Consistent Knowledge Discovery in Medical Diagnosis

Eliminating Contradictions Among Rules in Computer-Aided Systems, Experts Rules, and Databases

Boris Kovalerchuk[1], Evgenii Vityaev[2], James F. Ruiz[3]
[1]Department of Computer Science, Central Washington University
[2]Institute of Mathematics, Russian Academy of Science, Novosibirsk
[3]Department of Radiology, Woman's Hospital, Baton Rouge

Medicine is a science of uncertainty and an art of probability

—Sir William Osler (c.1904)

There are several modern approaches for knowledge discovery in the medical field, some of which have originated in the artificial intelligence area. In this article, we discuss the application of these methods for medical diagnosis, using features extracted from mammograms. We describe a method that can be used to discover a consistent set of logical diagnostic rules for breast cancer diagnosis. These rules may serve as the core of a comprehensive computer-aided diagnostic system, which has the ultimate purpose of providing a second diagnostic opinion. Consistency of the system means that there are no contradictions among rules in a computer-aided diagnostic system, rules used by an experienced radiologist, and a database of pathologically confirmed cases. We have developed a method for discovering a consistent set of diagnostic rules, and we show advantages of the method for development of a breast cancer computer-aided diagnostic system.

Overview: Breast Cancer Diagnosis and Knowledge Discovery

In the US, breast cancer is the most common female cancer [1]. The most effective tool in the battle against breast cancer is screening mammography. However, it has been found that intra- and interobserver variability in mammographic interpretation is significant (up to 25%) [2]. Additionally, several retrospective analyses have found error rates ranging from 20% to 43%. These data clearly demonstrate the need to improve the reliability of mammographic interpretation.

The problem of identifying cases suspicious for breast cancer using mammographic information about clustered calcifications is considered here. Examples of mammographic images with clustered calcifications are shown in Figs. 1-3. Calcifications are seen in most mammograms and commonly indicate the presence of benign fibrocystic change. However, certain features can indicate the presence of malignancy. These figures demonstrate the broad spectrum of appearances that might be present within a mammogram.

Figure 1 shows calcifications that are irregular in size and shape. These are biopsy-proven, malignant-type calcifications. Figure 2 presents a cluster of calcifications within a low-density, ill-defined mass. Again, these calcifications vary in size, shape, and density, suggesting that a cancer has produced them. Finally, Fig. 3 is an example of a carcinoma, which has produced a high-density nodule with irregular spiculated margins. While there are calcifications in the area of this cancer, they are all nearly spherical in shape and quite uniform in their density. This high degree of regularity suggests a benign origin. At biopsy, the nodule proved to be a cancer, while the calcifications were associated with a benign fibrocystic change.

There is promising computer-aided diagnostic research aimed to improve the situation [3-8]. Knowledge discovery in medical diagnosis includes two major steps: (S1) extracting diagnostic features and (S2) extracting diagnostic rules based on these features.

Typical knowledge discovery research in breast cancer diagnosis includes:

- (C1) a few hundred data units,

- (C2) about a dozen diagnostic features given or extracted from images,
- (C3) knowledge discovery process (KD process)

Neural networks, nearest neighbor methods, discriminant analysis, cluster analysis, linear programming, and genetic algorithms are among the most common knowledge discovery tools. Data mining in other fields tends to use larger databases and discover larger sets of rules using these techniques. At the same time, mammography archives at hospitals around the world contain millions of mammograms and biopsy results. Currently, the American College of Radiology (ACR) supports the National Mammography Database (NMD) Project (http://www.eskimo.com/~briteoo/nmd) with a unified set of features [9]. Several universities and hospitals have developed mammography image bases that are available on the Internet. Such efforts provide the opportunity for large-scale data mining and knowledge discovery for breast cancer diagnosis. Data mining experience in business applications have shown that a large database can be a source of useful rules, but the useful rules may be accompanied by larger set of irrelevant or incorrect rules. A great deal of time may be required for experts to select only nontrivial rules. In this article, we address this problem by offering a method of rule extraction consistent with expert opinions.

Traditional expert systems rely on diagnostic rules extracted from experts. Systems based on machine-learning techniques rely on an available databases for discovering diagnostic rules. These two sets of rules may contradict each other. A radiologist may not trust rules, as they may contradict his/her rules and experience. Also, a radiologist may have questionable or incorrect rules, while the data and image base may have questionable or incorrect records. These contradictions make the design of a computer-aided diagnostic system extremely complex.

There are two tasks:

- (T1) Identify contradictions among diagnostic rules.
- (T2) eliminate contradictions.

If the first task is solved, the second one can be approached by cleaning the records in the database, adding more features, using more sophisticated rule extraction methods, and testing the competence of a medical expert.

In this article, we concentrate on the extraction of rules from an expert and from a collection of data and then attempt to identify contradictions. If rule extraction is performed without this purpose in mind, it is difficult to recognize a contradiction. Also, rules generated by an expert and data-driven rules may be incomplete, as they may cover only a small fraction of possible feature combinations. This limitation may make it impossible to confirm that rules are consistent with an available database. Additional new cases or features can make the contradictions visible. Therefore, the major problem here is discovering sufficient, complete, and comparable sets of expert rules and data-driven rules. Completeness is critical for comparison. For example, suppose that an expert and data-driven rules cover only 3% of possible feature combinations (cases) and assume that there are no contradictions between these rules. Then, there is still plenty of room for contradiction in the remaining 97% of the cases.

We are developing methods to discover complete sets of expert rules and data-driven rules. This objective presents us with an exponential nontractable problem of extracting diagnostic rules. A brute-force method may require asking the expert thousands of questions. Such a dialog is a well-known problem for expert system development [10]. For example, for 11 binary diagnostic features of clustered calcifications, there are ($2^{11} = 2048$) feature combinations, each representing a new case. A brute-force method would require questioning a radiologist on each of these 2048 combinations.

A related problem is that, in attempting to analyze a complex system, experts may find it difficult or impossible to articulate confidently the large number of interactions among features. For such problems, it becomes increasingly impractical to conduct knowledge acquisition and to extract meaningful rules. In general experience, about 60 to 70% of the time taken to develop rule-based systems is spent on knowledge acquisition. Thus, knowledge engineering to extract hundreds of rules becomes the bottleneck in the process. Perhaps the most important reason for considering an expert system approach to a problem is that a rule-based system approach seeks to behave like an expert. It exhibits the "feel" of an expert and can explain and justify a conclusion. The expert ponders alternative scenarios, and thus might say: "I think that under the circumstances, X, the most likely conclusion is Y. But if an additional fact, say F, were present, the more likely conclusion might be P." If a problem is "decomposable," where the interactions among variables are limited and experts can articulate their decision process with confidence, a rule-based approach is a good candidate and the system may scale well [10].

