12

MIDAS – Mammographic Image Database for Automated Analysis

Fabiano Fernandes[1], Rodrigo Bonifácio[2], Lourdes Brasil[3], Renato Guadagnin[4] and Janice Lamas[5]
[1]*Instituto Federal de Brasília,*
[2]*Computer Science Department, University of Brasília,*
[3]*Post-Graduate Program in Biomedical Engineering, University of Brasília at Gama*
[4]*Post-Graduate Program in Knowledge Management and Information Technology, Catholic University of Brasília,*
[5]*Janice Lamas Radiology Clinic*
Brazil

1. Introduction

The CAD (Computer-aided Diagnosis) systems have been experiencing an exponential growth in the last decades. Since the mid-1980s of time consuming film digitization on a limited number of cases to its present status on large FFDM (Full-Field Digital Mammography) databases Giger et al. (2008). The current usage of CAD systems has brought about the need for breast cancer detection and classification efficient mechanisms. The CAD algorithm sensitivity and specificity are influenced directly by database characteristics such as image size, lesion distribution, and location of the lesion, biopsy results, BI-RADS™ classification and consensus opinion Giger et al. (2008). The use of existing mammographic databases such as DDSM (Digital Database for Screening Mammography) Heath et al. (2001b) with female patients mainly from the Massachusetts General Hospital mammography program with the statistics of 56.18% (whites), 30.34% (unknown races), 2.06% (asians), 4.12% (blacks), 6.55% (spanish surnames), 0.75% (other races), to tune the CAD algorithms can avoid time consuming and expensive achievement but on the other hand can bias the CAD algorithm to a women population not reflected in the image database, therefore causing a ripple effect in the results Heath et al. (2001a). In the current study, we developed an unprecedented database, the MIDAS (Mammographic Image Database for Automated Analysis) covering a sample of Brazilian women population - such sample is primarily based on FFDM images and whereas the population are not divides by races but on the other hand are formed by a Brazilian unique mixing of indian south american, africans and european populations along with genetic informations. The initial database contains about 600 digital mammograms including two images of each breast, associated patient information, masses, architectural distortion, special cases, calcification, associated findings, breast composition, BI-RADS™ categories and overall impression. The MIDAS mammogram images are obtained from the Janice Lamas Radiology Clinic.

1.1 The Janice Lamas Radiology Clinic

The Janice Lamas Radiology Clinic is a image diagnosis clinic founded in 1993 that performs mammography exams for diagnosis and screening, general ultrasound exams, biopsies and bone densitometry exams. The medical director is Janice Magalhães Lamas, M.D., PhD. and her clinical interests include all aspects of breast imaging and intervention including digital mammography with the use of CAD systems, breast ultrasound and dedicated breast MRI (Magnetic Resonance Imaging). Dr. Janice's expertise is requested regularly to speak at brazilian meetings dedicated to radiology. In addition to her speaking engagements, she has developing scientific research in breast cancer area and bone mineral density evaluation, and she has published numerous articles on breast imaging. The clinic is also certified by the Brazilian College of Radiology and Image Diagnosis.

1.2 The MIDAS approach

Unlike the others well established mammographic databases, the MIDAS database approach includes all mammogram images covering a sample of Brazilian women population, patient genome sequences data, open source image processing algorithms available, open access and open collaborative environment. The unique characteristic of containing mammographic imagens covering the Brazilian women population associated with the genomic sequences is a brand new inovation that enhances the scientific research and discovery. The MIDAS database also welcomes existing and new open source image processing algorithms that allow the tests and validation on the mammographic images. All scientists worldwide can participate in the MIDAS database with new algorithms and tools in a open and collaborative environment.

1.3 Motivation

The lack of a genuine brazilian mammographic database and the collaborative and open use and test of new CAD tools are the main motivation for MIDAS project. The MIDAS database is an innovative database that integrates phenotypes - the mammograms and increasingly their respective genotypes, according to patient biopsy and subsequent genome sequencing. Complex studies for mapping the effect of distinct genes and their results in the mammogram image pattern will enable the discovery of unknown cancer genes and help to understand their pathways of behavior. The BI-RADS™ assessment for all mammograms is an additive factor that enables better algorithm tunning.

2. Mammography screening

Screening to control chronic-degenerative diseases and diseases of neoplastic nature can be defined as an examination on asymptomatic women, carried out with the intent of classifying them as likely or unlikely to develop the disease Morrison (1992). The goal of screening is to define women with preclinical disease as positive and women without preclinical disease as negative. The result from the screening - positive or negative definition - reflects the efficacy of the test in showing the signs of preclinical disease as well as correct interpretation of the findings. An error, i.e. a positive definition for someone without preclinical disease, or a negative definition for someone with the condition, can result from either low test efficacy or incorrect interpretation Morrison (1992) Barratt et al. (2005). In a screening test, the matter that is under investigation is the capability to correctly distinguish between diseased and healthy individuals. Thus, to be certain that the disease is really present or absent, it is frequently

the case that elaborate, expensive or risky tests, such as biopsies, surgical exploration or autopsy, must be carried out Fletcher et al. (1996). Estimation of diagnostic test validity in relation to a standard is certified by knowing the proportions of right diagnoses (true positives and true negatives) and wrong ones (false positives and false negatives) Pereira (2000b). In relation to mammography, the level of logic validity is based on consensus and expert opinions Basset et al. (1994) Fletcher & Elmore (2005), and evaluations of accuracy are based on histopathological examination of the suspected lesion. The ability of mammography to define women with preclinical diseases as positive is referred to as its sensitivity. If measured, it is around 78 to 85%. The specificity of the test is its ability to define as negative women who do not have the disease. If measured, it should be greater than 90% Morrison (1992). The sensitivity of the screening test is a determining factor for disease control programs and the specificity directly influences the costs and feasibility of the screening program Morrison (1992) Fletcher et al. (1996) Basset et al. (1994). All mammograms must be categorized as one of the alternatives below Basset et al. (1994):

- True Positive: when cancer is diagnosed within a one-year period after a biopsy is recommended.
- True Negative: when no cancer is diagnosed up to one year after a normal mammogram is reported.
- False Positive: when no cancer is diagnosed within a one-year period after an abnormal mammogram from which a biopsy is recommended; and when there is a benign finding in a biopsy within a one-year period after an abnormal mammogram Fletcher & Elmore (2005).
- False Negative: when cancer is diagnosed within a one-year period after a normal mammogram is reported.
- Positive Predictive Value: refers to diferent rates, depending on the definition of false positive. Based on an abnormal screening examination: 5-10%. Based on a recommendation for biopsy or surgery: 25-40%. Based on the result from biopsies carried out at the clinic, from an abnormal mammogram or from another diagnostic procedure carried out on the breast, after a negative mammogram Basset et al. (1994); Sickles et al. (2002).

