoke a mental schema that covers the salient abnormal features, resulting in one or more tentative diagnoses. This triggers a process of active search for other cues in the image along with case data and medical knowledge that might constrain the interpretation. Lastly, they articulate their findings as a verbal report. A more recent study by Azevedo and Lajoie [18] indicates that perception and problem solving are tightly coupled, with experts creating an active mapping between their perception of the visual image and their evolving hypotheses about competing diagnoses.

Our own investigations support these accounts of radiological reporting. They also confirm the two main teaching events identified by Azevedo and Lajoie of guiding junior residents systematically through the interpretation process and of scaffolding the interpretation of more senior trainees by offering hints and indications of where to direct attention. In summary, our investigations indicated that a computer system to support the training of radiologists should:

− base the training on a large library of cases representative of radiological practice;

− provide a means of making rapid comparisons between cases by similarity of diagnostically relevant features;

− expose the trainee to cases in an order that promotes understanding and retention;

− help the trainee to make rapid, accurate initial judgements;

− help the trainee to integrate fragmentary knowledge into more general structural schemata;

− help the trainee to reflect on experience gained and to integrate general and situated knowledge;

− be implemented on a personal computer, for use as part of self study at home or work.

The current MR Tutor, as described in this paper, meets these criteria. The training is based on an archive of 1200 cases, fully described in the terms of the IDL, that constitute a representative sample of the patient clientele from two different imaging centres (a specialist tertiary referral centre and a more general imaging centre) dealing with very varied disease. The current system displays all cases for two diseases from the archive and further disease sets are being added. The trainee can make rapid comparisons between cases by selecting them from the overview plot and viewing their associated images and structured descriptions. The system teaches a consistent method of reporting based on a structured language, supported by a context-sensitive help system giving precise definitions of terms, to help the trainee acquire a structural schema for describing abnormal images. It relates the trainee's description to other cases in the archive by means of a visual display that helps the trainee reflect on the experience of describing each case. It is implemented in Java and runs on personal computers.

Overview of the MR Tutor

The training provided by the MR Tutor has a number of pedagogic aims. The first is to help the trainee to acquire and apply a structured language for describing abnormal features of MR images of the head. The second aim is to help trainees to gain an appreciation of the range, distribution and variability of features presenting with a particular disease. A third aim is to help the trainee to discriminate between confusable diseases. We intend to carry out a series of studies, using the MR Tutor as a test environment, to investigate these aims.

The system is intended for trainees in the specialism of neuroradiology. These are highly educated professionals who have already gained a medical qualification and experience in general radiology. They have intrinsic motivation and at least a basic competence in working with computer-based technology. Computer-based "coached practice" environments such as SOPHIE [19] SHERLOCK [20] and IDM [21] are examples of successful systems for training in complex skills. The theory of cognitive apprenticeship [22] that underlies these systems is appropriate for training in the professions and has informed the design of the MR Tutor.

Figure 1 here

Structurally and in presentation to the user the MR Tutor is divided into two main modules (shown on the left and right sides of Figure 1), to enable the trainee to explore the archive of cases and to engage in case based training. Both modules request images and structured descriptions from the case archive and interact with the user through a common graphical interface. The basic operation is that a) the user explores the case archive, b) either the user or the computer selects a case for teaching, c) the user describes the case using the terminology of the IDL and receives tutorial feedback from the system. Figure 2 shows the main flow of tasks supported by the system, though the user may at any point break the flow and switch tasks.

Figure 2 here

The interface of the prototype MR Tutor is divided into two halves that indicate its two main functions (see Figure 3). The right side of the screen allows the trainee to interact with the image archive through an interactive "overview plot" that provides a direct representation of typicality, similarity and category membership of the cases. It is computed automatically from the structured descriptions of the lesions and enables a trainee to view the distribution of cases within and across diseases. The plot can be interpreted as a "radar screen" with each point representing a case from the archive. The small crosses represent the centres of disease categories and the ellipses show scaled "contours of typicality" for the categories. The closer a case point lies to the centre of a disease category, the more typical it is of that disease (in terms of the abnormal image features it shares with other members of that disease category). The closer any two cases are to each other, the more similar are their abnormal features. Thus, a trainee can see directly whether any case is typical or atypical of a disease, the similarity between any pair of cases, and whether an outlier (atypical) case belongs unambiguously to one disease or if lies in the space between two or more disease regions and thus may be difficult to diagnose.

Figure 3 here

For the initial prototype system, fifteen T_2 weighted representative examples of glioma and eight examples of infarcts have been selected from the archive. The images are displayed at 256 by 256 pixel size. The overview plot for the glioma and infarct case space is shown at the lower right of Figure 3, with gliomas shown as dark points and infarcts as light points.

The trainee selects a case to examine by clicking on a point in the overview plot. The associated stack of images then appears in the pane at the top middle right of the screen. A control panel enables the trainee to move up and down the slices and to "window" an image by adjusting its grey levels. The trainee can call up similar cases, and more or less typical ones, by selecting other appropriate points on the plot. Slices can be dragged to the "gallery" below the main image pane.

The left half of the screen is the interface to the tutoring component. Either the trainee or the system can select a case for tutoring (a "target case"). This can be viewed and windowed as for the reference images, so the trainee can make visual comparisons between the target case and any reference cases. The trainee describes the target case by selecting terms from pop-up menus of feature descriptors. A context-dependent help system offers a definition of each term from the image description language, to assist the trainee to enter a full and accurate description.

As the trainee works through a description the system gives a response in two forms. When the trainee moves the mouse cursor over a menu of feature descriptors (e.g. the terms for "location" shown in Figure 3) the system "lights up" those points on the overview plot that correspond to cases in which the indicated feature is present. This shows at a glance the position of the case being tutored relative to those with a particular feature. User trials have suggested that this is a useful facility for beginning users but makes the task of describing a case too easy for more expert trainees, so future versions will provide a means to disable the feature as the teaching progresses.

