8-1998

# Prototype internet consultation system for radiologists

Boris Kovalerchuk

James F. Ruiz

Evgenii Vityaev

Steven Fisher

# Prototype Internet consultation system for radiologists

**Boris Kovalerchuk,**
Department of Computer Science, Central Washington University, Ellensburg, WA 98926-7520, USA
borisk@tahoma.cwu.edu
**James F. Ruiz**
Department of Radiology, Woman's Hospital, Baton Rouge, LA 70895-9009, USA
MDJR@womans.com
**Evgenii Vityaev,**
Institute of Mathematics, Russian Academy of Science, Novosibirsk, 630090, Russia
vityaev@math.nsc.ru
**Steven Fisher**
Department of Computer Science, Central Washington University, Ellensburg, WA 98926-7520, USA
fishers@tahoma.cwu.edu

## ABSTRACT

The overall purpose of this study is to develop a prototype radiological consultation system. We concentrate our work on **prototype software environment for the system.** The system provides a second diagnostic opinion based on similar cases, incorporating the experience of radiologists, their diagnostic rules and a database of previous cases. The system allows a radiologist to enter the description of a particular case using the lexicon such as BI-RADS of American College of Radiology and retrieve the ***second diagnostic opinion*** (probable diagnosis) for a given case. The system also allows a radiologist to get other important information too. These advances are based on a new computational intelligence technique and first-order logic [Mitchell, 1997].

We implemented a rule-based prototype diagnostic system. Two experimental Internet versions are currently available on the web and are under testing and evaluation of design. The diagnosis is based on the opinions of radiologists in combination with the statistically significant diagnostic rules extracted from the

available database.

## 1. INTRODUCTION

We present current implementation and general design of the system. Currently radiologists can retrieve the second diagnostic opinion (probable diagnosis) for a given case using two sets of features:

BI-RADS of the American College of Radiology and

Set 2. This set includes: (1) the number of calcifications/$cm^2$ ; 2) the volume (in $cm^3$) ; 3) total number of calcifications ; 4) irregularity in shape of individual calcifications; 5)variation in shape of calcifications ; 6) variation in size of calcifications; 7) variation in density of calcifications; 8) density of calcifications; 9) ductal orientation; 10) comparison with previous exam; 11) associated findings . See [Kovalerchuk, Triantaphyllou, Ruiz, 1996] for more details.

The radiologist can retrieve:

(1) description of similar cases in terms of BI-RADS and Set 2, including clinical and pathological data for these cases.

(2) digital reproduction of the mammograms for these similar cases (currently about 100 images in JPEG format)

(3) diagnostic rules used by the system.

The breast cancer database is developed as a part of our system and is open to incorporate other databases).

Our main purpose in this project is to extend the traditional design of ***medical expert systems*** to make the system closer to a real ***medical consultation system***, where several radiologists discuss their vision of the case and diagnosis. The traditional design of medical expert systems is focused on delivering a second diagnostic opinion by the system automatically. The Internet technology makes the idea of interactive medical consultation system realizable. In this project we are creating a software system as a medium allowing radiologists to ***interact*** via the web. A radiologist can send to the system a description of a case and transmit X-ray images of this case. The radiologist removes any real patient identification before sending. The radiologist creates an artificial ID number or the system assigns this unique ID number to the case. The consultation system already has a technical capability to deliver simulated opinion of another radiologist presented through his/her diagnostic rules. Currently the system is loaded with the rules presented in [Kovalerchuk, Tantaphyllou, Ruiz, J. 1996]. This experimental part of the system is available on the web (http://www.cwu.edu/~borisk/diagtool/diagtool.html) and is under testing by Dr. Ruiz. Students M. Klatt and M. Kovalerchuk were involved in programming using Java. We developed a technology (Kovalerchuk et al, 1996 a, b) which allows us to "extract" experts diagnostic rules relatively quickly (see section results). Currently we are in the process of extending the rule base to include rules of other radiologists.

The next option within the consultation systems that is already technically realized and available on the web is delivering the second diagnostic opinion based on the rules extracted using SIPINA software system (http://eric.univ-lyon2.fr/~ricco/sipina.html). This part is based on BI-RADS lexicon. The current rule base was extracted using relatively small amount of cases (less than 100). Software system is designed in such way that it can be corrected at any moment when larger database or/and more sophisticated methods of rule extraction will be available.