2.1 Prevalence and incidence

A program for early detection and treatment is applicable in the case of diseases that present a preclinical phase that cannot be diagnosed but is detectable, and for which the treatment must offer some advantage over late treatment. The proportion of the population that has a detectable preclinical phase is the prevalence. Prevalence depends, primarily, on the incidence rate of the condition, which, in turn, reflects the action of causal factors. Prevalence varies with the length of the preclinical phase. The greater the duration is, greater the proportion of affected women will be Pereira (1996) Fletcher & Wagner (1996). Finally, prevalence depends on whether or not previous screenings have been carried out Morrison (1992). A screening test with greater sensitivity can detect tumors at the beginning of this phase. Prevalence represents the stock of cases, new and old, and usually expresses damage of a chronic nature Fletcher & Wagner (1996), such as breast cancer. Incidence is recognized as the most important measurement in epidemiology because it relates to the dynamics of occurrence of a certain event, over a specific observation time Pereira (1996). For example, among asymptomatic women with previous mammography examinations that showed no suspicious signs of

malignancy, it informs how many of them present subclinical breast cancer. Incidence is one of the determining factors of prevalence Fletcher & Wagner (1996). In reality, estimation of both incidence and prevalence enables better knowledge of the situation and consequently enables adequate orientation of actions, with regard to implementation of new programs. Implementation of an early detection program for breast cancer by means of mammography must be preceded by studies that can evaluate the existing situation. Thus, before the possible beneficial impact on mortality of early detection of malignant but asymptomatic lesions can be measured, it is important to ascertain the breast cancer rate among women who are apparently healthy Warren-Burhenne (1996). In Brazil, there are few reports measuring the distribution of malignant lesions detected by mammography screening among asymptomatic women Koch (1998) Lamas & Pereira (1998).

2.2 Factors related to detection of breast cancer

Several factors influence early detection through mammography or delayed diagnosis of malignant lesions among apparently normal women. Among these are the biological behavior of the tumor, the type of equipment used, the technical ability of those who produce and interpret the mammograms, the time interval between subsequent screenings and the existence of a quality control program at the clinic Koch (1998) Sickles (1995). Technological advances and greater knowledge of breast radiology have made it possible to identify breast lesions with suspected malignancy earlier on Haus et al. (1990) Taplin et al. (2004). The biological behavior of the tumor, which is inherent to each type of neoplasia and the relationships established with the human organism, is one of the factors that determine the length of the subclinical phase and hence favors or hinders early detection Taplin et al. (2004). Length bias The growth speed of the tumor is a crucial matter, since the likelihood of cancer detection during the preclinical phase depends on the length of this phase. There is little chance of detecting tumors with a short preclinical phase before they are manifested clinically. On the other hand, tumors with preclinical phases that last for years are more likely to be discovered through screening Morrison (1992). Studies have shown that the mean duration of the preclinical phase of cases diagnosed through mammography screening tends to be longer than the mean duration of the cases that are routinely identified through the appearance of symptoms Warren-Burhenne (1996) and that tumors that grow slowly during the preclinical phase have the same behavior when symptomatic. Thus, screening would tend to detect tumors with good prognoses, as a result of the bias from the length of the preclinical phase, regardless of how much time is gained through early detection or how much benefit comes from early treatment Black & Ling (1990) Zelen (1976). Prevalence bias Breast cancer has faster growth among young women and slower growth among older women. It can remain in the subclinical phase for an indeterminate length of time Moskowitz (1986). Research has indicated that older women, aged over 50 years, have a longer preclinical phase of the disease, compared with women aged less than 50 years, in whom the biological behavior of the tumor is more aggressive Baker (1982) Kopans (1995).

However, mammography examinations contribute towards elevation of the prevalence rate, through indiscriminately detecting tumors with progressive growth and those that may remain without clinical manifestation for indeterminate periods Morrison (1992). This constitutes a distortion, known as prevalence bias, caused by excessive representation of long-duration cases and can occur in cross-sectional studies Fletcher & Wagner (1996). Although these distortions can contribute towards high prevalence of subclinical lesions, these rates have not been shown to be statistically different at ages of under and over 50 years in Brazil, which seems to indicate that prevalence bias is not the only explanation for

the measurement found, especially among younger women Lamas & Pereira (1998). These results are consistent with some other studies that observed that there was no drastic change in cancer rates from under to over the age of 50 years Kopans (1995). Since the duration of subclinical disease is greater among older women, it is expected that there will be a proportionally greater number of affected women over the age of 50 years, represented by the prevalence rates. The data from Brazil do not show statistically significant differences in the proportion of women with cancer at the subclinical phase, at an initial stage of development, comparing women under and over 50 years of age (P = 0.52). These studies in Brazil indicate that detection of a greater number of tumors with slow growth and good prognostics is not, in this particular case, related to age Lamas (2000). By contributing towards a higher detection rate for malignant tumors, mammography includes lesions of different biological behavior with distinguishing between them: lesions with rapid growth and those with slow development Liff et al. (1991). The latter may remain asymptomatic for an indeterminate period of time Feuer & Wun (1992). Regarding the frequency of malignant lesions, there is a consensus that the tumor prevalence rate among asymptomatic women is between six and ten cases per thousand women screened using mammography Warren-Burhenne & Burhenne (1992) . Rates lower than this standard range indicate that mammography is adding little to clinical examination of breasts, in which case there could be many false negatives. In Brazil, the data indicate high rates, in relation to population-based studies Vizcaíno et al. (1998) Thurfjell & Lindgren (1994). The measurements are similar to the rates observed by some institutional studies in developed countries at the beginning of their early detection programs Warren-Burhenne & Burhenne (1992) Maya et al. (2006). Proposals that are feasible for the public healthcare system need to be drawn up from quantitative data on disease frequency. Such knowledge provides information on the magnitude and importance of the damage to health Pereira (1996). The decision on whether it is appropriate or not to begin mammography screening as a secondary prevention strategy must be based on relevant measurements of breast cancer frequency among asymptomatic women.