The second type of feedback is provided by a written tutorial response generated by the system. The system gives a confirming or remedial message each time the trainee completes a section of the structured description. A typical response is as follows:

> You have correctly identified the following locations:
>
> > Cerebral white matter
>
> However, you have incorrectly identified the lesion as being in the following areas:
>
> > Cortical grey matter
>
> And omitted the following areas:
>
> > Periventricular
> >
> > Central grey
> >
> > Brainstem or cerebellum
>
> Now describe the exterior margin of the lesion. Indicate to what extent the lesion is distinct form normal tissue

When the description is complete, the system adds a point to the overview plot that corresponds to the trainee's description. This enables the trainee to make a visual comparison with the point that represents the expert description for the case. It shows, for example, whether the trainee's description is more or less typical than the expert description and where it lies in relation to cases for candidate diseases. The trainee can then make further attempts to describe the case until the description matches that of the expert, when the trainee's case point will appear in the same position as the expert's. At any time the trainee can give up and request the Tutor to show the expert's description of the case in the menu window.

The MR Tutor has been implemented in Java. All the facilities described in the Overview of the MR Tutor section have been implemented. Running on a Toshiba Tecra 8000 laptop computer (Pentium II 233 MHz processor, 127Mb RAM) the delay for the slowest operation (loading a new case and displaying the first image) is less than 1 second.

Structured Reporting and the Image Description Language

Although there is no standard method of describing abnormalities in MR neuroradiological images, there is widespread agreement on the need to develop a more structured approach to the description of lesions, so that radiologists can exchange findings in an agreed language using terms that have been precisely defined.

CT and MRI have provided opportunities to describe in formal terms and in great detail the positions and features of brain lesions and to calculate disease probabilities or to index image features for teaching purposes. Despite the hopes of some for a unified language to describe medical images [23], most progress has been made in description menus applied to a single organ or system, imaged by one imaging modality.

The early work by G. du Boulay on diagnostic assistance for brain diseases evolved into a collaborative research project to develop image description languages for CT and MR images of the head. This has continued over eighteen years and involves the Medical Systems Group, De Montfort University, and the Institute of Neurology, London. A study of CT brain scans led to the development of a structured description process for CT images and a menu-based computer advisor (BRAINS) to aid in image interpretation and cerebral disease diagnosis [24]. An independent evaluation of that project [25] involved 10 users of various levels of expertise, each describing up to 20 cases. It identified the training benefits of a structured approach to reporting linked to a reference set of annotated example images and diagrams.

The language for MR has been derived using an iterative prototyping approach. The initial specification of the description language was based on observation, transcription and protocol analysis of focused discussions between G. du Boulay, Emeritus Professor of Neuroradiology at the Institute of Neurology, and D. P. Kingsley, a senior neuroradiologist, during the examination of a representative sample of abnormal MR images. The language was then applied to a set of ten cases which were reviewed by both observers for accuracy and completeness. The language was subsequently further developed in consultation with Kingsley and G. du Boulay. A new sample of 66 cases was described in the terminology of the IDL by du Boulay, and rated by him for

accuracy and completeness. This led to further refinements as a result of information regarding the inadequacies of the prototype IDL in describing these cases.

The IDL describes the appearance of the images rather than the underlying disease, though the ontology of the language is influenced by a knowledge of diagnostically important disease processes. There are 183 individual values in the full description language. Figure 4 gives a list of its main category headings.

Figure 4 here

The analytical power of the IDL has been partly tested by its application to the differentiation between multiple sclerosis (MS) and vascular disease [3] and the effects of HIV infection on the brain [4]. Further insights into the predictive power of combinations of features will emerge as part of the continuing statistical analysis of the data, including the application of Multiple Correspondence Analysis [26] (the basis of the overview plot methodology described later).

The language is particularly suited to computer implementation, since it provides a canonical set of feature descriptors that can be stored as schemas for each case where the image features are represented by slots whose fillers indicate feature values (e.g. lesion margin: graded). The terms of the IDL are supported by precise definitions and, for lesion shape, by indicative examples.

To test the methodology, a simplified version of the description language has been used in the current prototype training system, based on the main category headings of the full IDL. It provides an initial set of terms to support discussion and sharing of knowledge amongst trainee neuroradiologists and their supervisors. It also serves as a structured representation of knowledge for the MR Tutor, enabling it to generate remedial responses to student errors. Figure 5 shows the terms of the simplified language and Figure 6 shows a description of a case using the simplified language.

Figure 5 here

Figure 6 here

The Overview Plot: Visualisation and Interaction with the Case Library

The acquisition of skill in radiology involves integrating knowledge gained from viewing and reporting individual cases into general schemas covering diagnostic categories, enabling judgements to be made of the likelihood of possible abnormalities. We propose that the process of building disease schemas can be aided by providing, as part of training, visual indications of the distribution of cases by disease category, spatially organised according to similarity, typicality and category membership.

To test this proposal we have developed a novel interactive overview plot (see Figure 3) that enables a trainee to view a given case in relation to other cases of the same and related diseases. The plot is based on a statistical model of the distribution of cases within categories, where each category represents one disease. The model provides operational definitions of typicality, similarity and graded category membership. It enables estimates to be computed of the probability of feature profile given disease $p(f|D)$. The model can also be used in a non-independent Bayesian procedure to give estimates of disease given a feature profile $p(D|f)$, thus providing diagnostic advice for a given case.