The dialog pages of the system for BI-RADS and set 2 are presented in appendix 1.

The system has an option to show a radiologist the similar cases for each set of rules. This way we support a dialog similar to a real consultation where radiologists can argue each using previous cases with proven pathological analysis results.

The next innovative part of the Consultation system is a comparison tool. With this tool radiologist compares rules extracted from data with his/her rules and with rules of other radiologists or rules extracted from database. We think that this is a more realistic approach than traditional approach of automatic generation of the second diagnostic opinion. It is very questionable that a radiologist can rely on the second diagnostic opinion without extra justification.

We present extra justification for diagnostic rules using the comparison tool and provide statistical information about reliability of the rules. See section 2.

## 2. METHOD

There are many promising Computer-Aided Diagnostic (CAD) approaches ([F. Shtern, 1996], [SCAR, 1996], [TIWDM, 1996] and [CAR, 1996].) based on neural networks, nearest neighbor methods, discriminant analysis, cluster analysis, linear programming and genetic algorithm. We concentrate our approach on a less actively used first-order logic approach [Mitchell, 1997].

A radiologist enters into the CONSULTATION SYSTEM the description of a particular case. This can be done in terms of the BI-RADS lexicon of American College of Radiology. Using a CAD system, the radiologist retrieves on the screen the <u>second diagnostic opinion</u> (probable diagnosis) for a given case.

Diagnostic opinions of other radiologists about similar cases are stored in the database and will be available on the Internet.

Different techniques can be used to find similar cases in CAD system. We use a rule-based approach. A diagnostic rule applicable to the study is identified in the knowledge base. Then all cases in the database for which the premise of the rule is true are retrieved. For example for a rule: "IF x and y then z", all cases with x and y are displayed for the radiologist.

An advanced mode is designed to allow the radiologist to analyze diagnostic rules from a Computer-Aided Diagnostic system. These rules are **applicable** for a given case, and used by the system to deliver a probable diagnosis. These **rules are understandable** by any radiologist without sophisticated knowledge of the mathematical methods inside of the CAD system.

This analysis can become the most effective way for radiologists to share their experience to improve the reliability of interpretation. These rules are stored in the rule base. The method to find these rules in the rule base is based on comparison of premises of rules with a case entered by a radiologist into the system. Also a radiologist is able to enter his/her diagnosis for a studied case and to obtain a comparison of his/her diagnosis with computer simulated opinions of other radiologists for the same case. **(**Consultation system does not have actual opinions of other radiologists for this case, but using their rules, the system can simulate the opinion, i.e. the consultation system will apply diagnostic rules of other radiologists for this case description.

The user also has access to:

-the rationale of diagnostic rules by these experienced radiologists;

-the significance of rules from statistical perspective.

-a comparison of radiologist's diagnostic opinion with data-based diagnosis.

There are no serious obstacles to achieving these goals. The technique for obtaining the rationale behind the rules used by radiologists has been developed. Also we use the Consultation System to broaden the base of experience in this part of the database.

Determining **statistical significance** is difficult for many CAD systems [Kovalerchuk *et al*, 1996]. The most popular methods are based on Neural Networks, but do not have a mechanism to evaluate statistical significance of diagnosis. Therefore the reliability of diagnosis is based only on the performance on training and testing data [Gurney, 1994]. These populations may or may not be sufficiently representative for the entire population [Kovalerchuk et al, 1996]. We use an original method [Vityaev, Moskvitin, 1993], which allows the derivation of diagnostic rules and evaluates statistical significance of these rules. Examples of rules extracted from 156 cases (77 malignant and 79 benign) are given below. A radiologist may check his/her diagnostic opinion by comparing this opinion with the diagnosis made with rules derived from the data base of the Consultation System .The system diagnosis is inferred from rules discovered in the data base by "data mining" software [Vityaev, Moskvitin, 1993]. The breast cancer database is developed as a part of consultation system and is open to incorporate other databases.

The interview of a radiologist to extract rules is realized using an original method [Kovalerchuk et al, 1996 a,b] and the comparison of rules is performed by translating the rules into monotone Boolean functions and then comparing these functions [Kovalerchuk et el, 1996 a,b]. The demonstration of cases is the most important and is realized by comparing the simulated diagnosis of a given radiologist and pathologically confirmed diagnosis in the database.

## 3. RESULTS

Diagnostic Rule Acquisition. Examples of diagnostic rules extracted in a pilot study are presented below. Expert diagnostic rules were extracted from specially organized interviews of a radiologist (J. Ruiz, MD). For details of the method see [Kovalerchuk, et al, 1996 a,b]. This method is based on the theory of Monotone Boolean Functions and hierarchical approach. One of the extracted rules is presented below. RULE 1: IF NUMber of calcifications per $cm^2$ ($w_1$) is large AND TOTal number of calcifications ($w_3$) is large AND irregularity in SHAPE of individual calcifications is marked ($y_1$) THEN highly suspicious for malignancy. The mathematical expression for this rule is $w_1w_3y_1$=>"highly suspicious for malignancy".

In this study of calcifications found on mammograms we used the following features: 1) the number of calcifications/$cm^2$; 2) the volume (in $cm^3$); 3) total number of calcifications; 4) irregularity in shape of individual calcifications; 5)variation in shape of calcifications ; 6) variation in size of calcifications; 7)variation in density of calcifications ; 8) density of calcifications; 9) ductal orientation; 10) comparison with previous exam ; 11) associated findings .

To restore all diagnostic rules thousands of questions might be needed if questions are not specially organized. For 11 diagnostic features of clustered calcifications there are ($2^{11}$=2,048) feature combinations, representing cases. The questioning procedure required only about 40 questions, i.e. 50 times fewer questions than the full set of feature combinations [Kovalerchuk et al, 1996 a,b]. Note that practically all studies in CAD systems derive diagnostic rules using significantly less than 1,000 cases [Gurney, 1994]. This is the first attempt to work with such a large number of cases (2,000).

Diagnostic Rules extracted from Database. A study used 156 cases (77 malignant, 79 benign). Cases were described with 11 features of clustered calcification listed above and with two extra features: Le Gal type and density of parenchyma. The diagnostic classes were "malignant" and "benign".

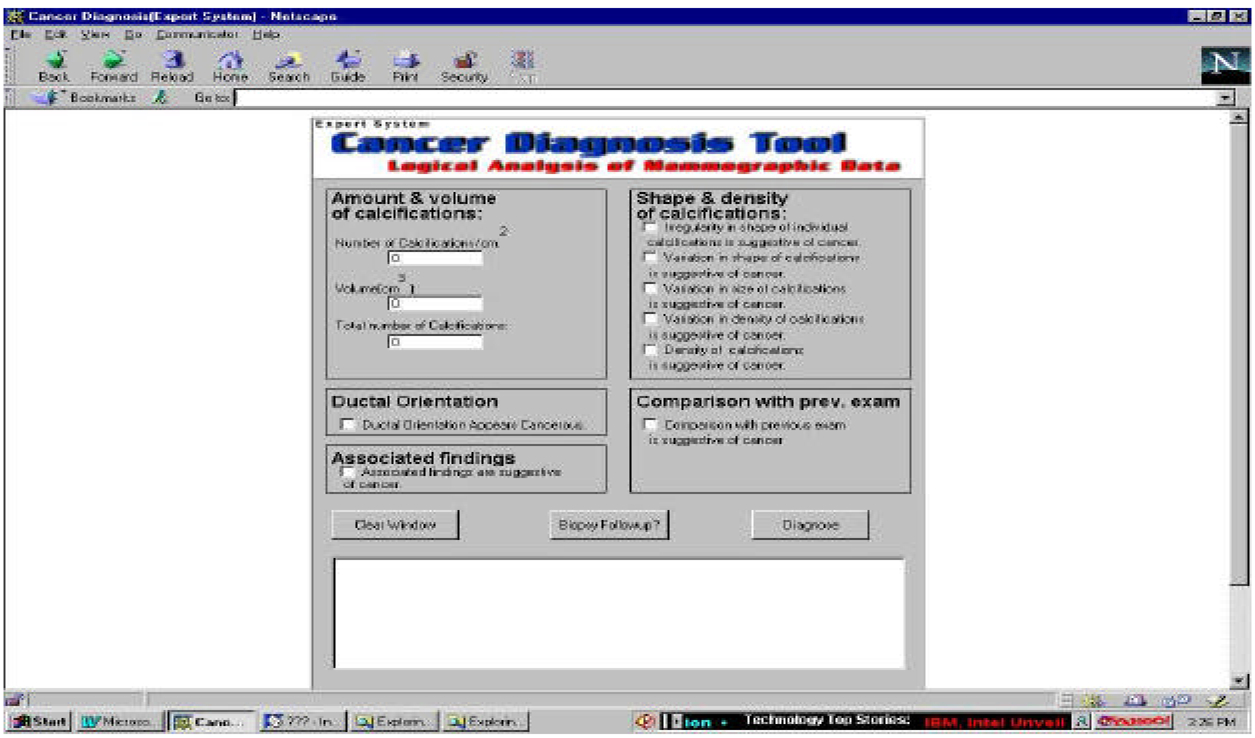
With Logical Analysis of Data method (LAD) [Vityaev, Moskvitin, 1993] 44 statistically significant diagnostic rules were extracted with the conditional probability greater than 0.7. There were 30 regularities with the conditional probability greater then 0.85 and 18 rules with conditional probability more then 0.95. The total accuracy of diagnosis was 82%. The False/negative rate was 6.5% (9 malignant cases were diagnosed as benign) and false/positive rate was 11.9% (16 benign case were diagnosed as malignant). For the 30 more reliable rules we obtained 90% total accuracy, and for the 18 most reliable rules we obtained 96.6% accuracy with only 3 false positive cases (3.4%). Neural Network ("Brainmaker") software had given 100% accuracy on training data but for Round-Robin test the total accuracy fell to 66%. The main reason for this low accuracy is that NN *do not have a mechanism to evaluate statistical significance /reliability of the performance on training data.* Poor results (76% on training data test) were also obtained with Linear Discriminant Analysis ("SIGAMD" software). Decision Tree approach ("SIPINA" software) has performed with accuracy of 76%-82% on training data. This is worse than what we obtained for the LAD method with the much more difficult Round-Robin test. The extremely important false-negative rate was 3-8 cases (LAD), 8-9 cases (Decision Tree), 19 cases (Linear Discriminant Analysis) and 26 (NN).

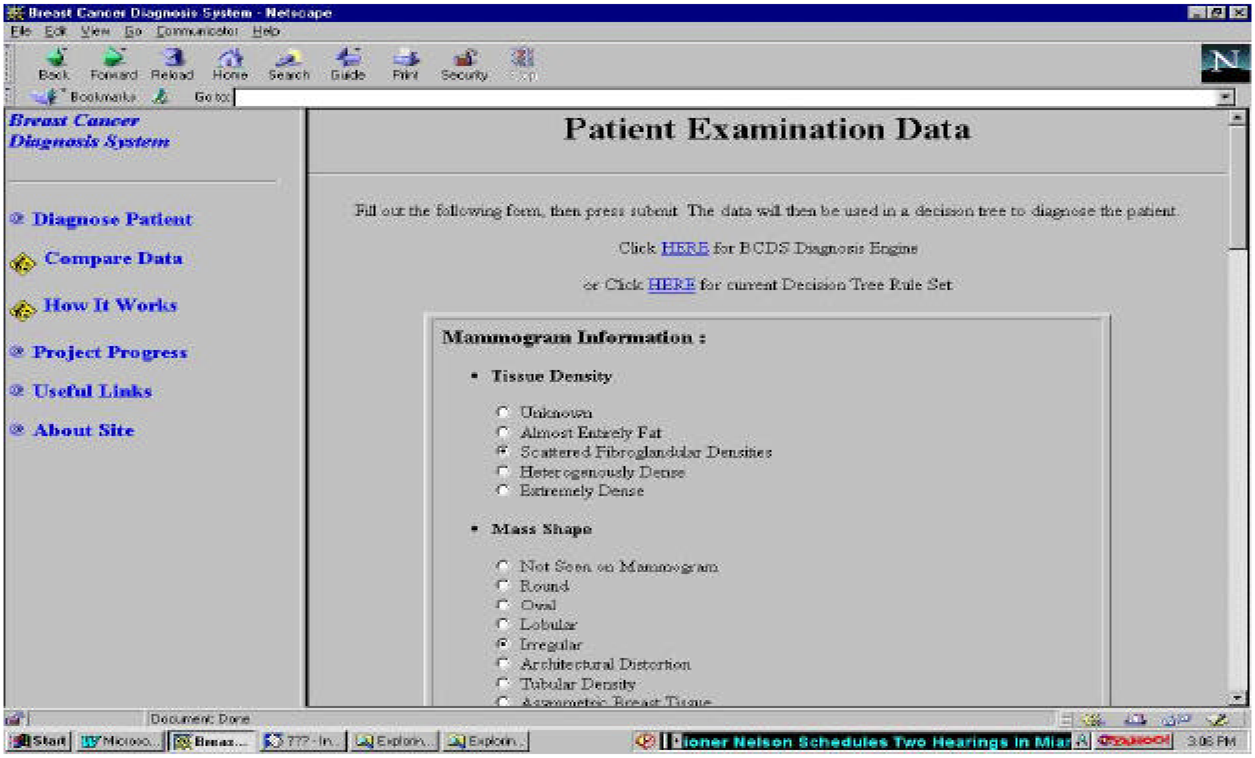
Note also that LAD and decision trees produce diagnostic rules. These rules make a CAD decision process visible to radiologists. With these methods radiologists can control the decision making process. Linear discriminant analysis gives an equation, which separates benign and malignant classes. For example, 0.0670x1-0.9653x2+... represents a case. How would one interpret the weighted number of calcifications/cm2 (x1) plus weighted volume (cm3)(x2)? There is no direct medical sense in this formula. For more technical details of these results see **appendices 2-4 below**

**4. CONCLUSION**

Our study has shown that used Logical Data analysis approach based of first-order logic is appropriate for designing a consultation diagnostic system under requirements presented in section 1. This approach can be used for development of a full-size consultation system.

Appendix1




Breast Cancer Diagnosis System - Netscape
File Edit View Go Communicator Help
Back Forward Reload Home Search Guide Print Security Stop
Bookmarks Go to:
Breast Cancer Diagnosis System
Diagnose Patient
Compare Data
How It Works
Project Progress
Useful Links
About Site
Patient Examination Data
Fill out the following form, then press submit. The data will then be used in a decision tree to diagnose the patient.
Click HERE for BCDS Diagnosis Engine
or Click HERE for current Decision Tree Rule Set
Mammogram Information :
Tissue Density
Unknown
Almost Entirely Fat
Scattered Fibroglandular Densities
Heterogenously Dense
Extremely Dense
Mass Shape
Not Seen on Mammogram
Round
Oval
Lobular
Irregular
Architectural Distortion
Tubular Density
Document: Done
Start
3:06 PM

Appendix 2

**TABLE 1**
**Computational Experiments with Different CAD Methods.**

| Method | statistical significance (0.05) | empirical test | true cancer c->c | false negative c->b | True benign b->b | false posi-tive b->c | refused (unclas-sified) | total accuracy without refused |
|---|---|---|---|---|---|---|---|---|
| LAD1 (0.75) 44 rules | tested F-criterion | Round-Robin | 60 | 9 | 54 | 16 | 17 | 82% |
| LAD2 (0.85) 30 rules | tested F-criterion | Round-Robin | 54 | 4 | 55 | 7 | 46 | 90% |
| LAD3 (0.95) 18 rules | tested F-criterion | Round-Robin | 43 | 3 | 45 | 0 | 68 | 96.6% |
| Neural Network | ? test unknown | Round-Robin | 51 | 26 | 52 | 27 | 0 | 66% |
| Decision Tree 1 | ? test unknown | Training data | 61 | 14 | 56 | 22 | 3 | 76% |
| Decision Tree 2 | ? test unknown | Training data | 57 | 14 | 56 | 18 | 11 | 78% |
| Decision Tree 3 | ? test unknown | Training data | 45 | 8 | 48 | 9 | 43 | 82% |
| Linear Disc. Analysis | testable | Training data | 58 | 19 | 60 | 19 | 0 | 76% |

Appendix 3.

## Examples of Extracted Diagnostic Rules

### Expert Diagnostic Rules

**EXPERT RULE 1:**

**IF** NUMBER of calcifications per $cm^2$ ($w_1$) is large
AND TOTAL number of calcifications ($w_3$) is large
AND irregularity in SHAPE of individual calcifications is marked
**THEN** highly suspicious for malignancy

**EXPERT RULE 2:**

**IF** NUMBER of calcifications per $cm^2$ (w1) large
AND TOTAL number of calcifications is large ($w_3$)
AND variation in SIZE of calcifications ($y_3$) is marked
AND VARIATION in Density of calcifications ($y_4$) is marked
AND DENSITY of calcification ($y_5$) is marked
**THEN** highly suspicious for malignancy.

**EXPERT RULE 3:**

**IF** (SHAPE and density of calcifications are positive for *cancer* AND Comparison with previous examination is *positive for cancer*)
OR (the number and the VOLUME occupied by calcifications are *positive for cancer* AND SHAPE and density of calcifications are *positive for cancer*)
OR (the number and the VOLUME occupied by calcifications are *positive for cancer* AND comparison with previous examination is *positive for cancer*)
OR (DUCTAL orientation is *positive for cancer* OR associated FINDINGS are *positive for cancer*)
**THEN** Biopsy is recommended.

Appendix 4

**Diagnostic Rules extracted from Database**

**DB RULE 1:**

**IF** TOTAL number of calcifications >30
AND VOLUME >5 $cm^3$
AND DENSITY of calcifications is moderate
**THEN** Malignant.
F-criterion **--** significant for 0.05.
Accuracy of diagnosis for test cases --100%.
Radiologist's comment -- This rule might have promise, but I would consider it risky**.**

**DB RULE 2:**

**IF** VARIATION in shape of calcifications is marked
AND NUMBER of calcifications is between 10 and 20
AND IRREGULARITY in shape of calcifications is moderate
**THEN** Malignant.
F-criterion **--** significant for 0.05.
Accuracy of diagnosis for test cases -- 100%.
Radiologist's comment -- I would trust this rule.

**DB RULE 3:**

**IF** variation in SIZE of calcifications is moderate
AND variation in SHAPE of calcifications is mild
AND IRREGULARITY in shape of calcifications is mild
**THEN** Benign.
F-criterion **--** significant for 0.05.
Accuracy of diagnosis for test cases -- 92.86%.
Radiologist's comment -- I would trust this rule.

**REFERENCES**

1**.** Kovalerchuk B., Triantaphyllou E., Despande A., Vityaev E: Interactive learning of Monotone Boolean

Functions, Information Systems, 94;1-4; 87-118,1996,

2. Kovalerchuk B., Triantaphyllou E, Ruiz J: Monotonicity and logical analysis of data: a mechanism for evaluation of mammographic and clinical data, in Kilcoyne RF, Lear JL, Rowberg AH (eds): Computer applications to assist radiology, Carlsbad, CA, Symposia Foundation,1996, pp.191-196.

3. Mitchell, T: Machine learning. NY, McCraw Hill, 1997

4. [TIWDM, 1996]. Third International Workshop on Digital Mammography, University of Chicago, Chicago, IL, June 9-12, 1996,

5. Shtern, F. Novel digital technologies for improved control of breast cancer. in: CAR'96, Computer Assisted Radiology, Proceedings of the International Symposium on Computer and Communication Systems for Image Guided Diagnosis and Therapy, Lemke HU, Vannier MW, Elsevier Science, pp.357-361.

6. CAR'96, Computer Assisted Radiology, Proceedings of the International Symposium on Computer and Communication Systems for Image Guided Diagnosis and Therapy, Lemke HU, Vannier MW, Inamura K, Farman AG, (eds.) Paris, France, June 26-29, 1996, Elsevier Science, 1996.

http://www.elsevier.nl/inca/publications/store/6/0/0/0/1/2/index.htt

7. Vityaev, E., Moskvitin, A. Introduction to discovery theory: Discovery: software system. Computational Systems, Institute of Mathematics, Novosibirsk, 1993, n. 148, pp.117-163 (in Russian).

8. Gurney J. Neural Networks at the crossroads: caution ahead, Radiology, v. 193, n. 1, pp. 27-28, 1994.