2.3 Limitation of epidemiological studies

The variation in morbidity due to breast cancer, for which there are multiple causal factors and some are still little known, limits epidemiological studies, even in populations with similar characteristics, which may invalidate comparisons between studies Pereira (1996) Sickles (1992). It has to be borne in mind, regarding the distribution of morbidity, that the disease affects women who are less favored in socioeconomic terms. As also seen with infectious diseases, chronic-degenerative diseases and especially the advanced stages of breast cancer are a greater scourge among less favored individuals. In Brazil, breast cancer is diagnosed at advanced stages: more than 70 % of the cases are found in stages II to VI, a situation in which the chances of cure are much smaller Maya et al. (2006) . According to data from the National Cancer Institute, the estimated risk is 50 cases of this disease for every 100,000 women, including asymptomatic and symptomatic patients. It is therefore important to have public policies that establish nationwide strategies for diagnosing breast cancer, such as the breast cancer information system (SISMAMA) that has been implemented since 2008, and the Mammography Quality Program. There are still flaws in screening for breast cancer in Brazil. SISMAMA has not yet been implemented in all public clinics, and the quality control program is not available in all clinics. However, some important tools for implementing public policy regarding breast cancer prevention have been developed within the Brazilian National Health System (SUS) . Higher social classes are better informed about primary and secondary prevention mechanisms, as well as having greater access to health services, which influences

morbidity Pereira (1996). This unequal access is in addition to unequal quality at diagnostic centers, which is another condition responsible for the distorted picture of morbidity. Screening attracts people who are more conscious about healthcare, among whom the disease tends to take a more favorable clinical course, regardless of the time gained through early detection or the benefit of timely treatment. This might not have such an influence on the results if all members of the population were screened. One important point that also influences the measurement of prevalence is sample selection. Samples composed by women who seek medical care or are referred to a clinic dedicated to mammography do not constitute a random sample, but rather, attendance of the demand. One important point to be discussed in investigations from which the aim is to extrapolate the results is the use of randomization for selection of elements for a sample. Age, social levels and ethnic characteristics are factors associated with diseases, and thus, comparisons between groups with different characteristics will be biased. Differences relating to sample selection The institutional nature of some studies may introduce distortions and be responsible for differences in rates, because they constitute attendance of the demand. To minimize this sample selection bias, one option is to use women who undergo mammography examinations as periodic routine examinations required by the companies in which they work. The higher breast cancer prevalence rate at clinics that attend to the demand Sickles (1995) is explained by the greater incidence of the disease in such groups, which expresses the presence of causal factors Kopans (1995) Lopes et al. (1996). There is evidence that women exposed to a greater number of risk factors more frequently seek specialized clinics, which explains some of the differences in the frequencies of breast cancer between populations Smith (1993) Colditz et al. (1993). Women from privileged economic classes are exposed to a greater number of risk situations, such as stress, lack of breastfeeding, use of hormones, nulliparity and use of oral contraceptives, as well as other factors that are associated with greater risk of developing other diseases, including diets rich in animal fat and alcohol, among other factors Koch (1998) MacPherson et al. (1983). These conditions are associated with greater risk of having breast cancer, and personal antecedents of this neoplasia are the factor most strongly related to the disease Roubidoux et al. (1997). Atypical hyperplasia, also known as a high-risk type of lesion, is similarly strongly associated, and it has been estimated that the risk of developing in situ carcinoma or subsequent invasive carcinoma is five to ten times greater Dupont et al. (1993) Boecker et al. (2002). To affirm the causal relationship between the two events, it is necessary to rule out alternative explanations, in order to avoid erroneous conclusions. Confounding factors are one of these explanations, and therefore should be a matter of constant concern during the development of a study: in its planning, in the statistical analysis and in the interpretation of the results Pereira (1996). Differences relating to the existence of previous screenings Another factor that influences the results is the absence of previous screenings. In Brazil, unlike what is found in developed countries, there is no early detection program on a national scale, and only recently has there been any encouragement for asymptomatic women to undergo periodic mammography examinations, as a form of secondary prevention. This explains the small percentage of women who had already undergone screening in the samples of studies published on this issue. The yield from a screening test decreases proportionally as it is repeated among a given group of people Chamberlain (1984). Prevalence studies may differ in relation to the proportion of the women in the sample who had already undergone mammography examinations, since the prevalence measurement includes both new and old cases. The length bias that was described by some English authors Smith (1995) Fletcher & Wagner (1996) may contribute to the high cancer rate found in samples composed predominantly by older women, a situation in which the tumor presents slow growth. Length bias is greater among the cases detected through the initial screening, a

situation in which the prevalence of the tumors in this phase is predominantly represented by lesions that have long preclinical phases. As the control examinations are repeated, especially at short intervals, the distribution of the preclinical phase duration among the detected subclinical cases becomes more similar to the distribution routinely diagnosed in a population that did not undergo screening. In these circumstances, it could be that the length bias is less important in the selection of the prognostic factors. However, if the preclinical phase of the disease is short, continuous screenings would be necessary to reach the objective of detecting lesions at these stages Morrison (1992). Differences relating to the measurement procedures Methodological errors can give false interpretations to the results. Although the possibility is smaller, the prevalence rate may be altered by distortions introduced into the measurement of the procedures. Thus, in frequency investigations, these measurement biases appear when the findings obtained from the sample data differ from those of the population, simply through measurement problems. These deviations can occur when the event lacks definition, when poorly elaborated questionnaires are used, when uncalibrated low-resolution equipment is used, and when several data-gatherers, interviewers or observers are used, among other situations Pereira (2000a) Sickles et al. (2002). Randomized clinical trials are considered to be the standard of excellence, because they produce direct and unequivocal evidence to explain a cause-effect relationship between two events Pereira (1996).

2.4 Comparative efficacy of self-examination, clinical breast examination and mammography in screening for breast cancer

There is no way to separate the effects of mammography and clinical breast examination in relation to reduction of mortality. Even with technological advances, there are no reasons to suppose that a mammography screening program, in isolation from a clinical examination, would have the same effect as when they are combined. Since the time of the first randomized study, which began in 1960, a time at which the equipment used had low resolution, mammography has been shown to be 1.6 times more sensitive than the examination carried out separately, for mammography screening Baker (1982). It is important to emphasize that clinical examination has been recommended as an early detection method for breast cancer . Clinical breast examination is important for ruling out clinically evident tumors that do not have a corresponding mammographic configuration. Although mammography is more sensitive than clinical breast examination, 9% of tumors are only detected through palpation Baker (1982). Nevertheless, with the technological evolution of mammography equipment, some studies have observed a decrease in the rate of negative mammograms among patients with palpable nodules . In these cases, non-visualization of the tumor through mammography may result from a variety of factors, but in most cases is a consequence of increased breast density Tabár (1993) Kopans et al. (1996) Fitzgerald (2001). There is evidence that clinical breast examination contributes towards the screening of suspected lesions Baker (1982) Miller et al. (1992) Chu & Connor (1991). However, on a larger scale, it does not have an impact relating to decreased risk of mortality due to cancer Humphrey et al. (2002).