Each case in the archive is associated with an expert description formed from the terms of the image description language. The case description is stored as an array of features where each feature value is coded as a binary variable indicating its presence or absence in the image (e.g. "lesion margin graded: present"; "lesion margin sharp: absent"). Ordinal attributes (such as "size") are coded by dividing them into ranges of values and representing each range as a corresponding binary variable. The benefits of this approach compared to scalar representations are that nominal and ordinal features can be combined into a single feature list, without the necessity to create contrived and incommensurable scales.

The description language defines a multi-dimensional space, where each dimension represents one feature value with two states, present or absent. We can consider a case as representing a single point in the space and a category as occupying a region. For the simplified language used in the prototype MR Tutor, the space has 30 dimensions.

To allow a trainee to visualise the distribution of cases, it is necessary to simplify the presentation of these multiple dimensions whilst retaining the most useful information. The statistical technique of multiple correspondence analysis (MCA) [26] provides such a visualisation.

Simple two dimensional scatterplots are a well established statistical technique for viewing bivariate data. The benefit of the MCA is that it can project a multi-dimensional space onto a two-dimensional surface, such that the data points are maximally spread over the surface. Computed values of typicality and similarity can be presented as direct perceptual relationships. A case that lies near the centroid of a category can be interpreted as being highly typical of that category, whereas a case nearer to the periphery is less typical. The similarity of any two cases within a particular category is indicated by the spatial distance between their two points. A cluster of points indicates a group of cases of similar appearance.

The plot is scaled such that the same perceptual distance between cases represents an increasing degree of similarity as one moves out from the centre of a category. This matches the psychological finding that people can make finer similarity judgements for more typically encountered cases [27, 28]. To assist judgement, we have augmented the plot by showing "contours of typicality" for disease categories, based on a bivariate normal model of the distribution of cases for a particular disease in the overview plot [29].

As a means of visualising and interacting with libraries of cases, the overview plot offers a number of benefits:

− It is a representation across all the variables rather than a small selection of terms.

− It can be applied to a mixture of ordinal and categorical features.

− The data points are maximally spread out over the display surface. The degree of spread does not depend merely on the position of outliers. These can influence the spread, but the method summarises the main source of variation of all the variables. The plot allows outliers to be identified, in a more sophisticated manner than with univariate plots.

− It gives a direct perceptual overview of the typicality, similarity, spread and clustering of cases, and overlap of categories.

— An image description can, in principle, be generated for any position in the plot. If the position does not correspond to one of the stored cases, then the description can be computed by interpolation from nearby cases. The mapping from a position in the plot to a description is one-to-many and heuristic methods would be required to select the most plausible description.

— The plots can be pre-computed automatically from the feature arrays.

— For a large sample, new cases can be added without the need to recompute the plot (the distribution of a large sample is not markedly affected by the addition of a small number of extra cases).

The simplified version of the image description language was used to generate the initial overview plots for the Tutor, but the method can be scaled to the full IDL. The plot for a single disease category, glioma, is shown in Figure 7. Annotations have been added to the figure to indicate the distribution and clustering of cases. As the figure shows, the MCA has separated out a cluster of cases with lesions that exhibit a focal structure. The vertical dimension has further divided the cases, primarily according to size of lesion.

Figure 7 here

Cases from further diseases could be added to the plot until it displays the entire archive. However, there is no diagnostic value in plotting within a single space of dissimilar diseases. Instead, we have adopted the "small worlds" metaphor [30]. We divide the archive into small sets of confusable diseases and compute separate overview plots for each of these disease sets.

Figure 8 shows an example of an overview plot for combined glioma and infarct cases. The display indicates that there is a cluster of cases in the centre of the plot that cannot easily be distinguished as gliomas or infarcts. It also shows a number of outlying cases that, although they are atypical, should present no difficulty in a differential diagnosis of glioma or infarct (they may, however, be confusable with other diseases). It should be noted that the two axes of the plot summarise the main source of variability in the data. Adding a third orthogonal axis might provide some improved separation of the diseases; this could be investigated for pairs of diseases using statistical methods and, if appropriate, presented as an additional plot of the 2nd and 3rd principal dimensions.

Figure 8 here

The display provides an easy method of accessing cases from the archive. A trainee can call up the images and structured description for a case by clicking on its point in the overview plot. The display can also enable a trainee to explore the relationship of a case description to cases in the archive or to disease categories. For example, a trainee can investigate and eliminate differential diagnoses by describing a target case and viewing its position on the overview plot for candidate diseases.

Evaluation

A heuristic evaluation [31, 32] of the MR Tutor was undertaken to assess the overall usability, intuitiveness and efficacy of the system and its interface. The evaluation involved three expert neuroradiologists from The Institute of Neurology, London and

four Human Computer Interaction (HCI) experts from the School of Cognitive and Computing Sciences, University of Sussex.

Each subject was given a 20 minute tutorial on the interface actions to demonstrate the functionality of the MR Tutor. The subjects were then asked to undertake a set of three tasks for which the system was designed (for example to retrieve a typical example of a given disease from the archive and describe the appearance of the lesion). The subjects were asked to comment on any problems they encountered with regard to eight heuristics, based on those proposed by Nielson and Molich [31]. The heuristics were: visibility of system status; match between the system and the real world; user control and freedom; consistency and standards; intuitiveness; speed and ease of use; aesthetic and uncluttered design; help and documentation.

The subjects also filled in a questionnaire to ascertain their overall impression of the system. The questionnaire provided evidence relating to overall usability and intuitiveness of the user interface of the MR Tutor. The HCI subjects recorded a mean score of 3.96 for overall level of satisfaction (on a 5 point scale) and the radiologists gave a mean score of 4.13. The HCI subjects indicated that the system conforms to good practice in interface design and were pleased with the overall look and feel of the interface. All subjects described the screen layout as good and aesthetically pleasing

Although the interface and system were in general well received, the subjects offered useful suggestions for improvement. For example, radiologists suggested providing additional imaging sequences to give context to the target case, and also providing a slightly larger image. The radiologists indicated that the image description language is adequate for initial tutoring, but suggested that some of the terminology used to describe the lesion appearance needed to be further validated, especially with regard to lesion size and shape.