We have developed an effective mechanism for decomposition and to exploit monotonicity so as to make this problem tractable.

Creating a consistent rule base includes the following steps:

1. Finding data-driven rules *not* discovered by asking an expert.

2. Analysis of these new rules by a medical expert using available proven cases. A list of these cases from the database can be presented to an expert. The expert can check:
   - Is a new rule discovered because of *misleading cases*? The rule may be rejected and training data can be extended.
   - Does the rule *confirm* existing expert knowledge? Perhaps the rule was not sufficiently transparent for the expert. The expert may find that the rule is consistent with his/her previous experience, but he/she would like more evidence. The rule can increase the confidence of his/her practice.
   - Does the rule *identify new relationships* that were not previously known to the expert? The expert can find that the rule is promising.

3. Finding rules that are *contradictory* to his/her knowledge or understanding. Rules express the interconnections of the features presented within training cases. This means that there are two possibilities:
   - The rule was discovered using misleading cases. This rule must be rejected and training data must be extended.
   - The expert can admit that his/her ideas have no real basis. The system improves expert experience.

This article is based on and extends our previous research [11-18].

Method For Discovering Diagnostic Rules From a Database

A machine-learning method, called machine methods for discovering regularities (MMDR) [18], can be applied for the discovery of diagnostic rules for breast cancer diagnosis. The method expresses patterns in first-order logic and assigns

probabilities to rules generated by composing patterns. Learning systems based on first-order representations have been successfully applied to many problems in chemistry, physics, medicine, finance, and other fields [11-14,18]. As with any technique based on logic rules, this technique allows one to obtain human-readable forecasting rules that are interpretable in medical language and also provides a diagnosis [19]. A medical specialist can evaluate the correctness of the diagnosis as well as the diagnostic rule. The critical issue in applying data-driven forecasting systems is generalization. MMDR and related "discovery" software systems [18] generalize data through "law-like" logical probabilistic rules.

Conceptually, law-like rules come from the philosophy of science. These rules attempt to mathematically capture the essential features of scientific laws: (1) high level of generalization, (2) simplicity (Occam's razor), and (3) refutability. The first feature—generalization—means that any other regularity covering the same events would be less general; i.e., applicable only to a subset of events covered by the law-like regularity. The second feature—simplicity—reflects the fact that a law-like rule is shorter than other rules. The law-like rule (R1) is more refutable than another rule (R2) if there are more testing examples that refute (R1) than (R2) but the examples fail to refute (R1).

Formally, we present an IF-THEN rule C as $A_1 \& ... \& A_k \Rightarrow A_0$, where the IF part, $A_1 \& ... \& A_k$, consists of true/false logical statements $A_1,...,A_k$, and the THEN part consists of a single logical statement A_0. Statements A_i are some given refutable statements or their negations, which are also refutable. Rule C allows us to generate subrules with a truncated IF part; e.g., $A_1 \& A_2 \Rightarrow A_0$, $A_1 \& A_2 \& A_3 \Rightarrow A_0$, and so on.

For rule C, its conditional probability $Prob(C) = Prob(A_0/A_1 \& ... \& A_k)$ is defined. Similarly, conditional probabilities $Prob(A_0/A_{i1} \& ... \& A_{ih})$ are defined for subrules C_i of the form $A_{i1} \& ... \& A_{ih} \Rightarrow A_0$.

We use conditional probability, $Prob(C) = Prob(A_0/A_1 \& ... \& A_k)$, for estimating forecasting power of the rule to predict A_0. The rule is "law-like" if all of its subrules have a statistically significant lower conditional probability than the rule. Each subrule C_i generalizes rule C; i.e., potentially, C_i is true for a larger set of instances [19]. Another definition of "law-like" rules can be stated in terms of generalization. The rule is "law-like" if it cannot be generalized without producing a statistically significant reduction in its conditional probability. "Law-like" rules defined in this way hold all three properties of scientific laws. They are (1) general from a logical perspective, (2) simple, and (3) refutable. Below, we present some rules extracted using this approach.

The "discovery" software searches all chains C_1, C_2, ..., C_{m-1}, C_m of nested "law-like" subrules, where C_1 is a subrule of rule C_2, $C_1 = sub(C_2)$, C_2 is a subrule of rule C_3, $C_2 = sub(C_3)$, and finally C_{m-1} is a subrule of rule C_m, $C_{m-1} = sub(C_m)$. Also, $Prob(C_1) < Prob(C_2)$, ..., $Prob(C_{m-1}) < Prob(C_m)$. There is a theorem [17] that all rules that have a maximum value of conditional probability can be found at the end of such chains. The algorithm stops generating new rules when they become too complex (i.e., statistically insignificant for the data), even if the rules are highly accurate on training data. The Fisher statistical criterion is used in this algorithm for testing statistical significance. The obvious other stop criterion is time limitation.

Theoretical advantages of MMDR generalization are presented in [12], [17] and [18]. This approach has some similarity with the hint approach [20]. We use mathematical formalisms of first-order logic rules described in [21]-[23]. Note that a class of general propositional and first-order logic rules covered by MMDR is wider than a class of decision trees [19].

Figure 4 describes the steps of MMDR. In the first step, we select and/or generate a class of logical rules suitable for a particular task. The next step is learning the particular first-order logic rules using available training data. Then we test first-order logic rules on training data using the Fisher statistical criterion. After that we select statistically significant rules and apply Occam's razor principle: the simplest hypothesis (rule) that fits the data is preferred [19]. The last step is creating interval and threshold forecasts using selected logical rules: IF A(x,y,...,z) THEN B(x,y,...,z).

Method for Extracting Diagnostic Rules from Medical Experts

Hierarchical Approach

The interview of a radiologist to extract rules is managed using an original method of monotone Boolean function restoration [12]. One can ask a radiologist to evaluate a particular case when a number of features take on a set of specific values. A typical query will have the following format:

"If feature 1 has value V_1, feature 2 has value V_2,..., feature n has value V_n, then should biopsy be recommended or not?

"Or, does the above setting of values correspond to a case suspicious of cancer or not?"