3. Database image and data aquisition

All images are generated in DICOM format by digital mammography equipments at Janice Lamas Radiology Clinic. At first moment the MIDAS database is available to general public containing about 100 mammograms and 600 images and an image sample is shown in Figure 1. The high definition images are available under personal request, after MIDAS team approval. The low definition images are freely available to the general public. All

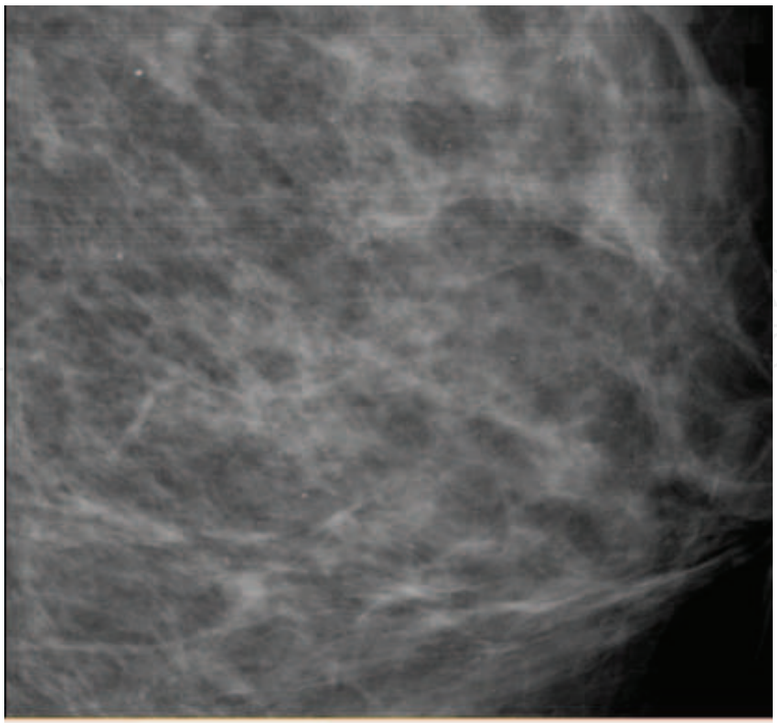


Fig. 1. A Reduced Example of MIDAS Dicom Image

the genomic information will be extracted after breasts biopsies and genomic sequencing at Genome Analyser Illumina GAIIx. All the images are obtained from the following equipment:

- Equipment: Mammomat Inspiration
- Voltage: 23 Kv to 35 Kv
- Large focus: 0.3 mm and small focus: 0.15 mm
- Pixel size: 85 μm and 70 μm
- Samples per pixel: 1
- Image size 2082 x 2800
- Bits allocated: 16
- Photometric interpretation: monochrome-1

The DBMS open source object-relational database system PostgreSQL is used to receive the MIDAS database. The database conceptual modeling is presented in Figure 2.

4. Software tools

We provide a Web application (Midas-Web) for getting access to the database content and for manipulating the database' images through several digital image processing algorithms. Therefore, at a high-level point of view, Midas-Web is both an information system (with CRUD[1] operations for the mammography database) as well as an extensible environment for

[1] In the software community, CRUD operations correspond to basic functions that allow users to create, update, delete, and query data base content

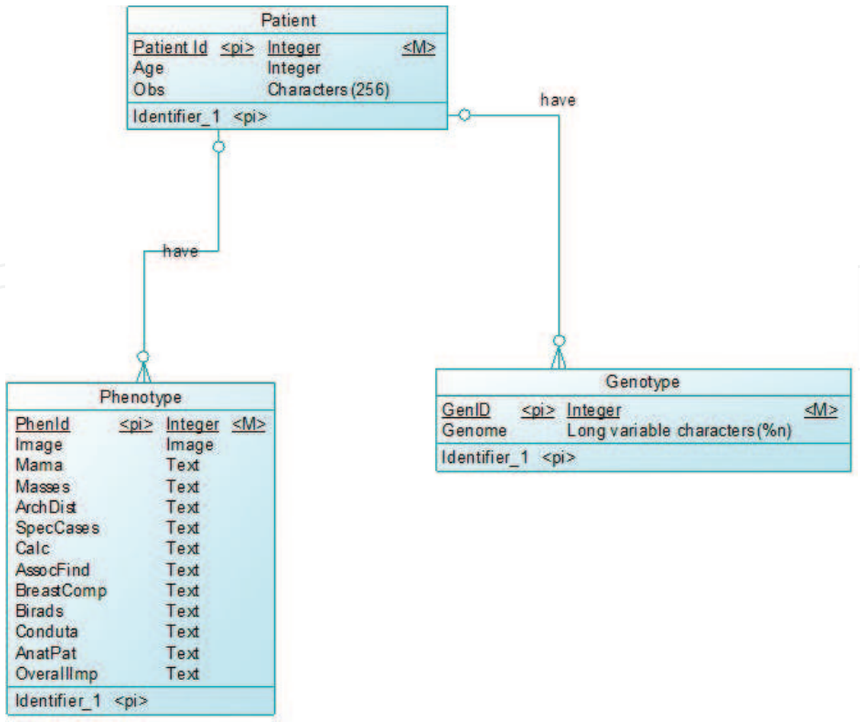


Fig. 2. MIDAS Database Conceptual Model

playing with and experimenting digital image processing algorithms. Some goals have been used to guide the design and development of Midas-Web:

- Basic mechanisms for authentication and authorization.
- Rapid application development cycle
- Extensible set of algorithms for digital image processing and interpretation

We realize three main usage profiles within the Midas-Web context. First, medical practitioners might access Midas-Web to compare, through analogy, their diagnosis with existing BI-RADS™ diagnosis present in the Midas database. Such a comparison is useful for the purpose of increasing diagnosis' confidence and teaching. Another usage profile also corresponds to medical practitioners who want to share their findings, introducing new cases in the Midas database. This usage profile is restricted by security policies, and Midas administrators analyze all requests related to the introduction of new cases before making them publicly available. Another usage profile correspond to researchers that want to apply their algorithms for digital image processing and interpretation using the Midas database images.

In the remaining of this section we detail some design decisions related to the Midas-Web architecture, database schema and extensibility support for introducing new algorithms for digital image processing.

4.1 Architecture

Midas-Web follows a standard web based architecture, in which the system is decomposed in three principal layers Fowler (2002): the *web layer*, the *business logic layer* and the *data source*

layer. The web layer provides the presentation logic, processing the user requests, calling business operations, and forwarding to a proper *graphical user interface* (GUI) component that should render the results of a user request.

The business logic layer is responsible for implementing the application transactional logic, which usually involves algorithms for data validation and calculation. Actually, in the case of Midas-Web, this layer is really thin, once Midas-Web basically provides CRUD operations to the Midas database. Nevertheless, it is still important to consider this layer in the architecture, since it increases the opportunities for code reuse.

Finally, the data source layer provides an abstraction over the underlying Midas database system. Therefore, components in this layer implement services for accessing the database, in such a way that we are able to evolve the database (for instance from a SQL based to a non SQL database) without breaking the upper layers of the Midas-Web application.