The HCI subjects suggested that, although the menu area was clear, well structured and well laid out, there appears to be a basic flaw in the action of submitting an answer in the menu area by pressing the next major heading button. All the HCI subjects found this confusing, although it was surprising that none of the radiologist subjects commented on this problem and generally liked the menu area. The overview plot was found to be an intuitive and useful way of retrieving cases from the archive.

An initial evaluation of the overview plot [33] involved the expert providing subjective ratings (on a scale from 0 to 100) of similarity between the images for a sample of 8 cases taken from the Glioma disease space. The expert's subjective assessments of similarity were in good agreement the scores based on the Euclidean distances between the sample cases on the overview plot, with Spearman's Rank Correlation (rho) of 0.754 demonstrating a highly significant correlation $p<.001$. A Multidimensional Scaling Analysis (MDS) of the expert's subjective scores produced a two dimensional scatter plot which closely resembled the overview plot from the MCA of the image descriptions.

A second set of experiments involved seventeen subjects, comprising four novices (with no knowledge of radiology), nine intermediates (4th year medics and radiographers with some knowledge of anatomy and imaging) and four expert neuroradiologists. The subjects were presented with the overview plot (an unannotated version of Figure 6) for a single disease, glioma, on a computer screen with six of the cases removed. They were asked to explore all the presented cases by clicking on the points to bring up case images and associated descriptions. They were then shown the images and descriptions of the remaining six cases one at a time and asked to

place a marker representing each case at an appropriate position in the overview plot. The Euclidian distance between each individual's placement of a case and the corresponding statistical placement in the overview plot were computed. These distances were subject to an Analysis of Variance following log transformation to achieve the normality assumption required for statistical testing. The ANOVA provided evidence of a significant difference between the placement of cases by the three groups (expert, intermediates and novices) ($F_{2,60}$ = 3.150, $p<.05$), the expert's placement being statistically significantly closer to the statistical placement than that of the intermediates and novices.

Interviews with the subjects based on a structured questionnaire indicated that they found the overview plot easy to use and acceptable as a means for retrieving cases from the image archive. The evaluation suggests that the overview plot can provide a useful teaching device to assist a trainee in acquiring an understanding of the distribution of cases for a disease comparable to that of an expert.

Work in Progress and Future Work

The MR Tutor forms part of MEDIATE, a programme of research and development to produce a computer workbench for MR imaging. MEDIATE includes work in progress to extend the current MR Tutor, undertake a more detailed task analysis of radiological reporting, carry out an analysis of radiologists' comprehension and mental visualisation of the overview plot for multiple diseases, and explore methods of accepting and expressing levels of uncertainty for tutorial interactions.

A number of enhancements to the MR Tutor have been proposed in outline and are awaiting implementation. They will develop the teaching strategy in line with principles for concept tutoring and extend the range of tutoring tactics to match the teaching methods identified by Azevedo and Lajoie [18] such as articulating the reasoning process and "scaffolding" the interpretation to suit the expertise of the trainee. The highest priority is to create and validate a database containing all the cases and their associated descriptions from the archive, organised into small worlds of confusable diseases.

The teaching module needs to be developed so that it can tutor about the *process* of image description and reporting as well as the outcome. For example, it should be able to comment on a trainee's interactions with the archive, the particular cases chosen for examination, when a trainee has not indicated any initial hypotheses, not scanned through all of the slices, not called up and examined the slice best illustrating the lesion, or not made good use of the archive.

Another major line of development is to support trainees of differing levels of ability, by providing a staged approach to learning and gradual removal of the support given by the system. The trainee first learns the terminology of the IDL through practice in describing cases from a single disease, with the system responding to incorrect decisions and offering definitions for the terms of the IDL. The trainee then moves on to the second stage of viewing cases from sets of commonly confused diseases with their associated expert case descriptions, to gain an understanding of the overlap between diseases. At the third stage (not yet implemented) the trainee will be able to enter new cases and the system will provide tools to describe the cases and guidance to arrive at a diagnostic decision. The system has been designed so that with all scaffolding and tutoring removed, it could function as a decision-support system for an

expert radiologist, by displaying the relationship of a newly-described case to other cases and diseases in the archive and by providing advice on the conditional probability of differential diagnoses.

The current system implements a simplified version of the image description language. A further area of research will address the issue of whether and how the simplified language should gradually be extended for the more expert user. The aim is for a radiologist to be able to call up a new case direct from an MRI scanner and enter a report in the structured language by selecting appropriate descriptors from the menu of terms. Ultimately, the entire library of 1200 cases will be available, indexed by appearance and accessible by search for given terms or through the overview plot, along with reference material such as journal articles. Each case will be accompanied by a description using the full set of terms from the image description language.

Other developments being planned include integration of the system with a knowledge-based atlas of the brain [34] and providing communications to allow the tutor to be accessed over a wide-area network along with video conferencing facilities, to support tutoring at a distance by a combination of computer and human expert.

Conclusion

This paper describes a training and decision support system for MR neuroradiology. The MR Tutor differs from conventional computer-based training systems in providing symmetric access by the system and the trainee to a structured representation of diagnostically relevant image features associated with an archive of representative cases. The system implements a novel overview plot that gives an interactive visual display of the typicality, similarity, and disease membership of cases.

The teaching approach of structured reporting has a role in the reform of medical education but has been difficult to provide through current teaching methods. The MR Tutor provides a self-study approach to acquiring a systematic approach to radiological diagnosis.