Each set of values $(V_1, V_2,...,V_n)$ represents a possible clinical case. It is practically impossible to ask a radiologist to generate diagnoses for thousands of possible cases. A hierarchical approach combined with the use of the property of monotonicity makes the problem manageable.

We construct a hierarchy of *medically interpretable* features from a very generalized level to a less generalized level. This hierarchy follows from the definitions of the 11 medically oriented binary attributes. The medical expert indicated that the original 11 binary attributes, w_1, w_2, w_3, y_1, y_2, y_3, y_4, y_5, x_3, x_4, x_5, could be organized in terms of a hierarchy, with development of two new generalized attributes x_1 and x_2:

Level 1 (5 Attributes)		**Level 2 (All 11 Attributes)**
x_1	7	w_1, w_2, w_3
x_2	7	y_1, y_2, y_3, y_4, y_5
x_3	7	x_3
x_4	7	x_4
x_5	7	x_5,

We consider five binary features x_1, x_2, x_3, x_4, and x_5, on level 1. A new generalized feature:

x_1 — "Amount and volume of calcifications"

with grades (0 - "benign" and 1 - "cancer") introduced based on features:

w_1 — number of calcifications/cm^3,
w_2 — volume of calcification/cm^3 and
w_3 — total number of calcifications.

We view x_1 as a function $v(w_1, w_2, w_3)$ to be identified.

Similarly, a new feature:

x_2— "Shape and density of calcification"

with grades (1) for "marked" or "cancer" and (0) for "minimal" or "benign," generalizes features:

y_1 — "Irregularity in shape of individual calcifications"

y_2 — "Variation in shape of calcifications"

y_3 — "Variation in size of calcifications"

y_4 — "Variation in density of calcifications"

y_5 — "Density of calcifications"

We view x_2 as a function $x_2 = \psi(y_1, y_2, y_3, y_4, y_5)$ to be identified for cancer diagnosis. The described structure is presented in Fig. 5.

A similar structure was produced for a decision regarding biopsy. The expert was requested to review both the structure and answers for the questions:

- "Can function f_1 be assumed the same for both problems?"
- "Can function f_2 be assumed the same for both problems?"

The expert indicated that these two functions, ν and ψ, should be common to both problems: (P1) recommendation biopsy and (P2) cancer diagnosis. Therefore, the following relation is true regarding the f_i (for i = 1, 2) and the two φ and ψ functions:

$$f_i(x_1,x_2,x_3,x_4,x_5) = f_i(\varphi(w_1,w_2,w_3), y(y_1,y_2,y_3,y_4,y_5), x_3,x_4,x_5),\ i = 1,2.$$

Further levels of hierarchy can be developed for better describing the problem. For example, y_1 ("irregularity in shape of individual calcifications") may be found in three grades: "mild" (or t_1), "moderate" (or t_2) and "marked" (or t_3). Next, observe that it is possible to change (i.e., generalize) the operations used in the function $\psi(y_1,y_2,..,y_5)$. For instance, we may have mentioned function ψ as follows: $\psi(y_1,y_2,..,y_5) = y_1 \,\&\, \vee\, y_3 \,\&\, y_4 \,\&\, y_5$, where & and $\vee$ are the binary, logical operations for "AND" and "OR." respectively. Then, & and $\vee$ can be substituted for one of their multivalued logic analogs; for example, $x \,\&\, y = \min(x,y)$ and $x \vee y = \max(x,y)$, as in fuzzy logic (see, for example, [11]). This decomposition is presented in Fig. 5.

We assume that x_1 is the number and the volume occupied by calcifications, in a binary setting, as follows: (0-"against cancer,"1-"for cancer"). Similarly, let:

x_2 — {shape and density of calcifications}, with: 0-"benign,"1-"cancer"

x_3 — {ductal orientation}, with: 0 - "benign," 1-"cancer"

x_4 — {comparison w. previous examination}, with: 0 - "benign," 1-"cancer"

x_5 — {associated findings}, with: 0-"benign,"1-"cancer."

Monotonicity

To understand how monotonicity is applied to the breast cancer problem, consider the evaluation of calcifications in a mammogram. Given the above definitions, we can represent clinical cases in terms of binary vectors with five generalized features as (x_1,x_2,x_3,x_4,x_5). Next, consider the two clinical cases that are represented by the binary sequences (10110) and (10100). If one is given that a radiologist correctly diagnosed (10100) as a malignancy, then, by utilizing the property of monotonicity, we can also conclude that the clinical case (10110) should also be a malignancy. This conclusion is based on the systematic coding of all features "suggestive for cancer" as 1.

Observe that in (10100) we had two indications for cancer:

- $x_3 = 1$ (ductal orientation having value of 1; suggesting cancer) and
- $x_1 = 1$ (Amount and volume of calcifications with value 1 indicating cancer).

In the second clinical case, we have these two observations for cancer and also $x_4 = 1$ (a comparison with previous examinations suggesting cancer). In the same manner, if we know that (01010) is not considered suspicious for cancer, then the case (00000) should also not be considered suspicious. This is true because in the second case we have less evidence indicating the presence of cancer. The above considerations are the essence of how our algorithms function. They can combine logical analysis of data with monotonicity and can generalize accordingly. In this way, the weaknesses of the brute-force methods can be avoided.

It is assumed that if the radiologist believes that the case is malignant, then he/she will recommend a biopsy. More formally, these two subproblems are defined as follows:

The Clinical Management Subproblem (P1): One and only one of the following two disjoint outcomes is possible:

1) "Biopsy is necessary" or

2) "Biopsy is not necessary."

The Diagnosis Subproblem (P2): Similarly as above, one and only one of the following two disjoint outcomes is possible. That is, a given case is:

1) "Suspicious for malignancy" or

2) "Not suspicious for malignancy."

