

Chapter 4

Automated Diagnosis of Breast Cancer on Medical Images

Marina Velikova, Inês Dutra and Elizabeth S. Burnside

Abstract The development and use of computerized decision-support systems in the domain of breast cancer has the potential to facilitate the early detection of disease as well as spare healthy women unnecessary interventions. Despite encouraging trends, there is much room for improvement in the capabilities of such systems to further alleviate the burden of breast cancer. One of the main challenges that current systems face is integrating and translating multi-scale variables like patient risk factors and imaging features into complex management recommendations that would supplement and/or generalize similar activities provided by subspecialty-trained clinicians currently. In this chapter, we discuss the main types of knowledge—object-attribute, spatial, temporal and hierarchical—present in the domain of breast image analysis and their formal representation using two popular techniques from artificial intelligence—Bayesian networks and first-order logic. In particular, we demonstrate (i) the explicit representation of uncertain relationships between low-level image features and high-level image findings (e.g., mass, microcalcifications) by probability distributions in Bayesian networks, and (ii) the expressive power of logic to generally represent the dynamic number of objects in the domain. By concrete examples with patient data we show the practical application of both formalisms and their potential for use in decision-support systems.

4.1 Introduction

According to the American Cancer Society (ACS), breast cancer is the second leading cause of cancer death in women, exceeded only by lung cancer. The chance that breast cancer will be responsible for a woman's death is about 3%. Death rates from breast cancer have been declining since about 1990, with larger decreases in women younger than 50. These decreases are believed to be the result of earlier detection through screening and increased awareness, as well as improved treatment, changes in clinical procedures, for example, genetic testing, and innovation in technologies like digital mammography and tomosynthesis [7, 11]. The increased use of computerized

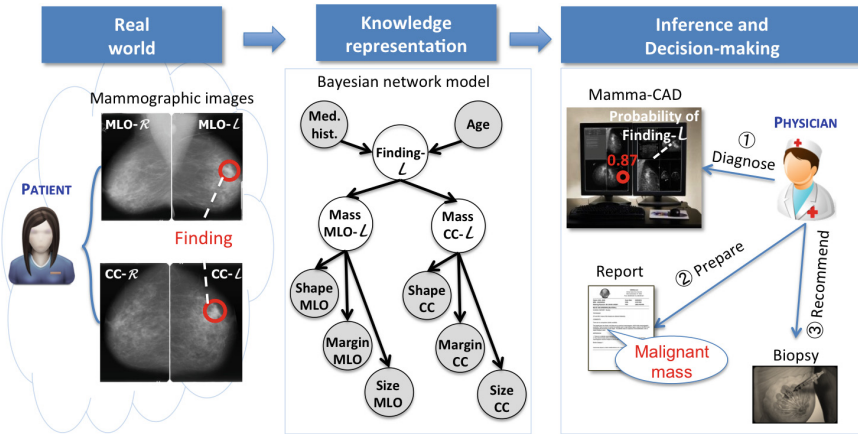


Fig. 4.1 Knowledge representation for decision-support in breast cancer diagnosis.

decision support systems that can detect breast cancer based on breast images or on the patient's history and clinical information, has the potential to contribute to improved outcomes [3, 5]. The severe consequences of breast cancer for many patients' health and life, and for their families well-being are still present and much room for improvement in the management of the disease is needed.

In this respect, a number of major challenges for clinical practitioners can be outlined, such as processing of huge amounts of data (e.g., interpretation of medical images) in short time, uncertainty in establishing a diagnosis or a treatment due to the variety of breast cancer pathologies. Another important problem is the lack of standardization and organization of what information to collect, which may be confusing and create delay in the diagnosis of diseases. This mostly concerns recording results in free text dictations, use of different terms for the same concepts and use of different metrics for the same values. Fortunately, and unique as compared to other medical fields, breast imaging has its own lexicon created by the American College of Radiology, the Breast Imaging Reporting and Data System (BI-RADS) [1], to facilitate the organization and standardization of information gathered. While this lexicon provides a good basis, it is not sufficient to support fully the management process of breast cancer.

Computer-based systems mitigate these problems by (1) efficiently organizing patient information, (2) preventing and eliminating errors and data inconsistencies; (3) extracting reliable statistics and non-trivial knowledge from the data, and (4) supporting clinical decision. Figure 4.1 presents a general scheme for such computerized support for the detection and diagnosis of breast cancer, where the knowledge about the parallel interpretation of two breast image views is represented once and in a consistent manner by means of a Bayesian network (probabilistic graphical model), and it is embedded into a computer-aided system for multiple use by a clinician.

In this work we assume that the data was entered correctly, is consistent, and is stored in some structured format. Our goal is to represent the “knowledge” attached to those data, which implies encoding not only primitive data (objects, attributes and their values), but also their relationships (causal, uncertain) that can convey useful information about the patient health conditions. In order to have knowledge with a good quality it is important to choose a good representation. In this chapter we will focus on logical and probabilistic knowledge representations.

4.2 The Domain of Breast Cancer

Breast cancer is a type of cancer originating from breast tissue, most commonly from the inner lining of milk ducts (ductal carcinomas) or the lobules (lobular carcinomas) that supply the ducts with milk. Any lump, abnormality, or alteration in the breast tissue’s integrity that may represent a breast cancer can be designated as a *finding*.

Figure 4.2 depicts the main tasks related to the identification and management of a finding, and the common methods used to perform them. The first task is called detection, which includes the identification of a finding as a physical object and its characterization (e.g., size, shape, density, and location). This is mostly done by a physical examination (e.g., palpation either by a woman herself or by a doctor) or by means of breast cancer (usually imaging) screening. The latter is performed regularly in asymptomatic women above certain age (usually between 40–50) to detect cancer at early stages and it is currently based on mammographic examinations. Such examination involves an X-ray of each breast—a *mammogram*—which is taken while carefully compressing the breast. On a mammogram, small changes in the breast tissue can be detected, which may indicate cancer that is too small to be felt. Mammograms are usually taken in two views: (1) mediolateral oblique (MLO), taken under 45° angle and showing part of the pectoral muscles, and (2) craniocaudal (CC), taken head to toe. Two main types of mammographic findings are distinguished: microcalcifications and masses. *Microcalcifications* are tiny deposits of calcium and are associated with extra cell activity in breast tissue. Microcalcifications that are scattered throughout the mammary gland are usually a non-cancerous sign, while their occurrence in clusters might indicate early stage breast cancer. According to the BI-RADS definition, “a *mass* is a space occupying lesion seen in two different projections.” When visible in only one projection, it is referred as a mammographic “asymmetry”. However, asymmetry may be a mass, perhaps obscured by overlying glandular tissue on the other view, and if it is characterised by enough suspicious features then it may indicate breast cancer.