Figure 3 presents the logical view of the Midas-Web architecture. In order to reduce the development cycle, and also motivated by the low complexity of the business logic, we developed Midas-Web using Grails Smith & Ledbrook (2009). Grails (or Groovy on Rails) is a web development framework focused on productivity gains through the confluence of the Groovy dynamic language Koenig et al. (2007) and the prevalence of *source code structure convention* over *configuration through XML files*. The main components of a Grails application are: (a) the Controller Classes, which handle user inputs and forward business' responses to suitable views; (b) The Grails Server Pages, which renders the graphical user interface; (c) the Service Classes, which implements the business logic; and (d) the Domain Classes, which describe the domain concepts and implement the data access layer to perform queries and updates into the database.

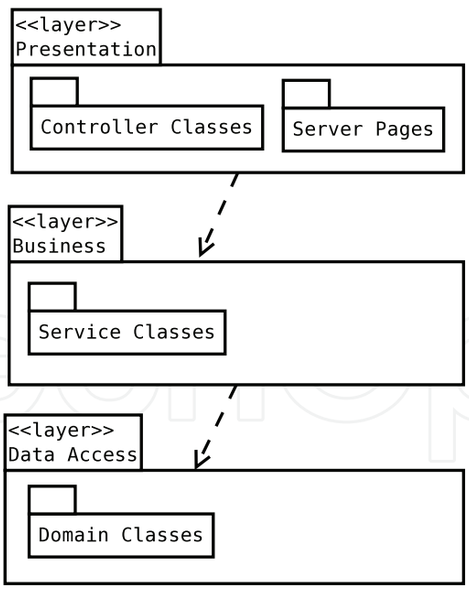


Fig. 3. Logical view of the Midas-Web architecture

Grails offers a powerful integration with the Java language, so that it is possible to call existing Java code from Grails, as well as Grails applications are package as a standard Java Web Archive (WAR) and, as such, they could be deployed into Java Application Servers (like Tom

Cat or JBOSS). The Java integration is useful because it supports the extensible architecture for digital image processing, one of the contributions of Midas-Web (Section 4.3).

4.2 Database structure

The Midas database schema was influenced, at a great extent, by the BI-RADS™ classification D'Osri et al. (2003). For this reason, each patient, whose identity must be preserved, might be associated with different studies (representing clinical cases or investigations); and each study has properties such as:

- breast composition
- histology
- date in which the exam was conducted
- the main findings of the exam

In addition, and also according to the BI-RADS™ classification, each study must provide a *lesion description* and at least four images (mammograms). Table 1 shows the attributes used to describe a lesion, whereas Figure 4 presents part of the Midas database schema.

Attribute	Domain
Assessment	Negative Benign Finding Probably Benign Finding Suspicious Abnormality Highly Suggestive of Malignancy
Mass Shape	Round Oval Lobulated Irregular Architectural distortion
Mass Margins	Circumscribed Microlobulated Obscured Ill Defined Spiculated
Calcification Type	Punctate Amorphous Pleomorphic Round and Regular Luscent Center Fine Linear Branching
Calcification Distribution	Clustered Linear Segmental Regional Diffuse

Table 1. Attributes of a lesion description according to BI-RADS™

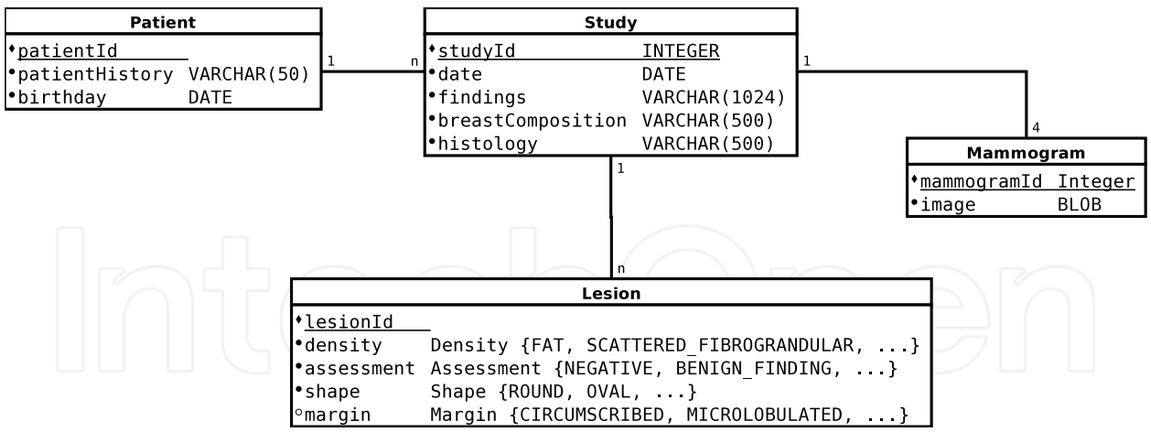


Fig. 4. Core relations of the Midas database schema

4.3 Extensibility support and collaborative environment

The MIDAS project group is a collaborative effort between Instituto Federal de Brasília, University of Brasília Computer Science Department, University of Brasília Post-Graduate Program in Biomedical Engineering, Catholic University of Brasília Post-Graduate Program in Knowledge Management and Information Technology and Janice Lamas Radiology Clinic. The MIDAS open and collaborative environment allows the users and developers to participate in the research process. After MIDAS team validating all new image processing algorithms, artifical intelligence tools, pattern recognition tools, and new findings can be added to MIDAS database.

5. Assessment and automated analysis

5.1 MIDAS assessment

The MIDAS database and application has been tested during the startup process and along its business life. The Information Technology infrastructure provided to MIDAS includes a Cyber Data Center host service and hardware usage monitoring. All the image processing algorithms and pattern recognition tools has been tested by Janice Lamas Radiology Clinic experts and reports will be generated in order to improve the software tools.

5.2 Automated analysis: Breast cancer image assessment using an adaptive network-based fuzzy inference system

The MIDAS database provides an ANFIS model algorithm Fernandes et al. (2010) for automated analysis through every 600 mammogram images. This algorithm presents an ANFIS model for a CAD (Computer Aided Diagnosis) prototype system to classify calcifications in mammograms, in order to aid the medical expert in breast cancer diagnosis. The proposed model embodies pre processing, detection, features extraction and classification phases, which proved adequate for the study domain, obtaining similar results to the indicated in the literature. This approach might be complemented with micro calcification shape analysis and image segmentation techniques. The neuro fuzzy ANFIS model, utilized in the mammogram ROI's classification phase, reached a maximum accuracy rate of 99.75% with Mini MIAS database and now can be tested with MIDAS database. This can be observed in the results presented by Fernandes et al. (2010), when the sigmoidal membership function

PSIGMF was chosen, with the training algorithm in back propagation, employing small values for epochs. The other membership functions analyzed also showed satisfactory accuracy rates, however they were not the best ones. The cross validation method allowed a higher formalism in the division of entry data (estimation, validation and test sets), which is necessary due to the small quantity of images with calcifications available in the database Mini MIAS, to prevent excessive training of the network and, consequently, a better generalization of the system. The proposed system is available as a knowledge management tool. This allows the dissemination of tacit and explicit knowledge of medical experts and their past experience in the field, allowing still a better performance in the evaluation of routine exams by means of a graphical tool.