Our goal here is to extract the way the system operates in the form of two discriminant Boolean functions, f_2 and f_1:

1. Function f_1 returns true (1) value if the decision is "biopsy is necessary," false (0) otherwise.

2. Function f_2 returns true (1) value if the decision is "suspicious for malignancy," false (0) otherwise.

The first function is related to the first subproblem, while the second function is related to the second subproblem. There is an important relation between subproblems P1 and P2 and functions $f_1(\alpha)$, $f_2(\alpha)$. The problems are nested; i.e., if the case is suggestive of cancer ($f_2(a) = 1$) then biopsy should be recommended ($f_1(\alpha) = 1$) for this case, therefore $f_2(\alpha) = 1 \Rightarrow f_1(\alpha) = 1$. Also, if biopsy is not recommended ($f_1(\alpha)=0$) then the case is not suggestive of cancer ($f_2(\alpha)=0$), therefore $f_1(\alpha) = 0 \Rightarrow f_2(\alpha) = 0$. The last two statements are equivalent to $f_2(\alpha) \geq f_1(\alpha)$ and $f_1(\alpha) \leq f_2(\alpha)$, respectively, for case α. Let $E^+_{n,1}$ be a set of α sequences from E_n, such that $f_1(\alpha) = 1$ (biopsy positive cases). Similarly, $E^+_{n,2}$ is a set of α sequences from E_n, such that $f_2(\alpha) = 1$ (cancer positive cases). Observe that the nested property formally means that $E^+_{n2} \subseteq E^+_{n1}$ (for all cases suggestive of cancer, biopsy should be recommended) and $f_2(\alpha) \geq f_1(\alpha)$ for all $\in E_n$.

The previous two inter-related subproblems, P1 and P2, can be formulated as a restoration problem of two nested monotone Boolean functions, f_1 and f_2. A medical expert was presented with the ideas of monotonicity and nested functions, as above, and he felt comfortable with the idea of using nested monotone Boolean functions. Moreover, the dialogue that followed confirmed the validity of this assumption. Similarly, the function $x_2 = \psi(y_1, y_2, y_3, y_4, y_5)$ for x_2 ("Shape and density of calcification") was confirmed to be a monotone Boolean function.

A Boolean function is a compact presentation of the set of diagnostic rules. A Boolean discriminant function can be presented in the form of a set of logical IF-THEN rules, but it is not necessary that

these rules stand for a single tree, as in the decision-tree method. A Boolean function can produce a diagnostic discriminant function that cannot be produced by the decision-tree method. For example, the biopsy subproblem is stated as:

$$f_1(x) = x_2x_4 \vee x_1x_2 \vee x_1x_4 \vee x_3 \vee x_5. \quad (1)$$

This formula is read as follows:

IF (x_2 AND x_4) OR (x_1 AND x_2) OR (x_1 AND x_4) OR (x_3) OR (x_5) THEN Biopsy is recommended

In medical terms this translates as:

IF (shape and density of calcifications suggests cancer AND comparison with previous examination suggests cancer) OR (the number and the volume occupied by calcifications suggests cancer AND shape and density of calcifications suggests cancer) OR (the number and the volume occupied by calcifications suggests cancer AND comparison with previous examination suggests cancer) OR (ductal orientation suggests cancer) OR (associated findings suggests cancer) THEN Biopsy is recommended.

Figure 6 presents the major steps in rule extraction from a medical expert: (1) develop a hierarchy of concepts and present them as a set of monotone Boolean functions, (2) restore each of these functions with a minimal sequence of questions to an expert, (3) combine discovered functions into a complete diagnostic function, and (4) present the complete function as a traditional set of simple diagnostic rules: *If A and B and...F then Z*.

Next, we describe step (2)—restoring each of monotone Boolean functions with minimal sequence of questions for the expert (Fig. 7).

The last block (2.5) in Fig. 7 provides for interviewing an expert with a minimal dynamic sequence of questions. This sequence is based on the fundamental Hansel lemma [11,24]. We omit a detailed description of the specific mathematical steps, which can be found in [11]. The general idea of these steps is given using an example of the interactive session in Table 1. A minimal sequence of questions means that we reach the minimum of the Shannon Function [11]; i.e., a minimal number of questions is required to restore the most complex monotone Boolean function with *n* arguments. This sequence is not a sequence written in advance. Rather, it depends on the previous answers of a medical expert; therefore, each subsequent question is defined dynamically, as illustrated in Table 1. Columns 2, 3, and 4 present values of the above-defined functions, f_1, f_2, and ψ (see the "Hierarchical Approach" section above). We omit a restoration of function $\varphi(w_1, w_2, w_3)$ because few questions are needed to restore this function, but the general scheme is the same as for f_1, f_2, and ψ, with consideration of all binary triples such as (010), (110), and so on. In Table 1, the first question is: "Does the sequence (01100) represent a case requiring a biopsy?" Here, x_1=0 and (01100) = (x_1, x_2, x_3, x_4, x_5). If the answer is "yes" (1), then the next question will be about biopsy for the case (01010). If the answer is "no" (0), then the next question will be about biopsy for (11100). This sequence of questions is not accidental. As mentioned above, it is inferred from the Hansel lemma [11]. All 32 possible cases with five binary features (x_1, x_2, x_3, x_4, x_5) are presented in column 1 of Table 1. They are grouped, and the groups are called Hansel chains. The sequence of chains begins from the shortest chain [#1—(01100) and (11100)]. This chain consists of two ordered cases, (01100) < (11100) for five binary features. Then the largest chain, #10, consists of six ordered cases: (00000) < (00001) <(00011) < (00111) < (01111) < (11111). Similarly, the cases are ordered as vectors in each chain.

To construct the chains presented in Table 1 (with five dimensions such as x_1, x_2, x_3, x_4, x_5 or y_1, y_2, y_3, y_4, y_5) a sequential process is used. First, all one-dimensional chains (in E_1) are generated, and then they are used to generate chains of higher dimensions, up to dimension five. Each step of chain generation consists of using current i-dimensional chains to generate (i+1)-dimensional chains. The generation of chains for the next dimension (i+1) is a five-step "clone-grow-cut-add" process. We clone an i-dimensional chain; e.g., having one-dimensional chain (0) < (1) we produce its copy: (0) < (01). Then we grow these chains, adding the second dimension, but differently:

Chain 1: (00) < (01)

Chain 2: (10) < (11).

Here, 0 is added to the left of both cases in chain 1; and 1 is added to the both cases in chain 2.

Next, we cut the head case (11) from chain 2 and add it as a head to chain 1, producing two two-dimensional Hansel chains:

New chain 1: (00) < (01) < (11) and

New chain 2: (10).