Based on the detection results of a finding, the physician may or may not request additional exams, for example, fine needle aspiration (FNA) or core needle biopsy (CNB) in order to perform the second task—diagnosis. It concerns the identification of a finding either as *benign (non-cancerous)* or as *malignant (cancerous)*. In benign tumors, the cells will not invade surrounding tissues or spread to distant organs. In most

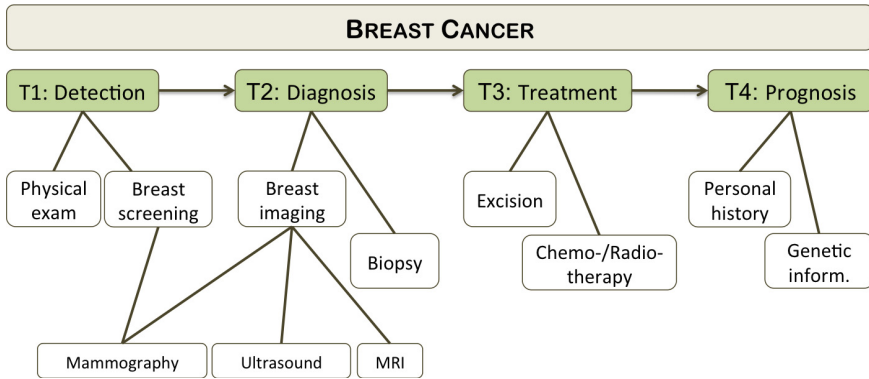


Fig. 4.2 Tasks (T#) and common methods involved in the management of breast cancer.

cases, a benign tumor can be removed. In a malignant tumor, the cells have the potential to behave aggressively, invading adjacent tissue and spreading to distant organs.

If the diagnosis is a malignant finding, the next task is to recommend and perform a treatment such as chemo-/radiotherapy or an excision surgery. Finally, the physician can study the effects of the treatment, and perform a prognostic analysis for cancer recurrence and chances for survival of the patient, by using, for example, genetic information or the patient's history.

Therefore, in this domain, we can count on information about the patient (demographics, personal history, family history, social information, and environmental exposures), about mammography images and reports, descriptors of abnormalities associated with a mammography, pathology information (details of histological analysis such as kind of breast cancer or cells associated with calcifications), and details about surgeries (kind of biopsy procedure, kind of needle, number of specimens collected etc.).

4.3 Knowledge Representation for Breast Cancer Diagnosis

4.3.1 Motivation

The information concerning the breast cancer diagnosis can come from various sources, e.g., image modalities, laboratory tests, and different medical experts, e.g., radiologists, surgeons, pathologists. As a result, we end up with heterogeneous type of information about the same patient that need to be represented and processed in a relational form as opposite to the traditional propositional approach that uses a single table to collect all information about a patient. One of the forms of representing relational data is to store it in relational databases. These databases allow only for querying the primitive (basic) data itself, and do not support queries about more

Human-defined knowledge

Probability of upgrade increases for stereotactic biopsy if the finding is microcalcifications and biopsy results reveal atypical ductal hyperplasia (ADH).

First-order logic

upgrade(A): IF
biopsyResults(A,ADH) AND
calcifications(A,a,f) AND
biopsyProcedure(A,stereo).

Bayesian network

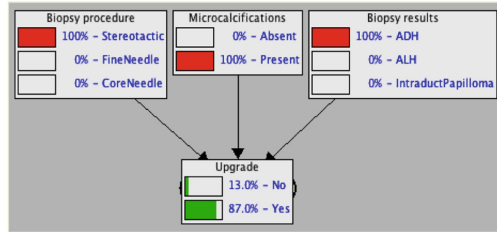


Fig. 4.3 Human-defined knowledge and its representation by FOL and a BN.

complex relationships such as “What is the relation between a malignant diagnosis and a combination of some patient attributes”, “What is the disease evolution along the time and what is the prognosis of a given patient?”, or even “What is the pattern of a discordant biopsy (the one that gave a result that is not agreeable by all physicians in a medical conference)?” From a clinical point of view, giving answers to these questions means to save patients from the inconvenience of undergoing invasive procedures and save other patients of being sent home without an adequate treatment, while reducing costs to patients and to hospitals.

To be able to answer such questions, more advanced approaches need to be used to represent relational knowledge. In this chapter we will focus on first order logic and graphical probabilistic models. To illustrate the basic knowledge representation principles of these methods, Fig. 4.3 presents an example in the domain of breast cancer. In the left-hand side, we have a first order logic (FOL) definition for an upgraded biopsy (an upgraded biopsy is the one that gave a negative result for malignancy, but proved to be malignant after recommended surgery). In the right-hand side we have a graphical probabilistic representation in the form of a Bayesian network (BN). Both representations make use of the attributes *A* related to an object *Biopsy* to build a relation among attributes. The first-order logic relates atypical ductal hyperplasia (ADH), microcalcifications (amorphous and fine-linear), and biopsy procedure to infer cases when the biopsy is an upgrade. The same information is represented in the Bayesian network, but in another format and uses additional probabilistic information.

To design a knowledge representation system, we need to identify the types of knowledge that exists in the domain of interest—the diagnosis of breast cancer. We distinguish between two categories of knowledge: (i) knowledge about primitive data, which are objects and attributes and (ii) knowledge about relationships between the primitive data.

4.3.2 Object-Attribute Knowledge

Table 4.1 presents examples of physical objects O relevant for our domain of interest, their attributes (features) A with the respective range of values $dom(A)$.

We distinguish between two main types of attribute domains:

- (i) *discrete* referring to a finite and countable set of values. It can be defined by categories or integers, e.g., the domain of age can be defined as the categorical set of “young”, “middle-aged”, “old” or as the integer set $\{1, \dots, 120\}$. Typical examples for discrete variables within the medical domain are risk factors such as gender $\in \{male, female\}$, history of a disease $\in \{no, yes\}$, and smoking (cigarettes per day) $\in \{0, 1 - 5, 6 - 20, > 20\}$.
- (ii) *continuous* referring to an infinite set of values between two points. Thus the domain is real-valued and values follow a distribution, e.g., Gaussian or Gamma. Typical examples of continuous attributes are the image features extracted by a computer-aided system or the size of a finding. From a knowledge representation point of view, continuous attributes are often *discretized*, i.e., their range is divided into a finite set of values that may or may not have a semantical meaning, but allow for an easier interpretation for human experts. For example, the size of a finding can be discretized into $\{<1\text{ cm}, 1-3\text{ cm}, >3\text{ cm}\}$ or $\{small, medium, large\}$. A recent work on discretization of mammographic features has shown the advantages of this data pre-processing method for improving the detection performance of a CAD system [8].