The overview plot and the structured approach to training and reporting have been well received by teachers of radiology. Further development and use of the system with trainee radiologists will show whether computer-assisted training in radiology can supplement conventional methods, by providing exposure to an archive of images indexed by lesion position, appearance and diagnosis, and by teaching a systematic approach to image description and interpretation.

Acknowledgements

References

[1] H. P. A. Boshuizen and H. G. Schmidt, "On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices," *Cognitive Science*, vol. 16, pp. 153-184, 1993.

[2] D. J. Getty, R. M. Pickett, C. J. D'Orsi, and J. A. Swets, "Enhanced Interpretation of Diagnostic Images," *Investigative Radiology*, vol. 23, pp. 240-252, 1988.

[3] G. H. du Boulay, B. A. Teather, D. Teather, N. P. Jeffery, M. A. Higgott, and D. Plummer, "Discriminating multiple sclerosis from other diseases of similar presentation – can a formal description approach help?," *Rivista di Neuroradiologia*, pp. 37 – 45, 1994.

[4] G. H. du Boulay, B. A. Teather, D. Teather, C. Santosh, and J. Best, "Structured reporting of MRI of the head in HIV," *Neuroradiology*, vol. 37, pp. 144, 1994.

[5] K. W. McEnery, "Interactive Instruction in the radiographic anatomy of the chest," *Comput. Methods Programs Biomed.*, pp. 81 – 86, 1986.

[6] J. A. Mitchell, A. S. C. Lee, T. Tenbrink, J. H. Cutts, D. P. Clark, S. Hazelwood, R. Jackson, J. Bickel, W. Gaunt, R. P. Ladenson, and G. C. Sharp, "AI/Learn: An interactive videodisk system for teaching Medical Concepts and Reasoning," *JMS*, pp. 421 – 429, 1987.

[7] C. E. Kahn, Jr., "Validation, clinical trial, and evaluation of a radiology expert system," *Methods of Information in Medicine*, pp. 268 – 274, 1991.

[8] D. B. Hayt, R. James, R. Knowles, and S. M. Erde, "A high-resolution networked computer system for radiological instruction of medical students," *Journal of Digital Imaging*, vol. 4, pp. 202 – 206, 1991.

[9] E. K. Fishman, J. G. Hennessey, and M. S. Nixon, "Computer-based radiology teaching programs: The challenge and trauma of development and implementation," *Proceedings of the Seminar of Ultrasound, CT and MR*, vol. 13, pp. 113 – 121, 1992.

[10] R. T. Macura, K. J. Macura, V. E. Toro, E. F. Binet, J. H. Trueblood, and K. Ji, "Computerized case-based instructional system for computed tomography and magnetic resonance imaging of brain tumors," *Investigative Radiology*, vol. 29, pp. 497 – 506, 1994.

[11] M. Sharples, "The Radiology Tutor: Computer-based teaching of visual categorisation," *Proceedings on the 4th International Conference on Artificial Intelligence and Education*, Amsterdam, 1989, pp. 252-259.

[12] A. I. Direne, "Methodology and tools for designing concept tutoring systems," *Proceedings of the World Conference on Artificial Intelligence in Education*, Edinburgh, 1993, pp. 58-65.

[13] R. Azevedo, S. Lajoie, M. Desaulniers, D. Fleiszer, and P. Bret, "RadTutor: The theoretical and empirical basis for the design of a mammography interpretation tutor," in *Artificial Intelligence in Education: Knowledge and Media in Learning*

Systems, B. du Boulay and R. Mizoguchi, Eds. Amsterdam: IOS Press, 1997, pp. 386-393.

[14] M. Sharples, N. Jeffery, J. B. H. du Boulay, D. Teather, B. Teather, and G. H. du Boulay, "Socio-Cognitive Engineering: A Methodology for the Design of Human-Centred Computer Systems," *Proceedings of HCP '99, Conference on Human Centred Processes*, Brest, France, 1999, pp. 67-73.

[15] M. Sharples, N. Jeffery, D. Teather, B. Teather, and G. du Boulay, "A Socio-Cognitive Engineering Approach to the Development of a Knowledge-based Training System for Neuroradiology," *Proceedings of the World Conference on Artificial Intelligence in Education*, Kobe, Japan, 1997, pp. 402-409.

[16] M. Myles-Worsley, W. A. Johnston, and M. Simons, "The influence of expertise on X-ray image processing," *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol. 14, pp. 553 – 557, 1988.

[17] A. Lesgold, H. Rubinson, P. Feltovich, R. Glaser, D. Klopfer, and Y. Wang, "Expertise in a complex skill: Diagnosing X-ray pictures," in *The Nature of Expertise*, M. Chi, R. Glaser, and M. J. Farr, Eds. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1988, pp. 453 – 494.

[18] R. Azevedo and S. P. Lajoie, "Learning styles underlying radiological expertise," McGill University, Montreal, Project Safari Report 005, 6 October 1995.

[19] J. S. Brown, R. R. Burton, and A. G. Bell, "SOPHIE: a step toward creating a reactive learning environment," *International Journal of Man-Machine Studies*, vol. 7, pp. 675 – 696, 1975.

[20] A. Lesgold, S. Lajoie, M. Bunzo, and G. Eggan, "SHERLOCK: A coached practice environment for an electronics troubleshooting job," in *Computer Assisted Instruction and Intelligent Tutoring Systems: Shared Goals and Complementary Approaches*, J. H. Larkin and R. W. Chabay, Eds. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1992, pp. 201 – 238.

[21] P. Fink and J. Lusth, "Expert systems and diagnostic expertise in the mechanical and electrical domains," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 17, pp. 340 – 349, 1987.

[22] A. Collins, J. S. Brown, and S. E. Newman, "Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics," in *Knowing, Learning and Instruction: Essays in Honor of Robert Glaser*, L. Resnick, Ed. Hillsdale: N.J.: Lawrence Erlbaum Associates, 1989, pp. 453 – 494.