This process continues and stops in the fifth dimension for <x_1, x_2, x_3, x_4, x_5> and <y_1, y_2, y_3, y_4, y_5>. Table 1 presents the result of this process. The chains are numbered there from 1 to 10, and each case has its number in the chain; e.g., 1.2 means the second case in the first chain. Asterisks in columns 2, 3, and 4 mark answers obtained from an expert; e.g., 1* for case (01100) in column 3 means that the expert answered "yes." The remaining answers for the same chain in column 3 are automatically obtained using monotonicity. The value f_1(01100) = 1 for case 1.1 is extended for cases 1.2, 6.3, and 7.3 in this way. Similarly, values of the third monotone Boolean functions ψ are computed using the Table 1. (The attributes in the sequence (10010) are interpreted as y_1, y_2, y_3, y_4, y_5 instead of x_1, x_2, x_3, x_4, x_5 used for f_1 and f_2. The Hansel chains are the same as long as the number of attributes is the same, five in this case).

Columns 5 and 6 list cases for extending function values without asking an expert. Column 5 is for extending function values from 1 to 1, and column 6 is for extending them from 0 to 0. If an expert were to give an answer opposite (f_1(01100) = 0) to that presented in Table 1 for function f_1 and case 1.1 (01100), then this 0 value could be extended in column 2 for cases 7.1 (00100) and 8.1 (01000). These cases are listed in column 6 for case (01100). There is no need to ask an expert about cases 7.1 (00100) and 8.1 (01000) because monotonicity provides the answer. The negative answer f_1(01100) = 0 cannot be extended for f_1(11100). An expert should be queried regarding f_1(11100). If his/her answer is negative, f_1(11100) = 0, then this value can be extended for cases 5.1 and 3.1 listed in column 6 for case 1.2. Relying on monotonicity, the value of f_1 for these cases will also be 0.

The total number of cases with an asterisk (*) in column 1 is equal to 13, and for columns 3 and 4 they are, respectively, 13 and 12. These numbers show that 13 questions are needed to restore each of f_1 and f_2 as functions of x_1, x_2, x_3, x_4, x_5 and that 12 questions are needed to restore as a function of y_1, y_2, y_3, y_4, y_5. This is only 37.5% of 32 possible questions and 60% of a possible maximum generated by the Hansel lemma.

Full restoration of either one of functions f_1 and f_2 with 11 arguments (see above) without any optimization of the interview process would have required up to $2^{11} = 2{,}048$ calls (membership inquires) to the medical expert. Note that practically all studies in breast cancer computer-aided diagnostic systems derive diagnostic rules using significantly less than 1000 cases [25]. However, according to the Hansel lemma and under the assumption of monotony, an optimal (i.e., a minimal) dialogue for restoring a monotone Boolean would require at most:

$$\binom{11}{5} + \binom{11}{6} = 2 \times 462 = 924$$

calls to a medical expert. This new value is 2.36 times smaller than the previous upper limit of 2048 calls. However, even this upper limit of 924 calls can be reduced further. The hierarchy presented in Fig. 5 reduces the maximum number of questions needed to restore monotone Boolean functions of 11 binary variables to 72 questions (nondeterministic questioning) and to 46 using the Hansel lemma. The actual number of questions asked was about 40, including both nested functions (cancer and biopsy) described below, (i.e., about 20 questions per function).

Discovering Diagnostic Rules from a Database

The next task is the discovery of rules from data. This study was accomplished using an extended set of features. The set of features listed in the "Hierarchical Approach" section was extended with two features: *Le Gal type* and *density of parenchyma*, with the following diagnostic classes: "malignant," "benign," and "high risk of malignancy." We extracted several dozen diagnostic rules that were statistically significant on the 0.01, 0.05, and 0.1 levels (F-criterion).

Rules were extracted using 156 cases (73 malignant, 77 benign, two highly suspicious, and four with mixed diagnosis). In the round-robin test, our rules diagnosed 134 cases and refused to diagnose 22 cases. The total accuracy of diagnosis is 86%. Incorrect diagnoses were obtained in 19 cases (14% of diagnosed cases). The false-negative rate was 5.2% (seven malignant cases were diagnosed as benign) and the false-positive rate was 8.9% (12 benign cases were diagnosed as malignant). Some of the rules are shown in Table 2, which presents examples of discovered rules with their statistical significance.

Figure 8 presents results for another selection criterion: level of conditional probability. We studied three levels: 0.7, 0.85, and 0.95. A higher level of conditional probability decreases the number of rules and diagnosed patients but increases accuracy of diagnosis. Results are marked as MMDR1, MMDR2, and MMDR3. We extracted 44 statistically significant diagnostic rules for the 0.05 level of F-criterion with a conditional probability no less than 0.75 (MMDR1). There were 30 rules with a conditional probability no less than 0.85 (MMDR2), and 18 rules with a conditional probability no less than 0.95 (MMDR3). The total accuracy of diagnosis was 82%. The false negative rate was 6.5% (nine malignant cases were diagnosed as benign), and the false positive rate was 11.9% (16 benign cases were diagnosed as malignant). The most reliable 30 rules delivered a total accuracy of 90%, and the 18 most reliable rules performed with 96.6% accuracy with only three false positive cases (3.4%).

Neural network software ("Brainmaker," California Scientific Software) had given 100% accuracy on training data, but for the round-robin test, the total accuracy fell to 66%. The main reason for this low accuracy is that neural networks do not evaluate the statistical significance of the perfect performance (100%) on training data. Poor results (76% on training data test) were also obtained with linear discriminant analysis ("SIGAMD" software, StatDialogue Software, Moscow, Russia). The decision-tree approach ("SIPINA" software, Université Lumière, Lyon, France) performed with accuracy of 76%-82% on training data. This is worse than what we obtained for the MMDR method with the much more difficult round-robin test (Fig. 8). The very important false-negative rate was 3-8 cases (MMDR), 8-9 cases (decision tree), 19 cases (linear discriminant analysis) and 26 cases (NN). In these experiments, rule-based methods (MMDR and decision trees) outperformed the other methods.

Note also that only MMDR and decision trees produce diagnostic rules. These rules make a computer-aided diagnostic decision process visible, transparent **(DOESN'T MAKE SENSE?)** to radiologists. With these methods, radiologists can control and evaluate the decision-making process. Linear discriminant analysis gives an equation that separates benign and malignant classes. For example, $0.0670x_1 - 0.9653x_2 + \ldots$ represents a case. How would one interpret a weighted number of calcifications/cm^2 ($0.0670x_1$) plus a weighted volume (cm^3); i.e., $0.9653x_2$? There is no direct medical sense in this arithmetic.

Rules Extracted from the Expert

Examples of Extracted Diagnostic Rules

Below, we present examples of rules discovered using the technique described above.

EXPERT RULE (ER1):
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 suspicious for malignancy.