More than one value can be assigned to some of the variables in breast cancer. For example, both values “fine” and “linear” can be assigned to the calcifications variable, or more than one pathology may be associated with a tumour. In that case, physicians may use a precedence list that indicates orders like Fine > Linear > ...

Table 4.1 Examples of objects, attributes, and their values in the domain of breast cancer

Object O	Attribute $A \in dom(A)$
Patient	age $\in \{\text{“young”}, \text{“middle-aged”}, \text{“old”}\}$
	gender $\in \{\text{“male”}, \text{“female”}\}$
Exam	time $\in \{\text{“prior”}, \text{“current”}\}$
	type $\in \{\text{“physical”}, \text{“screening”}\}$
Image	quality $\in \{\text{“low”}, \text{“medium”}, \text{“high”}\}$
	modality $\in \{\text{“MRI”}, \text{“mammography”}, \text{“CT”}\}$
	breast-view $\in \{\text{“MLO”}, \text{“CC”}\}$
Finding	location-in-breast (quadrant) $\in \{\text{“upper-outer”}, \text{“upper-inner”}, \dots\}$
	location (side) $\in \{\text{“left”}, \text{“right”}\}$
Mass	shape $\in \{\text{“oval”}, \text{“round”}, \text{“irregular”}\}$
	margin $\in \{\text{“circumscribed”}, \text{“indistinct”}, \text{“spiculated”}\}$

or $ALH < LCIS < ADH < DCIS$. This information can be used to give preferences to certain attribute values ranking their relevance.

Coding medical object-attribute knowledge is straightforward once there is an established convention for naming variables and terms, such as the BI-RADS lexicon for mammographic features and findings. For example, the shape (attribute) of a mass (object) using first-order logic (FOL) can be represented by the two-valued predicate $massShape(F, Value)$, where F is a variable referring to a mass object and $Value$ is a variable referring to one of the attribute values (see Table 4.1). In terms of probabilistic graphical models, such as Bayesian networks (BNs), the same knowledge is to be represented by a node called “massShape” whose domain will contain the three exclusive values describing shape.

Another way of coding the same attribute “massShape” is to use a boolean representation where a new attribute is created for each possible value of the original “massShape”. Therefore, if “massShape” could assume values “oval,” “round,” or “irregular,” the new representation would be done through three new variables, say, “massShapeOval,” “massShapeRound,” and “massShapeIrregular” with boolean values (for example, value 1 indicating presence and value 0 indicating absence). This kind of representation can be very useful when one attribute can assume several possible values or if the data is to be used for classification, as some classifiers work better with binary feature vectors. It is also helpful to improve the quality of data as each possible value of the variable will be properly discriminated. For example, assume the variable we have is “massShape.” If this variable is left blank for any reason, we can not conclude anything about “oval,” “round,” or “irregular.” On the other hand, if we represent this same variable by three new variables, chances are that at least one of them will not be left blank.

4.3.3 Relational Knowledge

Relational representations can be conceptualised as a binding between a relation symbol and a set of ordered tuples of elements. For example, the relation-symbol *larger* is bound to the set of ordered pairs: $\{(5, 2), (3, 1)\dots\}$. The symbol represents the “intension” of a relation and specifies which relation is intended; for example, elements are ordered by size. The ordered tuples represent the “extension” of a relation. They can include knowledge learned by experience, and can provide statistical knowledge of the world [4].

4.3.3.1 Causality

While object-attribute relationships are relatively straightforward to represent given a standardized naming, the relationships between the objects in the domain of interest may be more complex to formally express. One type of relationship concerns

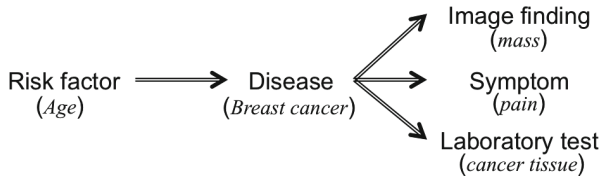


Fig. 4.4 Example of causal dependencies

causal dependencies. A typical example of such dependencies in a medical domain, including the breast cancer diagnosis, is presented in Fig. 4.4:

The concept on the left-hand side of each arc represents the cause whereas the concept on the right-hand side is the effect. While these causal arcs reflect the direction of influence, they do not necessarily express a deterministic dependence. In other words, the presence of a risk factor (elderly woman) increases the *chance* that a disease (breast cancer) may occur, but it does not imply that it will occur for sure. The same holds for the presence of a disease and its appearance on an image—breast cancer may or may not appear as a mammographic mass, for example. Clearly such relationships are inherent with uncertainty and they can be represented by probabilistic approaches such as Bayesian networks, where the network structure reflects exactly the direction of causality, and the probability distributions represent its strength. Certain causal relationships such as “Disease” \rightarrow “Laboratory tests” may be more probable and even in some cases deterministic, as in the example shown in Fig. 4.4, which can be expressed by the FOL rules.

Another type of relational knowledge that is more challenging to represent, especially in image interpretation, concerns aggregations such as the “part-whole” relations. A common assumption in this case is that given evidence about parts, the goal is to hypothesise and try to draw conclusions about the whole. In particular, evidence for certain characteristics in one or more parts increases the likelihood that the same characteristics are present in the whole. This type of relationship is illustrated in Fig. 4.5 where various levels of object image analysis are given, namely an image is “part-of” an exam, and the exam is “part-of” a patient case. Detecting cancer on the image will imply that the respective exam and patient case are also assigned a label of “cancerous”. The problem of this type of reasoning is, however, that the errors in the low(part)-level image analysis will be propagated to the higher(whole)-level analysis. An alternative is to represent and reason about additional knowledge such as spatial, temporal, and hierarchical relationships to better analyse the part-whole dependencies.