6. Related work

The Mammographic Image Analysis Society (MIAS) is a database of digital mammograms where the films are from the U.K. National Breast Screening Programme and they have been digitised to 50 micron pixel edge with a Joyce-Loebl scanning microdensitometer, a device linear in the optical density range 0-3.2 and representing each pixel with an 8-bit word Suckling et al. (1994). The database contains 322 digitised films and it also includes radiologist's "truth"-markings on the locations of any abnormalities that can be present. The database has been reduced to a 200 micron pixel edge and padded/clipped so that all the images are 1024x1024 Suckling et al. (1994). The Digital Database for Screening Mammography (DDSM) is a collaborative effort between Massachusetts General Hospital, Sandia National Laboratories and the University of South Florida Computer Science and Engineering Department. The database contains approximately 2,500 mammograms each includes two images of each breast, along with some associated patient information (age at time of study, ACR breast density rating, subtlety rating for abnormalities, ACR keyword description of abnormalities) and image information. Images containing suspicious areas have associated pixel-level "ground truth" information about the locations and types of suspicious regions. The DDSM also provides software for accessing the mammogram and truth images and for calculating performance figures for automated image analysis algorithms Heath et al. (2001b).

7. Future work

7.1 Genotypes and phenotypes

The search of new genes of cancer and its epigenomics implications are the main areas of future investigation. Discovering the human genomic variability and its complexes phenotypes are the major obstacles and where the efforts will be concentrated. New image processing algorithms and also new artificial intelligence methods will be used in order to offer the physicians more technical support in the breast cancer treatment.

7.2 Breast cancer visual modeling

7.2.1 Motivation

Visual characteristics of human body components are highly relevant as an input for a variety of decisions on health concerning activities. One can promptly access and eventually perform adjustments on information upon some injury or physiological change through user-friendly devices, as a valuable resource for therapeutic procedures and specialized training too.

One should remark that visualization is supposed to convey information compatible with viewer's perception and cognition skills Agrawala et al. (2011). Visual modeling of cells growth is an instance of data visualization that allows a fast information analysis for problem solving. This kind of data processing more and more becomes an efficient way to support decision-making. Visualization can be understood as a low-cost human cognitive process to create an image about a domain space. It enables insights about some context, say, qualitative and quantitative answers to existing problems and facts recognition that were previously not possible Fayyad & Grinstein (2002) Konofagou (2004). Although visual models are able to express just part of the features of the real object, they are enough informative and useful for an effective subsequent decision process. Typical instances of lesion that require visual information for treatment are cells abnormal growth that may constitute cancerous tissues as a result of an evolutionary process with genetic mutations Beckmann et al. (1997) Evan & Vousden (2001) Lux et al. (2006). Images provide relevant information about size and tonalities irregularities that may express different kinds of existing lesions. The availability of a non-invasive and inexpensive computational technique to model evolutionary process of breast cancer thus becomes highly relevant. Hence one expects to develop models of breast cancer, to recognize their properties, for supporting therapeutic processes and decisions, with experimental validation in mammography clinics. Its results should be suitable for actual use in units of health care.

7.2.2 Visual modeling

A visualization system is developed according to the following steps Agrawala et al. (2011). Initially, the principles of design-oriented field of interest are identified. Then the algorithms are developed to implement these principles. Afterwards the results are validated based on users' perception of visually modeled information. Segmentation is based on certain characteristic features that are common to the pixels that should make up each segment. If the image contains multiple objects with similar characteristics, as in the case of breast duct anatomy, the approximate tones of pixels belonging to the object are a segmentation criterion. After properly separated and identified the object or region, it becomes possible to capture information that can be used for subsequent classification according known categories. Thus it is essential to know all the properties that are necessary and sufficient to characterize an object. The image of an object is a projection of a three-dimensional object on a plane. Through the image we might infer the size of the projection of the object and thus estimate the size of the object. The position of the object is useful for the activation of devices able to perform any procedure directly on the scene, for example, microsurgery or application of medication. Each property should be able to unambiguously characterize the objects belonging to different classes and at the same time be able to accommodate all the variations that can occur for objects belonging to the same class. The decision about membership of an object to some class can be deterministic or probabilistic. In fact the value of a property in a set of standards can be a random variable with estimated mean and standard deviation. So such decision is based on the membership of the measured property in a pattern to a range of the known model. More accurate procedures for statistical classification also consider the cost of misclassification, which can be quite relevant, for example, at diseases diagnosis based on recognition of images of tissue samples. Concerning therapeutic procedures it is often necessary to recognize tissues with different textures, in order to assess the extent of changes in these tissues. This is the case of analysis of the development of breast cancer. Project activities are shown in Figure 5.

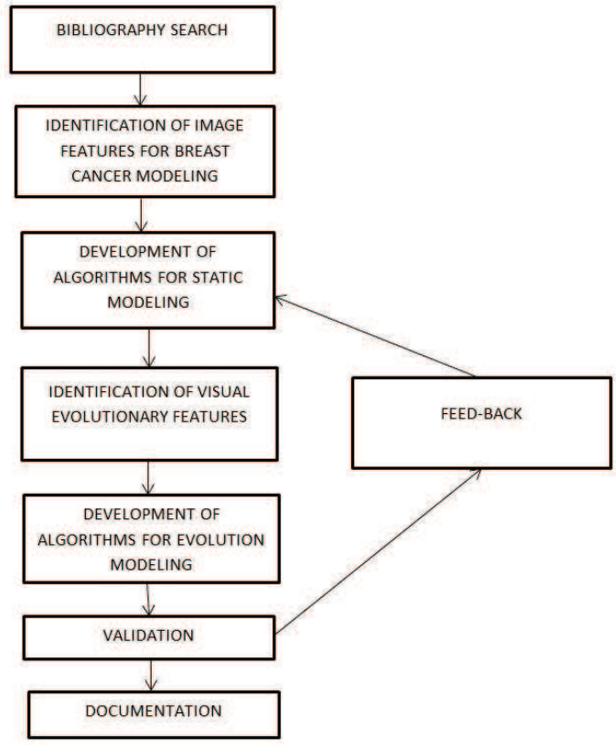


Fig. 5. Visual Modeling Project flow adapted from Guadagnin et al. (2011)

7.2.3 Expected results

One expects to have a system for visual modeling of evolutionary process of breast tumorous tissues. It will also identify the main requirements for training professionals in monitoring and management of patients supposed an appropriate use of computing resources.