EXPERT RULE 2 (ER2):
IF *NUMBER of calcifications per cm^2* (w_1) 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 suspicious for malignancy.

EXPERT RULE 3 (ER3):
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.

Below, we present briefly some other extracted rules in formal notation. MAL stands for suspicious for malignancy.

IF $w_2 \& y_1$ THEN MAL

IF w_2&y_2 THEN MAL
IF w_2&y&$_3$&y_4&y_5 THEN MAL
IF w_1&w_3&y_2 THEN MAL
IF w_1&w_3&x_5 THEN MAL

Rule Extraction Through Monotone Boolean Functions

We obtained Boolean expressions for shape and density of calcification $x_2 = \psi(y_1, y_2, y_3, y_4, y_5)$ from the information depicted in Table 1 (columns 1 and 4) with the following steps:

(i) Find all the maximal lower units for all chains as elementary conjunctions

(ii) Exclude the redundant terms (conjunctions) from the end formula. See expression (2) below. Thus, from Table 1 (columns 2, 4) we obtained:

$$x_2=\psi(y_1, y_2, y_3, y_4, y_5) = y_1y_2y_2y_3 \vee y_2y_4 \vee y_1y_3 \vee y_1y_4 \vee y_2y_3y_4 \vee y_2y_3y_5 \vee y_2 \vee y_1 \vee y_3y_4y_5$$

and then simplified it to $y_2 \vee y_1 \vee y_3y_4y_5$.

As above, from columns 2 and 3 we obtained the initial components of the target functions of x_1, x_2, x_3, x_4, x_5 for the biopsy subproblem as follows:

$$f_1(x) = x_2x_3 \vee x_2x_4 \vee x_1x_2 \vee x_1x_4 \vee x_1x_3 \vee x_3x_4 \vee x_3 \vee x_2x_5 \vee x_1x_5 \vee x_5$$

and for the cancer subproblem to be defined as:

$$f_2(x) = x_2x_3 \vee x_1x_2x_4 \vee x_1x_2 \vee x_1x_3x_4 \vee x_1x_3 \vee x_3x_4 \vee x_3 \vee x_2x_5 \vee x_1x_5 \vee x_4x_5.$$

The simplification of these disjunctive normal form (DNF) expressions allowed us to exclude some redundant conjunctions. For instance, in x_2 the term y_1y_4 is not necessary because y_1 covers it. Thus, the right-hand side of Eqs. (1) to (4) forms the minimal disjunctive normal form DNFs.

Using this technique, we extracted 16 rules for the diagnostic class "suspicious for malignancy" and 13 rules for the class "biopsy" [see Eqs. (5) and (6) below for mathematical representation]. All these rules are obtained from Eq. (6).

WHERE/WHAT IS EQUATION 1? Similarly, for the second subproblem ("highly suspicious for cancer") the function that we found was:

$$f_2(x) = x_1x_2 \vee x_3 \vee (x_2 \vee x_1 \vee x_4)x_5 \quad (2)$$

Regarding the second level of the hierarchy (which, recall, has 11 binary features) we interactively constructed the following functions (interpretation of the features is presented below):

$$x_1 = \nu(w_1,w_2,w_3) = w_2 \vee w_1w_3 \quad (3)$$

and

$$x_2 = \psi(y_1,y_2,y_3,y_4,y_5) = y_1 \vee y_2 \vee y_3y_4y_5 \quad (4)$$

By combining the functions in Eqs. (1)-(4), we obtained the formulas of all 11 features for biopsy:

$$f_1(x) = (y_2 \vee y_1 \vee y_3y_4y_5)x_4 \vee (w_2 \vee w_1w_3)(y_2 \vee y_1 \vee y_3y_4y_5) \vee (w_2 \vee w_1w_3)x_4 \vee x_3 \vee x_5 \quad (5)$$

and for suspicious for cancer:

$$f_2(x) = x_1x_2 \vee x_3 \vee (\xi_2 \vee x_1 \vee x_4)x_5 = (w_2 \vee w_1w_3)(y_1 \vee y_2 \vee y_3y_4y_5) \vee x_3 \vee (y_1 \vee y_2 \vee y_3y_4y_5) \vee (w_2 \vee w_1w_3 \vee x_4)x_5 . \quad (6)$$

Comparison of Data-Based and Expert Diagnostic Rules

Below, we compare some rules extracted from 156 cases using data mining algorithms and by interviewing the radiologist.

From the database we extracted the rule DBR1:

IF *NUMber of calcifications per cm*^{2} (w_1) is between 10 and 20 AND *VOLume* (w_2) < 5 cm^3
THEN *Malignant*

The closest expert rule is ER1:

IF *NUMber of calcifications per cm*^{2} (w_1) large AND *TOTal number of calcifications* (w_3) is large
AND *irregularity in SHAPE of individual calcifications* (y_1) is marked
THEN *Malignant*

There is no DBR1 rule among the expert rules, but this rule is statistically significant (0.01, F-criterion). Rule DBR1 should be tested by the radiologist and included in the diagnostic knowledge base after his verification. The same verification procedure should be done for ER1. This rule should be analyzed against the database of real cases. This analysis may lead to the conclusion that the database is not sufficient and that rule DB1 should be extracted from the extended database. Also, the radiologist can conclude that the feature set is not sufficient to incorporate rule DBR1 into to his knowledge base. This kind of analysis is not possible for linear discriminant analysis or neural networks. We also use fuzzy logic to clarify the meaning of such concepts as "total number of calcifications (w_3) is large."

We tested the reliability of the expert radiologist against 30 actual cases. He classified these cases into three categories:

1) "High probability of cancer, biopsy is necessary" (or CB).

2) "Low probability of cancer, probably benign but biopsy/short term follow-up is necessary" (or BB).

3) "Benign, biopsy is not necessary" (or BO).

These cases were selected from screening cases recalled for magnification views of calcifications. For the CB and BB cases, pathology reports of biopsies confirmed the diagnosis, while a two-year follow-up was used to confirm the benign status of BO.

The expert's diagnosis was in full agreement with his extracted diagnostic rules for 18 cases, and for 12 cases he asked for more information than that given in the extracted rule. When he was interviewed, he answered that he had cases with the same combination of 11 features but with different diagnosis. This suggests that we need to extend the feature set and the rule set to adequately cover complicated cases. Restoration of monotone Boolean functions allowed us to identify this need. This is one of the useful outputs from these functions.

We extracted from the database the following rule (DBR2):

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*. This rule was confirmed by the database of 156 actual cases using the round-robin test. We extracted from this database all cases for which this rule is applicable; i.e., cases where the variation in *SIZE of calcifications* is moderate; variation in *SHAPE of calcifications* is mild; and *IRREGULARITY in shape of calcifications* is mild. For 92.86% of these cases, the rule is accurate. The expert also had a rule with these premises, but the expert rule included two extra premises:

ductal orientation is not present and there are no associated findings [see Eq. (6)]. This suggests that the database should be extended to determine which rule is correct.

Radiologists Comments Regarding 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 I*RREGULARITY 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 I*RREGULARITY 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.

Discussion and Concluding Remarks

The study has demonstrated how consistent data mining in medical diagnosis can create a set of logical diagnostic rules for computer-aided diagnostic systems. Consistency avoids contradiction among rules generated using data mining software, rules used by an experienced radiologist, and a database of pathologically confirmed cases. We identified major problems: to find contradiction between diagnostic rules and to eliminate contradiction. We applied two complimentary intelligent technologies for extraction of rules and recognition of their contradictions. The first technique is based on discovering statistically significant logical diagnostic rules. The second technique is based on the restoration of a monotone Boolean function to generate a minimal dynamic sequence of questions to a medical expert. The results of this mutual verification of expert and data-driven rules demonstrate feasibility of the approach for designing consistent computer-aided diagnostic systems.

Boris Kovalerchuk, is a professor of computer science at Central Washington University, USA. Previously, he was a visiting scholar at several universities in the US and Europe (State University of New York; Louisiana State University; Linz University, Austria, and others). Dr. Kovalerchuk earned his M.S. in mathematics from Novosibirsk University, Russia, and his Ph.D. degree from the Soviet Academy of Sciences. Dr. Kovalerchuk is a member of the New York Academy of Sciences, INFORMS, the International Association of Fuzzy Systems, and the Society for Computer Applications in Radiology. He has published more than 60 papers on artificial intelligence and information technology and has been awarded several international and US research grants. Dr. Kovalerchuk is a member of the program committees of a number of international conferences, has lectured around the world, and has been successful in the application of computer technology to medical decision-making problems in collaboration with Drs. Vityaev and Ruiz.

Evgenii Vityaev is *a* Senior Scientist at the Institute of Mathematics of the Russian Academy of Science. Previously, he was a visiting scholar at several universities in the US and Great Britain. Dr. Vityaev earned his M.S. in mathematics from Novosibirsk University, Russia, and his Ph.D. degree from Soviet Academy of Sciences. Dr. Vityaev is a member of several scientific societies. He has published more than 50 papers on artificial intelligence and information technology and he has been awarded several international and Russian research grants. Dr. Vityaev is a member of the program committees of a number of international conferences and has been successful in the application of computer technology to medical decision-making problems in collaboration with Drs. Kovalerchuk and Ruiz.

James F. Ruiz is a staff radiologist at the Woman's Hospital in Baton Rouge, LA. Dr. Ruiz has a B.S. in biochemistry from Louisiana State University and earned his M.D. degree at Tulane University. Dr. Ruiz is a member of the American College of Radiology, the Radiological Society of North America, the American Roentgen Ray Society, and the American Institute of Ultrasound in Medicine. He has published more than 20 papers on radiology and use of artificial intelligence in radiology. Dr. Ruiz has been successful in the application of computer technology to various medical decision-making problems in collaboration with Drs. Kovalerchuk and Vityaev.

Address for Correspondence: Dr. Boris Kovalerchuk, Department of Computer Science, Central Washington University, Ellensburg, WA 98926-7520. Tel: +1 509 963 1438. Fax: +1 509 963 1449. E-mail: borisk@tahoma.cwu.edu.

References

1. **Wingo PA, Tong T, and Bolden S:** Cancer statistics. *Ca - A Cancer Journal for Clinicians* 45(1): 8-30, 1995.

2. **Elmore J, Wells M, Carol M, Lee H, Howard D, and Feinstein A:** Variability in radiologists' interpretation of mammograms. *New England J Med* 331(22): 1493-1449, 1994.

3. **Shtern F:** Novel digital technologies for improved control of breast cancer. In: *CAR'96, Computer Assisted Radiology, Proc Int Symp Computer and Communication Systems for Image Guided Diagnosis and Therapy*, pp. 357-361, 1996.

4. **Kilcoyne RF, Lear JL, and Rowberg AH (Eds):** Computer applications to assist radiology. In: *SCAR'96, Proc Symp Computer Applications in Radiology*, Carlsbad, CA, 1996. **(PAGE NUMBERS?)**

5. **AUTHORS?** SCAR'98, Proceedings of the symposium for computer applications in radiology. *J Digital Imaging* 11(3), Suppl., 1998.

6. **AUTHORS?** *Proc 3rd Int Workshop on Digital Mammography*. Chicago, IL, June 9-12, 1996.

7. **AUTHORS?** *Proc 4th Int Workshop on Digital Mammography*, Nijmegen, Netherlands, June 7-10, 1998. http://www.azn.nl/rrng/xray/digmam/iwdm98/Abstracts/node51.html

8. **Lemke HU, Vannier MW, Inamura K, and Farman AG (Eds.):** *CAR'96, Computer Assisted Radiology, Proc Int Symp Computer and Communication Systems for Image Guided Diagnosis and Therapy*. Paris, France, June 26-29, 1996.

9. BI-RADS, Breast Imaging Reporting and Data System. American College of Radiology, Reston, VA, 1998.

10. **Dhar V and Stein R:** *Intelligent Decision Support Methods*. Englewood Cliffs, NJ: Prentice Hall, 1997.

11. **Kovalerchuk B and Talianski V:** Comparison of empirical and computed fuzzy values of conjunction. *Fuzzy Sets and Systems* 46: 49-53, 1992.

12. **Kovalerchuk B, Triantaphyllou E, and 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, pp.191-196, 1996.

13. **Kovalerchuk B, Vityaev E, and Ruiz JF:** Design of consistent system for radiologists to support breast cancer diagnosis. In *Proc Joint Conf Information Sciences*, Durham, NC, 2: 118-121, 1997.