4.3.3.2 Spatial Knowledge

Another key knowledge used in breast cancer diagnosis on medical images are spatial relationships that indicate the context dependency to the objects locations. There are two general forms of spatial knowledge: (i) absolute position of the objects on the

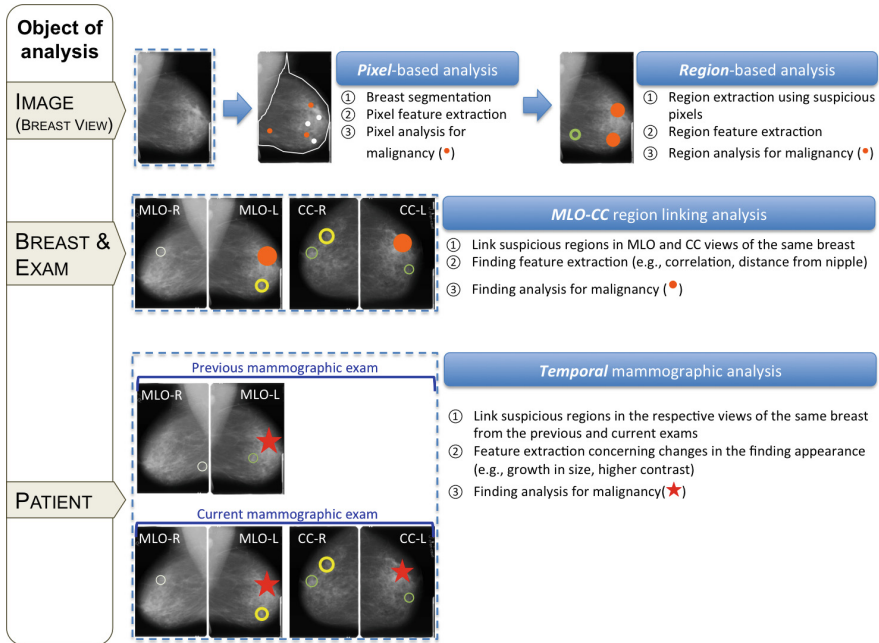


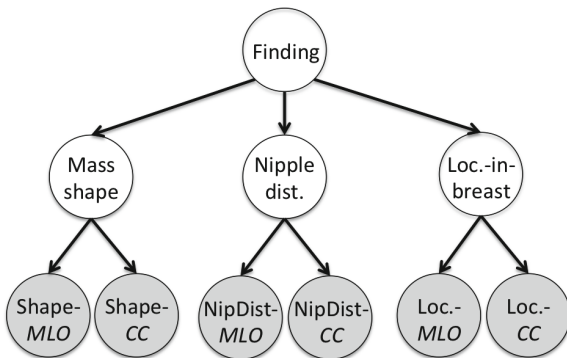
Fig. 4.5 Various levels of object image analysis by computer-aided detection systems.

image, usually in XY-coordinate system for 2D images, and (ii) relative positions of the objects to each other.

The first type of spatial knowledge in image interpretation for breast cancer diagnosis is relatively straightforward to represent. Let us consider a finding detected by a CAD system or a human reader in the MLO view of the left breast. The location of this finding will be represented by a node for each coordinate, e.g., “LocX-MLO” in BNs, and by a binary predicate, e.g., $locX_MLO(F, Value)$ in FOL with F referring to the finding and $Value$ to the X-location value. Depending on the available data, the range of values that location can take will be (i) continuous: obtained from the automated processing of the MLO image or (ii) discrete: based on a manual annotation (e.g., breast quadrant) or discretization of the continuous values.

The relation of objects in terms of space requires a more complex, and not necessary unique, representation. In mammographic analysis, it is well-known that two regions of interest (or findings) on MLO and CC views of the same breast that are approximately at the same distance from the nipple and exhibit similar features (e.g., mass shape is the same) are *very likely* to refer to one finding. In FOL, this knowledge concerning the findings F_1 and F_2 can be expressed as follows:

Fig. 4.6 A BN representing the linking between two findings on the MLO and CC views of the same breast. The grey circles represent the observed features of the findings on both views.



$$\begin{aligned}
 \text{same_finding}(F_1, F_2) \leftarrow & \text{MLOView}(F_1) \wedge \text{CCView}(F_2) \wedge \\
 & \text{nipple_distance}(F_1, D_1) \wedge \text{nipple_distance}(F_2, D_2) \wedge \\
 & (\text{abs}(D_1 - D_2) < \epsilon) \wedge \\
 & \text{side}(F_1, \text{left}) \wedge \text{side}(F_2, \text{left}) \wedge \\
 & \text{quadrant}(F_1, \text{upper_outer}) \wedge \text{quadrant}(F_2, \text{upper_outer}) \wedge \\
 & \text{massShape}(F_1, \text{oval}) \wedge \text{massShape}(F_2, \text{oval}).
 \end{aligned}$$

The problem with the representation above is that it is deterministic and it does not reflect a likelihood that F_1 and F_2 are the same finding. To do so, we can use a BN with probabilistic information as shown in Fig. 4.6.

The lowest network level captures the observed features O_i of an image finding on each breast view, modeled as effects of the unobserved finding features X_j (white circles). The top level node corresponds to finding F with values “no”, “benign,” and “malignant”. The conditional probability tables $P(O_i|X_j)$ and $P(X_j|F)$ can be obtained based on expert knowledge or statistics derived from image data. These can be expressed as qualitative or quantitative constraints as shown in Table 4.2.

Table 4.2 Probabilistic qualitative constraints and quantities

Probabilistic qualitative constraints
$P(\text{MassShape} = \text{'oval'} \text{Finding} = \text{'benign'}) >$
$P(\text{MassShape} = \text{'oval'} \text{Finding} = \text{'malignant'}) >$
$P(\text{MassShape} = \text{'oval'} \text{Finding} = \text{'no'})$
$P(\text{nipple_distance} = \text{'0 - 2cm'} \text{Finding} = \text{'benign'}) >$
$P(\text{nipple_distance} = \text{'0 - 2cm'} \text{Finding} = \text{'malignant'})$
Probabilities
$P(\text{MassShape} = \text{'oval'} \text{Finding} = \text{'benign'}) = 0.73$
$P(\text{nipple_distance} = \text{'0 - 2cm'} \text{Finding} = \text{'malignant'}) = 0.24$

4.3.3.3 Temporal Knowledge

Temporal knowledge implies a dependence to time and may lead to different inferences in different temporal contexts. In medical domain, including breast cancer diagnosis, modelling and reasoning about such knowledge is of particular importance due to a progressive nature of a disease. In breast screening programs, for example, it is typical that images of the same breast are taken over regular intervals of time. Detecting interesting changes amounts to recognising corresponding objects, if present, in these images.

We used the examples of mammographic patient data from Table 4.3 to illustrate knowledge representation principles of temporal knowledge using graphical models and logic. Table 4.3 contains observational data such as the column “Calc F/L” reporting if a radiologist saw fine or linear calcifications in the mammogram image, and the column “Location” reporting the quadrant in the breast image related to the finding.

Table 4.3 includes two interesting relations for patient P1, who has three mammographic exams. The first and the second exams seem to reveal the same finding, given the common location in the breast, and observed at different periods of time (5/02 and 5/04). This finding refers to a tumor that appears on the mammogram as a mass that has grown in size in the second examination and as newly observed microcalcifications—clearly signs for malignancy. At the same time, another tumor was found in patient P1 during the examination made in 5/04, which appears to be benign.

In terms of probabilistic graphical models, a common representation method of temporal knowledge are *dynamic Bayesian networks*—temporal models where the same variables of interest, describing both the state of the system, observables, conditions, and actions that may change the state at different points of time [6]. A usual assumption underlying these models is that: (i) the future state is conditionally independent of the past state given the present state (first-order Markov property), and (ii) the probabilistic temporal relations between adjacent states do not change over time (time invariance or stationarity condition). This way, a dynamic Bayesian network becomes a compact process representation that can be employed in forecasting.

Figure 4.7 presents a dynamic Bayesian network in the context of patient data shown in Table 4.3. We have two time slices representing, for example, mammographic exams taken over two years. Within each slice static causal relationships are represented by solid arcs whereas the temporal relationship between both slices is represented by the dashed line. The former expresses, for example, that the

Table 4.3 Examples with mammographic patient data

Patient	Month/Year	Finding	Calc F/L	Mass size	Location	Diagnosis
P1	5/02	1	No	0.03	RU4	Benign
P1	5/04	2	Yes	0.05	RU4	Malignant
P1	5/04	3	No	0.04	LL3	Benign
P2	6/00	4	No	0.02	RL2	Benign

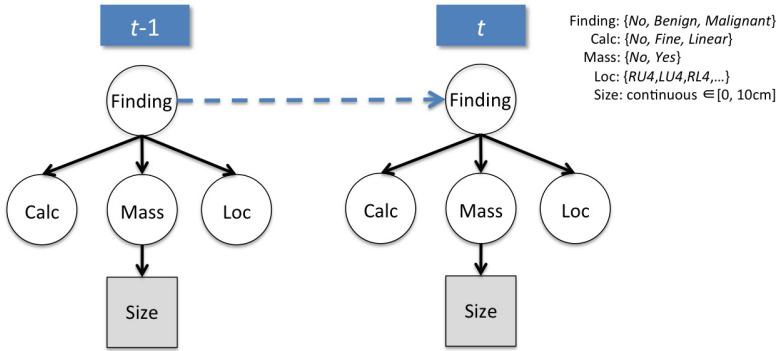


Fig. 4.7 A structure of a dynamic Bayesian network representing the relations in Table 4.3. The dashed arc represents a temporal relationship *between* two time slices whereas the solid arcs represent static relationships *within* a time slice.

presence of *Finding* is a causal factor for the presence of calcifications or a mass as well as for a location characteristic. Furthermore, *Mass* has a probabilistic influence on the distribution of the size attribute, which can be expressed, for example, as $P(\text{Size}|\text{Mass} = \text{yes}) = \mathcal{N}(0.03, 0.001)$, with \mathcal{N} denoting a normal distribution with a respective mean and standard deviation. A temporal relationship in the network expresses the fact that a finding detected in a previous time slice $t - 1$ increases the probability for a finding in the current time slice t , which is expressed by the conditional probability distribution $P(\text{Finding}_t|\text{Finding}_{t-1})$, e.g., $P(\text{Finding}_t = \text{benign}|\text{Finding}_{t-1} = \text{benign}) = 0.42$, and $P(\text{Finding}_t = \text{malignant}|\text{Finding}_{t-1} = \text{benign}) = 0.25$.

The relations in Table 4.3 can also be easily represented in logic as shown below, where names such as “previous_finding,” “mammo,” and “date” are regular first order logical predicates and P, F_1, F_2 are logical variables.

$$\begin{aligned} \text{previous_finding}(F_1, F_2) \leftarrow & \text{mammo}(P, F_1) \wedge \text{mammo}(P, F_2) \wedge \\ & \text{date}(F_1, D_1) \wedge \text{date}(F_2, D_2) \wedge \\ & (D_1 < D_2 \vee D_2 < D_1) \end{aligned}$$

This rule relates two findings F_1 and F_2 for the same patient P , separated in time (date of F_1 is before or after the date of F_2). It can be further used to simulate temporal reasoning in the context of other rules such as:

$$\begin{aligned} \text{is_malignant}(A) \leftarrow & \text{mass}(A, \text{present}) \wedge \text{previous_finding}(A, B) \wedge \\ & (\text{massSize}(A) < \text{massSize}(B)) \wedge \text{calc}(B, \text{present}) \wedge \\ & \text{previous_finding}(A, C) \wedge \text{calcFineLinear}(C, \text{yes}) \end{aligned}$$

In this rule, we have explicit relations among different rows of Table 4.3 with the use of the predicate *previous_finding* which relates finding A with finding B, each one having its own properties. This rule also relates finding A with a third finding C (not shown in Table 4.3), which has calcification fine-linear.

4.3.3.4 Hierarchies and Concept Aggregation

Up to now we discussed ways for representing mostly low-level image interpretation information, which concerns findings, manual annotations, and their features. Although this forms the basic step for automated decision-support in breast cancer diagnosis, the ultimate goal is that computerized systems should be able to analyze data and provide feedback at a patient level. In particular, as physicians are capable of simultaneous interpretation of various contexts (e.g., spatial and temporal), multiple types of findings (e.g., masses, calcifications, distortions) and modalities (e.g., X-ray, MRI, ultrasound), the systems should represent and reason with various sources and levels of information and knowledge. A useful representation scheme for systematic structuring of such variety of complex relationships and facilitating physician’s reasoning is a *concept hierarchy*, where knowledge and information sources are integrated both horizontally and vertically. Such a hierarchical structure in the domain of breast cancer image diagnosis is presented in Fig. 4.8.

The horizontal integration refers to combining various sources at the same level of processing, where each source supports part of an entire task. A typical example in the context of breast cancer image diagnosis is a parallel interpretation of multiple mammographic signs, such as microcalcifications *MCAL* and masses *MASS*, to provide a complete picture whether or not breast cancer *BC* (i.e., a malignant finding)

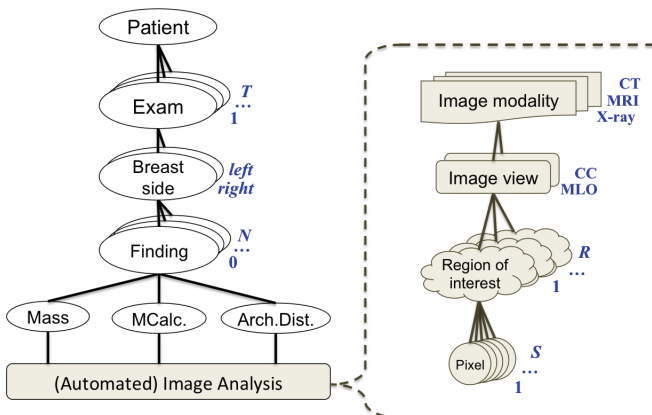


Fig. 4.8 A hierarchical structure of concepts used in breast cancer diagnosis. The left structure presents the semantical concepts as used by physicians whereas the right structure presents the top-down layers in image analysis.