7.2.4 Perspectives

Simulations could be used to advise patients and professionals about cancerous lesions evolution as well as to examine evolving response to specific treatments. The model could be also adjusted to simulate different cells mutation types Bankhead & Heckendorn (2007). The amount of image processing and analysis applications in medical diagnosis is very extent. Otherwise there are several other areas where these techniques are useful or even indispensable too, for instance, in computer-assisted surgery, post-surgery follow-up or therapy and monitoring of potentially dangerous evolution. Simulated images are helpful for healing supported by computational processes and medical diagnosis, and applications of telemedicine.

8. Conclusions

The MIDAS (Mammographic Image Database for Automated Analysis) is an innovative database containing mammogram images covering a sample of Brazilian women population, patient genome sequences data, open source image processing algorithms, open access and open collaborative environment. After its initial tunning and legal procedures it will be available to the Internet to worldwide access and collaboration. The MIDAS database project

aims to enhance the scientific activity regarding breast cancer research among brazilian women population and therefore raising the health prognosis.

9. References

Agrawala, M., Li, W. & Berthouzoz, F. (2011). Design principles for visual communication, *1.Communications of the ACM* 54(4): 60–69.

Baker, L. (1982). Breast cancer detection demonstration project: five - years summary report, *Cancer Clin.* Vol. 4(No. 32): 194–225.

Bankhead, A. & Heckendorn, R. (2007). Using evolvable genetic cellular automata to model breast cancer, *1.Proceedings of Genetic Programming and Evolvable Machines* pp. 381–393.

Barratt, A., Howard, K., Irwig, L., Salked, G. & Houssami, N. (2005). Model of outcomes of screening mammography: information to support informed choices, *Br Med J* 1(330): 936–941.

Basset, L., Hendrick, R. E. & Bassford, T. L. (1994). Quality determinants of mammography, *Technical report*, U.S. Department of Health and Human Services.

Beckmann, M. W., Niederacher, D., Schnurch, H. G., Gusterson, B. A. & Bender, H. G. (1997). Multistep carcinogenesis of breast cancer and tumour heterogeneity, *1.Journal of Molecular Medicine* 1(75): 429–439.

Black, W. C. & Ling, A. (1990). ls earlier diagnosis really better? the misleading effects of lead time and length biases, *AJR Am J Roentgenol* 1(No. 155): 625–630.

Boecker, W. et al. (2002). Usual ductal hyperplasia of breast is a committed stem (progenitor) cell lesion distinct from atypical ductal hyperplasia and ductal carcinoma in situ, *J. Pathol.* 1(No. 198): 458–467.

Chamberlain, J. (1984). Repeated screening for breast cancer, *J. Epidemiol Com Health* 1(No. 38): 54–57.

Chu, K. & Connor, R. (1991). Analysis of the temporal patterns of benefits in the health insurance plan of new york crial by stage and age, *Am. J. Epidemiol.* 1(No. 133): 1039–1049.

Colditz, G. A. et al. (1993). Family history, age, and risk of breast cancer, *JAMA* 1(No. 270): 338–343.

D'Osri, C., Bassett, L., Berg, W. et al. (2003). Breast imaging reporting and data system: Acr bi-rads, *Technical report*, American College of Radiology.

Dupont, W. D. et al. (1993). Breast cancer risk associated with proliferative breast disease and atypical hyperplasia, *Cancer* 1(No. 71): 1258–1265.

Evan, G. I. & Vousden, K. H. (2001). Proliferation, cell cycle and apoptosis in cancer, *Nature* 1(411): 342–348.

Fayyad, U. & Grinstein, G. (2002). *Information Visualization in Data Mining and Knowledge Discovery*, Academic Press.

Fernandes, F., Brasil, L., Lamas, J. & Guadagnin, R. (2010). Breast cancer image assessment using an adaptive network-based fuzzy inference system, *Pattern recognition and Image Analysis* .

Feuer, E. & Wun, L. (1992). How much of the recent rise in breast cancer incidence can be explained by increases in mammography utilization?, *Am J Epidemiol* 1(No. 136): 1423–1436.

Fitzgerald, R. (2001). Error in radiology, *Clin Radiol.* 1(No. 56): 938–946.

Fletcher, R. F. S. & Wagner, E. (1996). Prevenção, *Epidemiologia clinica: elementos essenciais.*, Artes Médicas, Porto Alegre, pp. 174–194.

Fletcher, R. H., Fletcher, S. W. & Wagner, E. H. (1996). *Epidemiologia Clínica: elementos essenciais*, Artes Médicas.

Fletcher, S. W. & Elmore, J. G. (2005). False-positive mammograms - can the U.S.A. learn from Europe?, *Lancet* 1(365): 7–8.

Fowler, M. (2002). *Patterns of Enterprise Application Architecture*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Giger, M. L., Chan, H.-P. & Boone, J. (2008). Anniversary paper: History and status of cad and quantitative image analysis: The role of medical physics and aapm, *Med. Phys.* .

Guadagnin, R., Santana, L., Brasil, L. M. & Neves, R. (2011). Visual modeling of skin wounds: A proposal to improve terapeutic processes and health management, *1.8-th Open German-Russian Workshop on PATTERN RECOGNITION and IMAGE UNDERSTANDING* .

Haus, A. G., Feig, S. A., Ehrlich, S. M. & others. others. others. others. (1990). Mammography screening: technology, radiation dose and risk, quality control and benefits to society, *Radiology* 1(No. 174): 627–656.

Heath, M., Bowyer, K., Kopans, D., Moore, R. & Kegelmeyer, W. P. (2001a). Anniversary paper: History and status of cad and quantitative image analysis: The role of medical physics and aapm, *Medical Physics Publishing* pp. 212–218.

Heath, M., Bowyer, K., Kopans, D., Moore, R. & Kegelmeyer, W. P. (2001b). The digital database for screening mammography, *Proceedings of the Fifth International Workshop on Digital Mammography* pp. 212–218.

Humphrey, L. L. et al. (2002). Breast cancer screening: a summary of the evidence for the u.s. preventive services task force, *Ann. Intern. Med.* 1(137): 347–360.

Koch, H. (1998). Projeto de detecção precoce do câncer de mama. in: Mamografia atual, *in* H. K. H. P. P. Pasquelete & C. Kemp (eds), *Comissão nacional especializada de diagnóstico por imagem da Febrasco*, Revinter, Rio de Janeiro, pp. 3–14.

Koenig, D. et al. (2007). *Groovy in Action*, Manning Publications.

Konofagou, E. (2004). *Ultrasonic Imaging*, CRC Press.