[23] D. A. B. Lindberg, B. L. Humphreys, and A. T. McCray, "The Unified Medical Language System," *Year Book of Medical Informatics*, pp. 41-51, 1993.

[24] D. Teather, B. A. Morton, G. H. du Boulay, K. M. Wills, D. Plummer, and P. R. Innocent, "Computer assistance for C.T. scan interpretation and cerebral disease diagnosis," *Statistics in Medicine*, pp. 311 – 315, 1985.

[25] D. Teather, B. A. Teather, K. M. Wills, G. H. du Boulay, D. Plummer, I. Isherwood, and A. Gholkar, "Evaluation of computer advisor in the interpretation of CT images of the head," *Neuroradiology*, pp. 511 – 517, 1988.

[26] M. J. Greenacre, *Correspondence Analysis in Practice*. London: Academic Press, 1993.

[27] C. L. Krumhansl, "Concerning the applicability of geometric models to similarity data: the interrelationship between similarity and spatial density," *Psychological Review*, vol. 85, 1978.

[28] R. M. Nosofsky, "Overall similarity and the identification of separable-dimension stimuli: a choice model analysis," *Perception and Psychophysics*, vol. 38, pp. 415-432, 1985.

[29] D. Teather, B. A. Teather, N. P. Jeffery, G. H. d. Boulay, B. d. Boulay, and M. Sharples, "Statistical Based Support for Uncertainty in Radiological Diagnosis," *Methods of Information in Medicine*, vol. 39, pp. 1-6, 2000.

[30] A. W. Kushniruk, V. L. Patel, and A. J. Marley, "Small worlds and medical expertise: implications for medical cognition and knowledge engineering," *International Journal of Medical Informatics*, vol. 49, pp. 255-271, 1998.

[31] J. Nielsen and R. Molich, "Heuristic evaluation of user interfaces," Proceedings of the ACM CHI'90 Conference, Seattle, WA, 1990, pp. 249-256.

[32] J. Nielsen, *Usability Inspection Methods*. New York: John Wiley & Sons, 1994.

[33] N. P. Jeffery, "Computer Assisted Tutoring in Radiology," in *Department of Medical Statistics*. Leicester: De Montfort University, 1997.