Table 4.4 Aggregated concepts in establishing the risk for breast cancer with a respective example for a representation in FOL

Concepts	Example representation in FOL
Proximity	$highRisk(F) \leftarrow calcifications(F, yes) \wedge location(F, X) \wedge location_calcifications(F, Y) \wedge distance(X, Y, E) \wedge E < Error$
Quantity	$lowRisk(F) \leftarrow numberOfFoci(F, N) \wedge N = 1$
Similarity	$lowRisk(F) \leftarrow previous_finding(F, BF) \wedge date(BF, D1) \wedge date(F, D2) \wedge D1 < D2 \wedge pathology(BF, benign), similar(F, BF)$
Timing	$lowRisk(F) \leftarrow previous_finding(F, BF) \wedge date(BF, D1) \wedge date(F, D2) \wedge D1 < D2 \wedge pathology(BF, benign)$
Association	$highRisk(F) \vee excise(F) \leftarrow calcifications(F, yes)$
Priority	$incidental(P) \leftarrow pathologyPriority(P, low) \wedge pathologyType(P', ARS')$

is present. In terms of probabilistic graphical models, this integration can be expressed in two ways depending on available knowledge and data:

- $MCAL \leftarrow BC \rightarrow MASS$: This *descriptive* representation expresses the causal knowledge that microcalcifications and masses are signs (effects) of the disease “breast cancer” (cause) and given that the disease is present then it is expected to appear as a mammographic sign. The uncertainty in this appearance (e.g., obscurity in the image due to high breast density) is provided by the conditional probability tables of *MCAL* and *MASS* based on domain knowledge, e.g., $P(MCAL = 'malignant' | BC = 'present') = 0.86$ and $P(MASS = 'malignant' | BC = 'present') = 0.93$. Once a sign is observed, the probability $P(BC | MCAL, MASS)$ can be computed using the Bayes theorem.
- $MCAL \rightarrow BC \leftarrow MASS$: This *discriminative* representation aims at predicting the probability for breast cancer given the mammographic observations. When sufficient data from image processing or human annotation reports are available, one can learn the conditional probabilities $P(BC | MCAL, MASS)$, expressing the combined effect of the signs in breast cancer diagnosis.

The vertical integration in a hierarchy, on the other hand, is a knowledge representation at different levels of abstraction. An example in the current context is the parallel interpretation of multiple two-dimensional breast projections, such as *MLO* and *CC*, to provide a complete picture whether or not a finding is present in the breast *B* as a whole. Similarly to the horizontal integration, the vertical knowledge representation can be expressed in various forms based on domain knowledge or available data.

Abstract concepts can be represented to help structure the physician’s reasoning. Table 4.4 presents a number of typical concepts in establishing the risk for breast cancer.

4.3.4 Observations and Hypotheses

From a knowledge representation point of view, we distinguish between observations and hypotheses. *Observations* are factual information obtained by means of a visual (physical) inspection, reporting, tests, or computer processing. Typical examples include risk factors (age, medical history), image features (location and shape of a finding), image findings (mass, microcalcifications), symptoms (pain, palpable mass) or laboratory results (breast biopsy).

A *hypothesis* is a possible explanation for the phenomenon we observe and it is often related to a variable of interest (output). Examples include the diagnosis of a disease (e.g., breast cancer) or determining the state of organ functioning (e.g., renal dysfunction). In the knowledge representation process, hypotheses may be included as separate entities that establish dependencies between the observations. In this case, we refer to hypotheses as “hidden variables”.

Despite this hard distinction between observed and hidden variables, in practice a variable can play the role of both, depending on available information or the problem at hand. For example, in certain situations, an image finding of mass may be reported by a human reader and be used as evidence for determining whether or not breast cancer is present, whereas in another situation the goal might be to predict whether mass is present given a number of observed image features.

4.4 Inference and Decision-Making in the Management of Breast Cancer

4.4.1 Deductive Inference

A deductive system uses the data combined with pre-defined rules to draw conclusions and to support the decision-making process. For example, after the first screening, a medical doctor can lookup the guideline on Breast Screening and Diagnosis produced by the National Comprehensive Cancer Center (NCCN)¹, to assist his/her the decision-making. With the guideline, depending on the symptoms found during a screening, the physician can follow different paths suggesting a possible follow-up to a patient. A guideline implements a limited form of deduction, where, given some knowledge about a patient, the physician infers a decision based on the paths followed in that guideline. This inference of deduction can be done automatically if we use formal languages such as mathematical logic that, for example, uses complete and sound proof procedures such as resolution [9]. In fact, there are several works in the literature that represent guidelines (or parts of) by means of logics [10, 13]. The knowledge represented in Sect. 4.3, using the logic formalism, can be used to automatically answer questions such as “what are the findings that are malignant?”

¹ http://www.nccn.org/professionals/physician_gls/pdf/breast.pdf, available after registration.

(in logic: $\exists F \text{malignant}(F)$) or “Is there a benign finding with a high mass density?” (in logic: $\exists F \text{pathologyType}(F, \text{benign}) \wedge \text{massDensity}(F, \text{high})$), using resolution.

4.4.2 Inductive Inference

Inductive systems, on the other hand, support the decision-making by automating the process of creating models based on available data or expert knowledge. Systems that fall in this category are usually called machine learning systems. In the case of creating rules, a machine learning algorithm can automatically produce a guideline as defined by NCCN or the rule presented in Fig. 4.9, or even complement a guideline with a newly created rule.

The example rule shown in Fig. 4.9, written with the Prolog syntax, was automatically extracted from a database containing more than 65,000 patients. This rule suggests that a set of patients may have had a delayed treatment, because they had obtained a BI-RADS category of 3 (low-risk benign, b3) in past exams, which later became 5 (high-risk malignant, b5) [2]. In fact, this rule was validated against the dataset, and this condition held true for seven positive patients and for none of the negative ones with benign findings.

Inductive learning with logic is very useful to extract readable and interpretable models from the data. Rather than producing a black-box classifier, logical rules can explain the classifier itself to the physician. This can further contribute to the refinement of the expert knowledge in a way that the inductive system learns rules, the physician can modify or refine them, then the system learns new rules from the refinements and the process continues.

One good side-effect of inductive learning is that the rules found during this interactive process can shed some light on the most relevant primitive features that

Human-defined knowledge

If finding A was classified as BI-RADS 5, had a mass present in a patient who was between the ages of 65 and 70, had two prior mammograms B and C, and prior mammogram B had no mass shape described, had no punctate calcifications, and prior mammogram C was classified as BI-RADS 3 then finding A is malignant.

Prolog-based rule

```
is_malignant(A):-
    birads_category(A,b5), massPA0(A,present), age(A,age6570),
    previous_finding(A,B,C), massesShape(B,none),
    calc_Punctate(B,notPresent),
    previous_finding(A,C), birads_category(C,b3).
```

Fig. 4.9 An example of knowledge representation using a Prolog rule.

can suggest a diagnosis. For example, some of the features may consistently appear in every learning step. The health professionals can then concentrate on studying these features and even improving the quality of the data values entered for these features by enforcing better data collection.

4.4.3 Application

In this section we demonstrate the application of knowledge representation formalism for mammographic diagnosis. We show two different formalisms. One is based on a probabilistic graphical model and the second one is based on first order logic. In the first one, features are automatically extracted from image processing. In the second one, features come from multiple tables generated by annotations performed by doctors when preparing medical reports about mammography, pathology analysis and biopsy procedures.

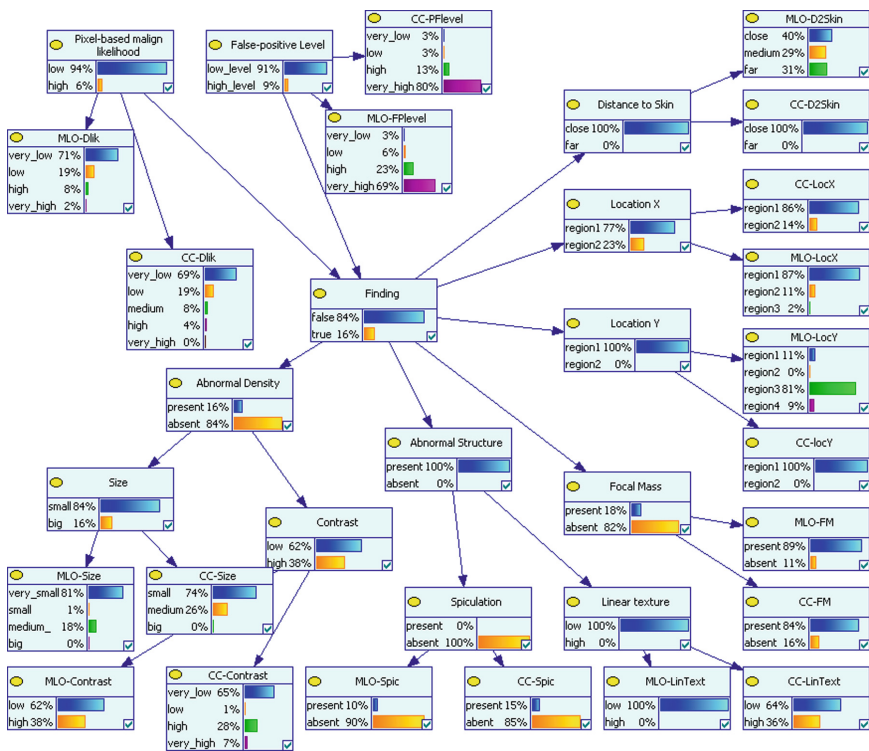


Fig. 4.10 A Bayesian network for interpretation of mammographic signs

4.4.3.1 Probabilistic Graphical Model

Figure 4.10 presents a Bayesian network model whose structure was manually built using domain knowledge and its parameters were learnt from real-world mammographic data. The model aims at detecting a malignant finding on a mammogram based on image features automatically extracted from a CAD system and following the two-view image interpretation as done by radiologists. For a more detailed description of the model, the reader is referred to [8, 12].

Table 4.5 presents the data for three real-world cases obtained from the Dutch mammographic screening program, which contains a number of automatically extracted regions of interest and their respective features on a breast view (image). The ground-truth of each region is provided by pathology reports. The last row in

Table 4.5 A sample of three real-world cases with mammographic regions of interest (ROI) and respective features extracted from a CAD system. Variable Finding is the ground-truth.

Cases (C#)	C1 (right breast)		C2 (left breast)		C3 (right breast)	
ROI#	ROI1	ROI2	ROI1	ROI2	ROI1	ROI2
Finding	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE
MLO-FPlevel	very high	low	very high	high	very high	low
CC-FPlevel	very high	very low	very high	very low	very high	very low
MLO-Dlik	very low	high	very low	very low	very low	low
CC-Dlik	very low	low	very low	high	very low	medium
MLO-Spic	present	present	present	present	absent	absent
CC-Spic	absent	present	present	present	absent	absent
MLO-FM	present	absent	present	present	present	present
CC-FM	present	absent	present	present	present	present
MLO-Size	very small	very small	medium	medium	very small	very small
CC-Size	small	small	small	medium	small	small
MLO-Contrast	low	low	low	low	low	low
CC-Contrast	very low	very low	very low	high	very low	very low
MLO-LinText	low	low	low	low	low	low
CC-LinText	low	low	low	low	low	low
MLO-D2Skin	far	medium	close	medium	medium	medium
CC-D2Skin	close	close	close	close	close	close
MLO-LocX	region1	region1	region1	region1	region1	region1
CC-LocX	region2	region1	region1	region1	region1	region1
MLO-LocY	region3	region3	region4	region3	region1	region3
CC-LocY	region1	region1	region1	region1	region1	region1
$P(\text{Finding} = T)$	0.07	0.89	0.11	0.77	0.11	0.62

Table 4.5 shows the Bayesian network (shown in Fig. 4.10) computed probability that a malignant finding is present, given the features in each view.

4.4.3.2 First Order Logic (FOL)

Another example using a logic representation is shown in Fig. 4.11, where each of the rules, automatically learned from data, is true for 30 out of 79 benign findings (with 40% Recall) while not missing any malignant finding out of 17 (with 100% precision). In other words, when these rules are used to classify new cases, a true malignant case is never missed and mistakenly sent home. On the other hand, some benign cases will may be misclassified, but not all. The dataset used to train the rules consists of non-definitive biopsies collected from the Medical School of the University of Wisconsin-Madison, USA. The relevance of this result is that the classifier is

$benign(F) \leftarrow$ $patientID(F,P) \wedge previousSurgery(P,ps_no)$ $marginSpiculated(F,no)$ $disappearanceAbn(F,N')$	\wedge The patient did not have a previous surgery, imaging did not present a spiculated mass margin, the abnormality remained in post-biopsy imaging \wedge
$benign(F) \leftarrow$ $biopsyAbn(F,Abn)$ $marginIndistinct(Abn,not_present)$ $marginSpiculated(Abn,not_present)$ $disappearanceAbn(F,N')$	\wedge Imaging did not present an indistinct mass margin, \wedge imaging did not present a spiculated mass margin, \wedge the abnormality remained in post-biopsy imaging

Fig. 4.11 FOL rules

$abnID(bio53,abn24).$ $patientID(bio53,patient771964).$ $biopsyDate(bio53,667).$ $age(bio53,68).$ $biopsySide(bio53,'R').$ $biopsyProcedure(bio53,'UsCore').$ $needleType(bio53,'vacuum-assisted').$ $needleGauge(bio53,9).$ $samplesNum(bio53,6).$ $clipInSiteBX(bio53,0).$ $disappearanceAbn(bio53,'Y').$ $outcome(bio53,'Y').$ $familyHistory(patient771964,fh_none).$ $personalHistory(patient771964,ph_no).$ $previousSurgery(patient771964,ps_yes).$ $patientID(bio53,patient771964).$ $marginIndistinct(abn24,not_present).$ $marginMicrolobulated(abn24,not_present).$ $marginCircumscribed'(abn24,not_present).$ $marginSpiculated(abn24,not_present).$ $marginObscure(abn24,not_present).$	$abnID(bio76,abn94).$ $patientID(bio76,patient60244).$ $biopsyDate(bio76,747).$ $age(bio76,63).$ $biopsySide(bio76,'L').$ $biopsyProcedure(bio76,'MriCore').$ $needleType(bio76,'vacuum-assisted').$ $needleGauge(bio76,9).$ $samplesNum(bio76,10).$ $clipInSiteBX(bio76,0).$ $disappearanceAbn(bio76,'N').$ $outcome(bio76,'N').$ $familyHistory(patient60244,fh_none).$ $personalHistory(patient60244,ph_yes).$ $previousSurgery(patient60244,ps_no).$ $patientID(bio76,patient60244).$ $marginIndistinct(abn94,not_present).$ $marginMicrolobulated(abn94,not_present).$ $marginCircumscribed'(abn94,not_present).$ $marginSpiculated(abn94,not_present).$ $marginObscure(abn94,not_present).$
--	---

Fig. 4.12 Instances represented in FOL

capable of sparing some women from excision while not missing any malignant finding. Currently, when biopsies are inconclusive (non-definitive), the common practice is to excise all women in this situation.

In order for that to work, data instances need also to be represented in FOL. Two examples are shown in Fig. 4.12. These instances are coded from medical mammography reports, and include extra information about the biopsy procedures and about the patient data. They also use the BI-RADS encoding. Applying the rules from Fig. 4.11 to the two instances, classifies correctly the left instance as malignant and the right instance as benign. We only show partial data for the instances, since the rules only describe mass margins (mammographic finding), biopsy features and patient data.

4.5 Discussion and Conclusion

We outlined various types of knowledge available in the domain of image interpretation of breast cancer diagnosis and their representation using two main formalisms from the field of artificial intelligence—Bayesian networks (BNs) and first-order logic (FOL). While both formalisms are capable of explicitly expressing domain knowledge, for example, in terms of causal, spatial and temporal relations, they differ in the form of this expression.

The power of Bayesian networks lies in their capabilities to deal in a probabilistic manner with uncertainty, which is often encountered in medical image interpretation due to, for example, image quality or resemblance in the image appearance between abnormalities and normal body structures. In the current context, we demonstrated how Bayesian networks can be used to model multi-view image interpretation by using a hierarchical representation following the human expert’s working principles.

As a propositional method, however, Bayesian networks are restricted in the representation of a dynamic number of objects and relationships, which is naturally done by FOL. In the context of breast cancer diagnosis based on medical images, we showed how the latter can be applied in formalizing expert knowledge in a compact manner.

Recent advances in medical imaging have led to a variety of modalities such as MRI, tomosynthesis, and ultrasound, to augment the current tools (primarily mammography) for breast cancer screening. The integrated interpretation of these modalities at a patient level imposes even more challenges for human readers and new modelling techniques are needed to handle both uncertainty and dynamics in findings. Probabilistic logics—the merge of probability theory and logic—is a promising direction for future research in this application domain.

References

1. D'Orsi, C.J., Bassett, L.W., Berg, W.A., et al.: BI-RADS: Mammography, 4th edn. American College of Radiology Inc., Reston (2003)
2. Davis, J., et al.: Knowledge discovery from structured mammography reports using inductive logic programming. In: 2005 Annual Symposium American Medical Informatics Association, pp. 86–100 (2005)
3. Fenton, J.J., et al.: Influence of computer-aided detection on performance of screening mammography. *N. Engl. J. Med.* **356**(14), 1399–1409 (2007). PMID: 17409321, <http://wwwnejm.org/doi/pdf/10.1056/NEJMoa066099>
4. Halford, G.S., Wilson, W.H., Phillips, S.: Relational knowledge: the foundation of higher cognition. *Trends Cogn. Sci.* **14**(11), 497–505 (2010). doi:[10.1016/j.tics.2010.08.005](https://doi.org/10.1016/j.tics.2010.08.005)
5. Horsch, K., et al.: Classification of breast lesions with multimodality computer-aided diagnosis: observer study results on an independent clinical data set. *Radiology* **240**(2), 357–368 (2006). doi:[10.1148/radiol.2401050208](https://doi.org/10.1148/radiol.2401050208). PMID: 16864666
6. Murphy, K.P.: Dynamic Bayesian networks: representation, inference and learning. PhD thesis, UC Berkeley (July 2002)
7. Pisano, E.D., et al.: Diagnostic performance of digital versus film mammography for breast-cancer screening. *N. Engl. J. Med.* **353**(17), 1773–1783 (2005). PMID: 16169887, <http://wwwnejm.org/doi/pdf/10.1056/NEJMoa052911>
8. Robben, S., et al.: Discretisation does affect the performance of Bayesian networks. In: Bramer, M., Petridis, M., Hopgood, A. (eds.) *Research and Development in Intelligent Systems XXVII*, pp. 237–250. Springer, London (2011)
9. Robinson, J.A.: A machine-oriented logic based on the resolution principle. *J. ACM* **12**(1), 23–41 (1965)
10. Shiffman, R.N., Greenes, R.A.: Improving clinical guidelines with logic and decision-table techniques: application to hepatitis immunization recommendations. *Med. Decis. Making* **14**(3), 245–254 (1994)
11. Skaane, P., et al.: Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology* **267**(1), 47–56 (2013)
12. Velikova, M., et al.: On the interplay of machine learning and background knowledge in image interpretation by Bayesian networks. *Artif. Intell. Med.* **57**(1), 73–86 (2013)
13. Wilk, S., et al.: Clinical practice guidelines and comorbid diseases: a MiniZinc representation of guideline models for mitigating adverse interactions. *Stud. Health Technol. Inform.* **192**, 352–356 (2013)