14. **Kovalerchuk B, Triantaphyllou E, Ruiz J, and Clayton J:** Fuzzy logic in computer-aided breast cancer diagnosis. *Analysis of Lobulation, Artificial Intelligence in Medicine*, **(VOLUME?)(11): 75-85, 1997.**

15. **Kovalerchuk B, Conner N, Ruiz J, and Clayton J:** Fuzzy logic for formalization of breast imaging lexicon and feature extraction. In: *Proc 4th Int Workshop on Digital Mammography*, Nijmegen, Netherlands, June 7-10, 1998, http://www.azn.nl/rrng/xray/digmam/iwdm98/Abstracts/node51.html

16. **Kovalerchuk B, Ruiz JF, Vityaev E, and Fisher S:** Prototype Internet consultation system for radiologists. *J Digital Imaging* 11(3): 22-26, Suppl., 1998.

17. **Vityaev EE:** Semantic approach to knowledge base development: Semantic probabilistic inference. *Computer Systems* 146: 19-49, Novosibirsk, 1992 (in Russian).

18. **Vityaev EE and Moskvitin AA:** Introduction to discovery theory: Discovery software system. *Computational Systems* 148: 117-163, Novosibirsk, 1993 (in Russian).

19. **Mitchell T:** *Machine Learning*. New York: McGraw Hill, 1997.

20. **Abu-Mostafa (INITIAL?):** Learning from hints in neural networks. *J Complexity* 6: 192-198, 1990.

21. **Russel S and Norvig P:** *Artificial Intelligence. A Modern Approach*. Englewood Cliffs, NJ: Prentice Hall, 1995.

22. **Halpern JY:** An analysis of first-order logic of probability. *Artificial Intelligence* 46: 311-350, 1990.

23. **Krantz DH, Luce RD, Suppes P, and Tversky A:** *Foundations of Measurement.* Volumes 1-3. New York Academic, 1971, 1989, 1990.

24. **Hansel G:** Sur le nombre des fonctions Boolenes monotones den variables. *C.R. Acad. Sci. Paris* 262 (20):1088-1090, 1966 (in French).

25. **Gurney J:** Neural networks at the crossroads: Caution ahead. *Radiology* 193(1): 27-28, 1994.

1. Clustered calcifications produced by breast cancer. Calcifications display irregular contours and vary in size and shape.

2. Low-density, ill-defined mass and associated calcifications.
3. Carcinoma producing mass with spiculated margins and associated benign calcifications.
4. Flow diagram for MMDR: Steps and technique applied.
5. Task decomposition.
6. Major steps for extraction of expert diagnostic rules.
7. Interactive restoration of each function in the hierarchy.
8. Performance of methods (round-robin test).

CALL-OUTS

The major problem here is discovering sufficient, complete, and comparable sets of expert rules and data-driven rules.

About 60 to 70% of the time taken to develop rule-based systems is spent on knowledge acquisition.

A rule-based system approach exhibits the "feel" of an expert and can explain and justify a conclusion.

It is practically impossible to ask a radiologist to generate diagnoses for thousands of possible cases.

The results demonstrate feasibility of the approach for designing consistent computer-aided diagnostic systems.

.

Table 1. Dynamic Sequence of Interview of an Expert							
Case	f_1 biopsy	f_2 Cancer	ψ shape and density of calcification	Monotone extension		Chain #	Case #
				$1 \rightarrow 1$	$0 \rightarrow 0$		
1	2	3	4	5	6	7	8
(01100)	1*	1*	1*	1.2;6.3;7.3	7.1;8.1	Chain 1	1.1
(11100)	1	1	1	6.4;7.4	5.1;3.1		1.2
(01010)	1*	0*	1*	2.2;6.3;8.3	6.1;8.1	Chain 2	2.1
(11010)	1	1*	1	6.4;8.4	3.1;6.1		2.2
(11000)	1*	1*	1*	3.2	8.1;9.1	Chain 3	3.1
(11001)	1	1	1	7.4;8.4	8.2;9.2		3.2
(10010)	1*	0*	1*	4.2;9.3	6.1;9.1	Chain 4	4.1
(10110)	1	1*	1	6.4;9.4	6.2;5.1		4.2
(10100)	1*	1*	1*	5.2	7.1;9.1	Chain 5	5.1
(10101)	1	1	1	7.4;9.4	7.2;9.2		5.2
(00010)	0*	0	0*	6.2;10.3	10.1	Chain 6	6.1
(00110)	1*	1*	0*	6.3;10.4	7.1		6.2
(01110)	1	1	1	6.4;10.5			6.3
(11110)	1	1	1	10.6			6.4
(00100)	1*	1*	0*	7.2;10.4	10.1	Chain 7	7.1
(00101)	1	1	0*	7.3;10.4	10.2		7.2
(01101)	1	1	1*	7.4;10.5	8.2;10.2		7.3
(11101)	1	1	1	5.6			7.4
(01000)	0*	0	1*	8.2	10.1	Chain 8	8.1
(01001)	1*	1*	1	8.3	10.2		8.2
(01011)	1	1	1	8.4	10.3		8.3
(11011)	1	1	1	10.6	9.3		8.4
(10000)	0*	0	1*	9.2	10.1	Chain 9	9.1
(10001)	1*	1*	1	9.3	10.2		9.2
(10011)	1	1	1	9.4	10.3		9.3
(10111)	1	1	1	10.6	10.4		9.4
(00000)	0	0	0	10.2		Chain 10	10.1
(00001)	1*	0*	0	10.3			10.2
(00011)	1	1*	0	10.4			10.3
(00111)	1	1	1	10.5			10.4
(01111)	1	1	1	10.6			10.5
(11111)	1	1	1				10.6
Total Calls	13	13	12				

Table 2. Examples of Extracted Diagnostic Rules						
Diagnostic Rule	**F-criterion for Features**		**Total Significance of F-criterion**			**Accuracy of Diagnosis for Test Cases (%)**
			0.01	**0.05**	**0.1**	
IF *NUMber of calcifications per* cm^2 is between 10 and 20 AND *VOLume* > 5 cm^3 THEN *Malignant*	NUM VOL	0.0029 0.0040	+ +	+ +	+ +	93.3
IF *TOTal number of calcifications >30* AND *VOLume* > 5 cm^3 AND *DENSITY of calcifications* is moderate THEN *Malignant*	TOT VOL DEN	0.0229 0.0124 0.0325	- - -	+ + +	+ + +	100.0
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*	VAR NUM IRR	0.0044 0.0039 0.0254	+ + -	+ + +	+ + +	100.0
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*	SIZE SHAPE IRR	0.0150 0.0114 0.0878	- - -	+ + -	+ + +	92.86

xx