[34] S. Garlatti and M. Sharples, "The Use of a Computerized Brain Atlas to Support Knowledge-based Training in Radiology," *Artificial Intelligence in Medicine*, vol. 13, pp. 181-205, 1998.
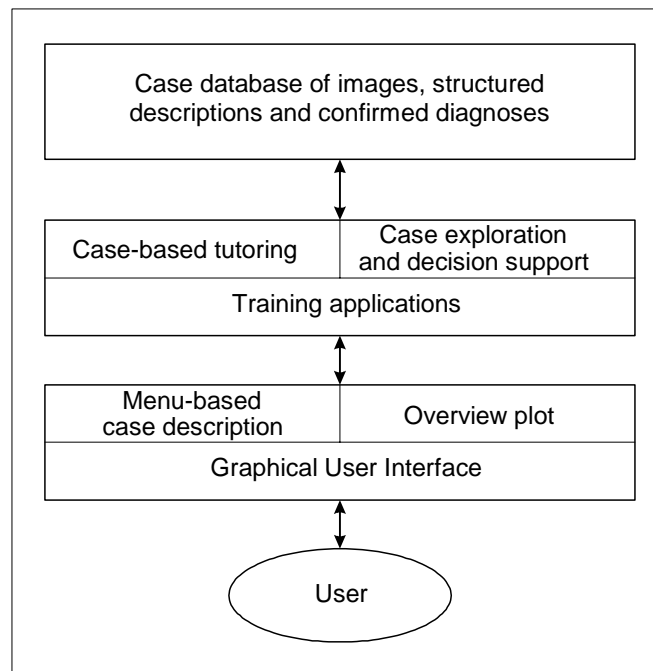
Figure 1. The System Architecture of the MR Tutor

```
┌─────────────────────┐      ┌─────────────────────┐      ┌─────────────────────┐
│ 1.                  │      │ 2.                  │      │ 3.                  │
│ Explore case archive│─────▶│   View tutorial case│─────▶│    Describe case    │
└─────────────────────┘      └─────────────────────┘      └─────────────────────┘
          │                            │                            │
          │                            │                            │
┌─────────────────────┐               │                            │
│ 1.1        ┌────────────────┐        │                            │
│ Select disease(s)│ 1.2      │        │                            │
│ to view    │ Examine and    │        │                            │
│            │ compare cases  │        │                            │
└────────────┴────────────────┘        │                            │
                                       │                            │
          ┌────────────────────────────────────────────────────┐   │
          │ 2.1          ┌──────────────┐  ┌──────────────────┐ │   │
          │ Select case  │ 2.2          │  │ 2.3 Select and view│ │   │
          │          ─────▶ Window and  ─▶│ related cases and │ │   │
          │              │ view slices  │  │ descriptors from archive│ │
          └──────────────┴──────────────┴──────────────────────┘   │
                                                                    │
          ┌────────────────────────────────────────────────────────┐
          │ 3.1 Enter description  ┌──────────────┐ ┌─────────────┐ │
          │ using menu of      ────▶ 3.2 Call up   ─▶ 3.3 View position on│
          │ IDL terms          │ expert         │  │ plot of expert and│
          │                    │ description    │  │ user descriptions │
          └────────────────────┴────────────────┴──────────────────┘
                    │
          ┌──────────────────────────────────────┐
          │ 3.1.1        ┌──────────────┐         │
          │ Describe feature │ 3.1.2      │       │
          │              ───▶ Receive tutorial│   │
          │              │ feedback       │       │
          └──────────────┴────────────────┘
```
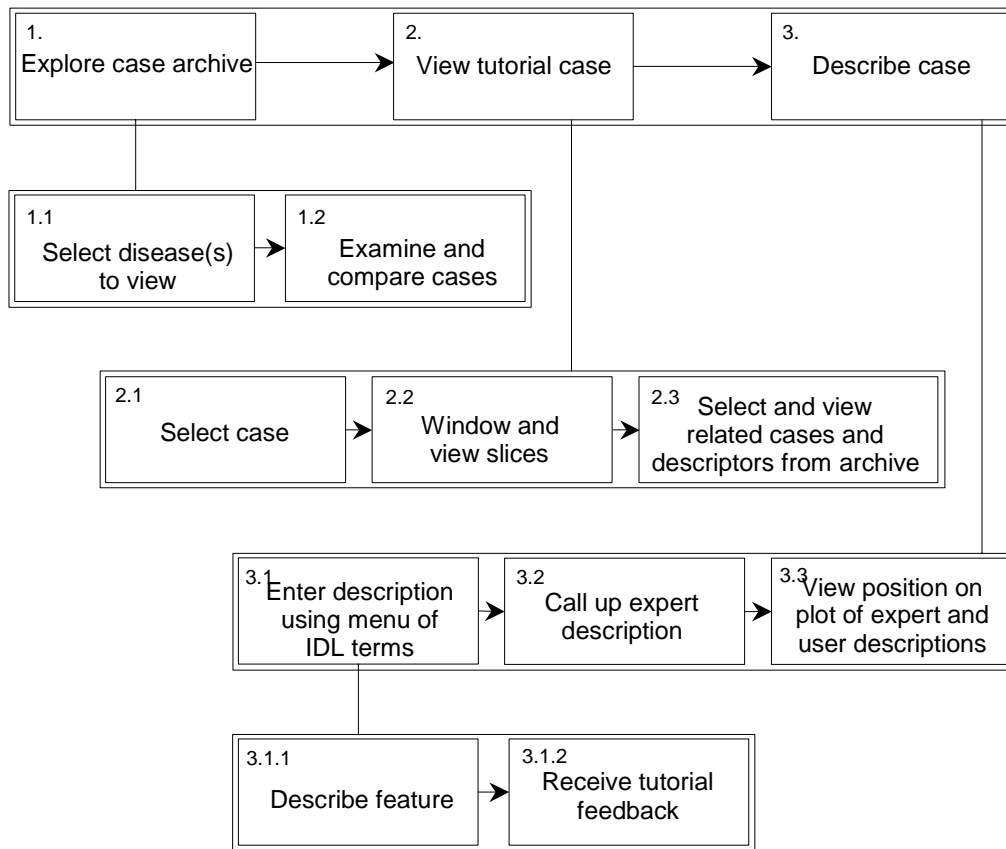
Figure 2. Main task interactions with the MR Tutor
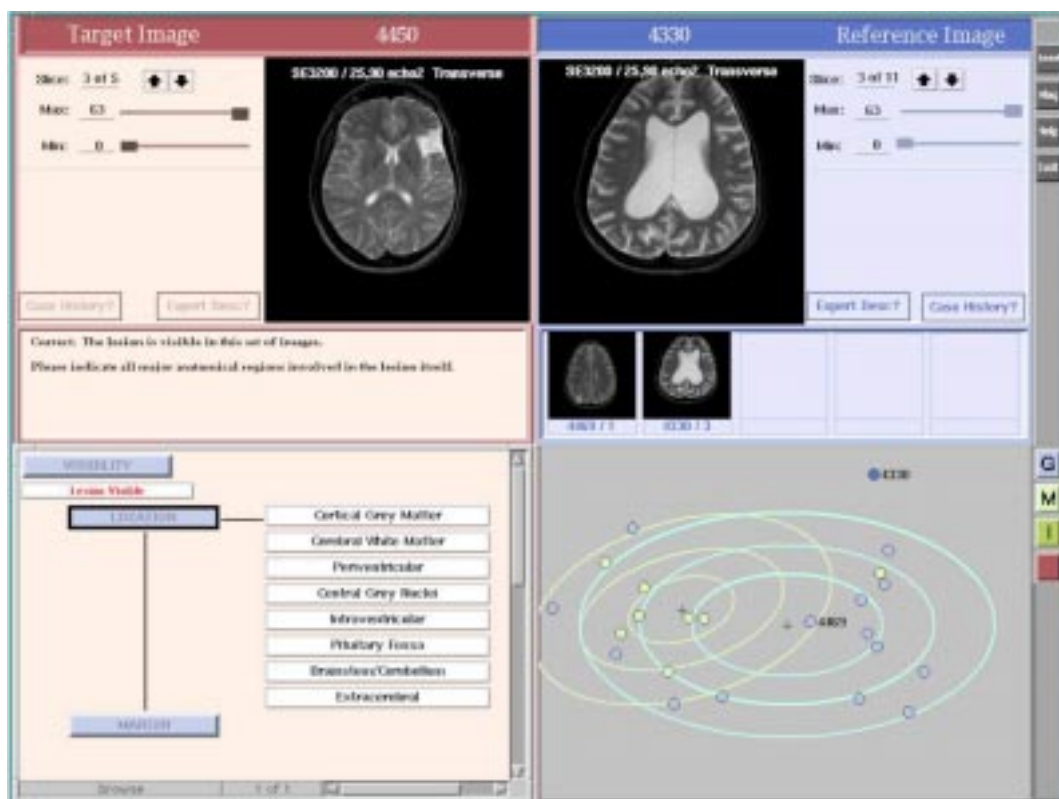
Figure 3. The main screen display of the MR Tutor

Presence of areas of abnormal signals
Single, Multiple, No discernible

If multiple specify
Lesion types : Characterise groups of lesions on the basis of appearance;
                  Specify Number, Size range, Dispersion (in terms of
                  the major anatomical locations indicated below)
                  of each type.

<u>Description of single lesion or largest lesion of each type</u> :

Major anatomical locations
Cortical grey matter; Cerebral white matter; Periventricular; Corpus callosum;
Central grey nuclei; Intraventricular; Pituitary fossa; Brain stem/Cerebellum; Basal cisterns;
Cortical subarachnoid space; Extracerebral; Skull/Neck; Orbits

Exact location
Exact location is specified from a detailed list specific to each of the major anatomical locations

Lesion shapes
Rounded; Irregular; Irregularly rounded; Oval; Linear; Mantle; Lentiform;
Conforming to anatomical structure(specify)

Geometric qualifiers
Margin of whole lesion : Sharp; Graded
Area of whole lesion : cm^2; Number of contiguous slices in which lesion is visible;
Proportion of structure occupied if conforming to anatomical structure

Overall appearance
Homogeneous; Containing focal structure; Lesion comprising two/three parts; Heterogeneous

Interior pattern of Lesion/Focal structure/Part
Homogeneous; Multiple repetitive; Concentric layers; Heterogeneous

Scales of intensity
Specify Reference Scale as: Cerebral or Cerebellar white
Specify intensities of lesion/focal structure if appropriate/parts if appropriate
using scale : +++, ++, +, iso, -, --, ---

Other signs include
Mass effect; Ventricular dilatation (specify); Sulcal compression;
Sulcal widening; Loss of brain substance; Other abnormal signals (specify)

Correspondence
Relates information provided by different sequences to identified parts/structures of lesions


Disease index
WHO list

Figure 4. Summary of Image Description Language

Abnormal Signals
Single/Multiple areas
Lesion visible/invisible
Cortical grey
Cerebral white
Periventricular
Central grey
Intraventricular
Pituitary fossa
Brainstem or cerebellum
Extracerebral
Graded/Sharp
Rounded
Irregular
Conforms to anatomical feature
Homogeneous
Unstructured heterogeneous
Containing a focal structure
Very small
Small
Medium
Large
Hyperintense
Hypointense
Isointense
Mass Effect
Expansion of cortical sulci
Dilation of Lat. Ventricles
Areas of diffuse signal
Localised loss of brain substance

*Figure 5. The terms of the simplified Image Description Language*

Visibility
> Lesion visible

Location
> Cortical grey matter
> Cerebral white matter
> Periventricular

Margin
> Graded

Shape
> Irregular

Interior pattern
> Unstructured heterogeneous

Size
> Large

Intensity
> Hyperintense

Other signs
> Mass effect

Figure 6. A description of a case using the simplified Image Description Language in the prototype Tutor
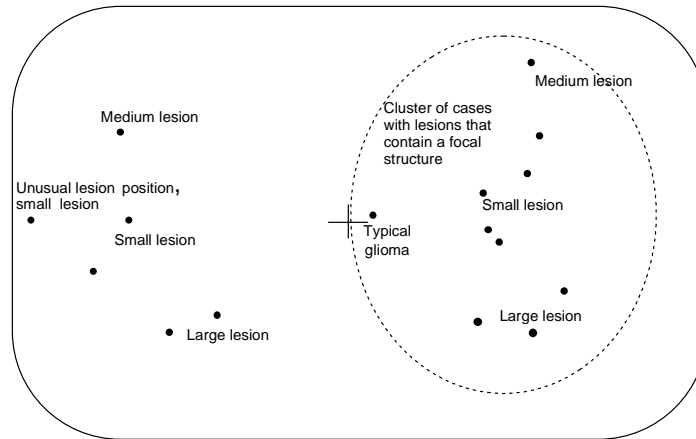
Figure 7. Annotated overview plot for gliomas
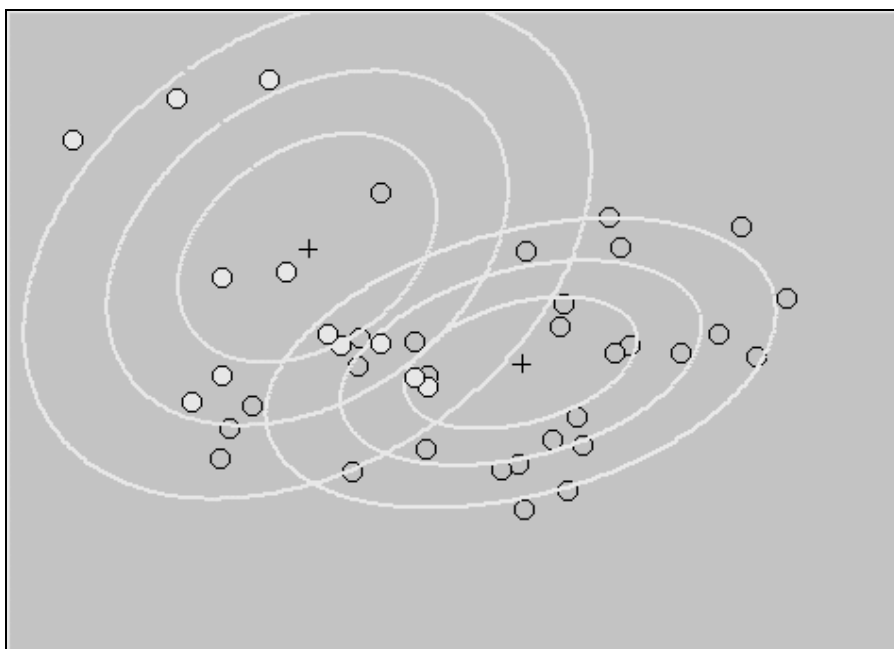
Figure 8. Overview plot for glioma (dark points) and infarct cases (light points).