Kopans, D. (1995). Screening mammography and the controversy concerning women aged 40 – 49 years, *in* D. Kopans & E. Mendelson (eds), *Syllabus: a categorical course in breast imaging*, Oak Brook, llI: Radiological Society of North America, North America, pp. 39–49.

Kopans, D. B. et al. (1996). Positive predictive value of breast biopsy peformed as a result of mammography: There is no abrupt change at age 50 ages, *Radiology* 1(No. 200): 357–360.

Lamas, J. (2000). *Avaliação dos resultados de exames mamográficos para detecção precoce do câncer de mama, no* D*istrito* F*ederal, segundo indicadores de qualidade.*, PhD thesis, UFRJ.

Lamas, J. & Pereira, M. G. (1998). *Prevalência de lesões malignas e pré-malignas subclínicas da mama em mulheres assintomáticas no* D*istrito* F*ederal.*, Master's thesis, UNB.

Liff, J. M. et al. (1991). Does increased detection account for the rising incidence of breast cancer?, *Am J Public Health* 1(No. 81): 462 – 465.

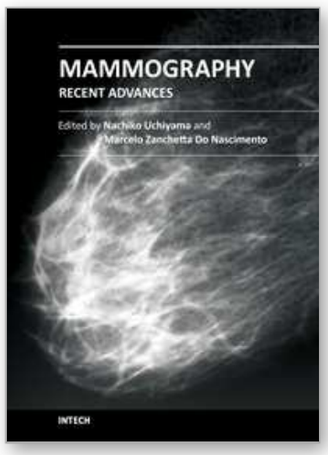
Lopes, E. R. et al. (1996). Câncer de mama: epidemiologia e grupos de risco, *Rev Bras Cancerol* Vol. 2(No. 42): 105–116.

Lux, M. P., Fasching, P. A. & Beckmann, M. W. (2006). Hereditary breast and ovarian cancer: Review and future perspectives, *1.Journal of Molecular Medicine* 1(84): 16–24.

MacPherson, K. et al. (1983). Oral contraceptives and breast cancer, *Lancet* pp. 1414–1415.

Maya, N. A. F. et al. (2006). Tendência de incidência e da mortalidade do cancer de mama em goiânia: Análise de 15 anos (1988-2002), *Rev Bras Mastol* 1(No. 1): 17–22.

Miller, A. B. et al. (1992). Canadian breast screening study: breast cancer detection and death rates among women aged 40 - 49 years, *Can Med. Assoc J.* 1(No. 147): 1459–1476.
Morrison, A. (1992). *Screening in chronic disease*, Oxford University Press.
Moskowitz, M. (1986). Breast cancer: age-specific growth rates and screening strategies, *Radiology* 1(No. 161): 37–41.
Pereira, M. G. (1996). Morbidade, *Epidemiologia:teoria e prática*, Guanabara Koogan, Rio de Janeiro, pp. 73–106.
Pereira, M. G. (2000a). Aferição dos eventos, *Epidemiologia: teoria e prática*, Guanabara Koogan, Rio de Janeiro, pp. 358–376.
Pereira, M. G. (2000b). *Epidemiologia: teoria e pratica*, Guanabara Koogan.
Roubidoux, M. A. et al. (1997). Women with breast cancer: histologic finding in the contralateral breast, *Radiology* 1(No. 203): 691–694.
Sickles, E. (1992). Quality assurance: how to audit your own mammography practice, *Radiol Clin North Arn* 1(No. 30): 265–275.
Sickles, E. (1995). Auditing your practice, *in* D. Kopans & E. Mendelson (eds), *Syllabus: a categorical course in breast imaging*, Oak Brook, lll: Radiological Society of North Arnerica, North Arnerica, pp. 81–91.
Sickles, E. A., Wolverton, D. E. & Dee, K. E. (2002). Performance parameters for screening and diagnostic mamography: specialist and general radiologists, *Radiology* 1(224): 861–869.
Smith, G. & Ledbrook, P. (2009). *Grails in Action*, Manning Publications.
Smith, R. (1993). Epidemiology ofbreast cancer, *in* A. Hans & M. Yaffe (eds), *Syllabus: categorical course in physics - technical aspects of breast imaging*, Oak Brook, lll: Radiological Society of North America, pp. 21–33.
Smith, R. (1995). The epidemiology of breast cancer, *in* D. Kopans & E. Mendelson (eds), *Syllabus: a categorical course in breast imaging*, Oak Brook, lll: Radiological Society of North America, North America, pp. 7–20.
Suckling, J. et al. (1994). The mammographic image analysis society digital mammogram database, *International Congress Series 1069* pp. 375–378.
Tabár, L. (1993). New swedish breast cancer detection results for women aged 40-49, *Cancer* 1(No. 72): 1437–1438.
Taplin, S. H., Ichikawa, L., Buist, D. et al. (2004). Evaluating organized breast cancer screening implementation: the prevention of late-state disease?, *Cancer Epidemiol Bio Prev* 1(No. 13): 225–234.
Thurfjell, E. & Lindgren, J. (1994). Population-based mammography screening in swedish clinical practice: prevalence and incidence screening in uppsala county, *Radiology* 1(No. 193): 351–357.
Vizcaíno, L. et al. (1998). Breast cancer screening: first round in the population-based program in valencia, spain, *Radiology* 1(No. 206): 253–260.
Warren-Burhenne, L. (1996). lmplications of international breast cancer screening of north american policies: The british columbia experience, *Supplement to Radiology*, Radiological Society of North Arnerica, North Arnerica, pp. 88–88.
Warren-Burhenne, L. H. T. & Burhenne, H. (1992). The british columbia mammography screening program: evaluation of the first 15 monrhs, *Am J Roentgenol* Vol. 1(No. 158): 45–49.
Zelen, M. (1976). Theory of early detection of breast cancer in the general population, *in* J. M. W. Heuson & M. Rozencweig (eds), *Breast Cancer: trends in research and treatment. New York*, Raven Press, pp. 287–300.



Mammography - Recent Advances
Edited by Dr. Nachiko Uchiyama

ISBN 978-953-51-0285-4
Hard cover, 418 pages
Publisher InTech
Published online 16, March, 2012
Published in print edition March, 2012

In this volume, the topics are constructed from a variety of contents: the bases of mammography systems, optimization of screening mammography with reference to evidence-based research, new technologies of image acquisition and its surrounding systems, and case reports with reference to up-to-date multimodality images of breast cancer. Mammography has been lagged in the transition to digital imaging systems because of the necessity of high resolution for diagnosis. However, in the past ten years, technical improvement has resolved the difficulties and boosted new diagnostic systems. We hope that the reader will learn the essentials of mammography and will be forward-looking for the new technologies. We want to express our sincere gratitude and appreciation?to all the co-authors who have contributed their work to this volume.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen