UNIVERSITA' DEGLI STUDI DI
NAPOLI FEDERICO II

Facoltà di Ingegneria
Corso di Studi in Ingegneria Informatica

Tesi di laurea specialistica

Automatic lesion detection in breast DCE-MRI

Anno Accademico 2012-2013

relatore
Ch.mo prof. Carlo Sansone

correlatore
Ch.mo prof. Mario Sansone

candidato
Gabriele Piantadosi
matr. 885/282

Contents

Introduction

According to the British oncologist R.A. Willis [1] a neoplasm "is an abnormal mass of tissue, the growth of which exceeds and is uncoordinated with that of the normal tissues, and persists in the same excessive manner after cessation of the stimulus which evoked the change"; this is the commonly recognized definition of "cancer" in the medical field. This plague is the first cause of death in developed countries and the second cause of death in developing ones [2]. This pathology has always scared population: the most ancient read testimony about cancer is dated 3000 B.C. and it refers to a breast cancer [3]; man learned to engage this pathology thanks to the scientific and cultral progresses: efficient diagnostic method, preventative measures and developing effective and less invasive therapies.

However the numbers of tumors, although mortality and late-stage incidence steadily decreasing [2, 4], remain unacceptable and therefore encourage the entire scientific community to engage significant resources in all sectors concerned.

Breast cancer is the most common form of cancer in women (tab.1 sorted by Age Standardized Rate - ASR) and it is globally the second one for the number of cancer deaths with a mortality rate of 12 per cent of all cancer deaths [5] (tab.2 sorted by Age Standardized Rate - ASR); these numbers are used to justify the rapidity with which the techniques of investigation, prevention and treatment evolve and adapt.

Cancer	number	ASR
Breast	1384155	38,9
Cervix uteri	530232	15,2
Colorectum	571204	14,6
Lung	515999	13,5
Stomach	348571	9,1

Table 1: First 5 cancers (by incidence) in women [6]

Cancer	number	ASR
Lung	1376579	19,3
Breast	458503	12,4
Stomach	737419	10,3
Liver	695726	9,9
Colorectum	1235108	8,2

Table 2: First 5 cancers (by mortality) in both sex [6]

Generally there are many options available for the treatment of various neoplasms, both benign and malignant, such as: surgery, chemotherapy, radiotherapy and palliative care. Therapy choice largely depends on the location and the type and on stage of cancer, as well as the health of the patient and his expectations. In any case, the medical community agrees that early detection decisively improves the bounty of prognosis [7]. For this reason, most of scientific research is dedicated to improving cancer screening techniques and different survey instruments. Oncological screening of the breast cancer involves extensive use of mammograms and ultrasound as primary imaging investigative tools [8, 9].

Since many years is growing using of Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) in the prevention and early diagnosis

of breast cancer [10].

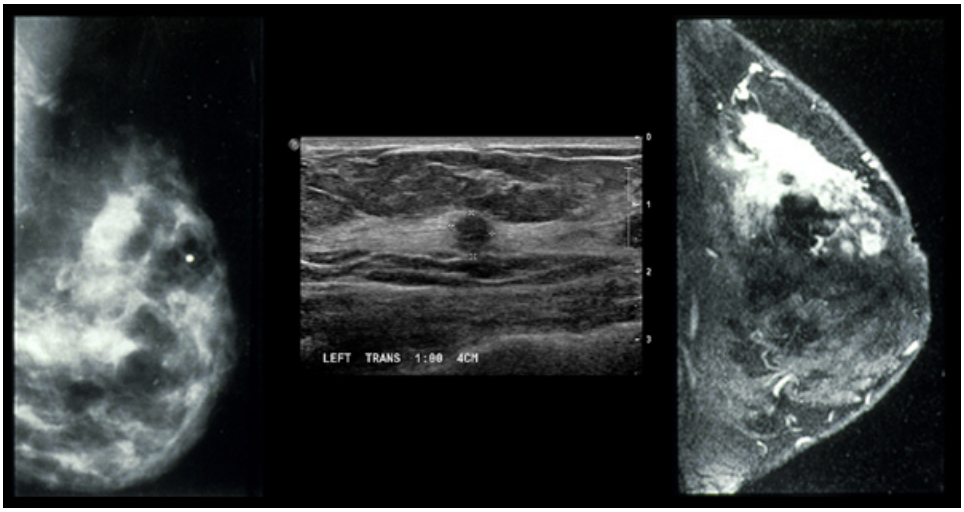


Figure 1: Visual comparing (in order: Mammography, Ultrasound, MRI)

Comparing this diagnostic technique to the previous and still valid tools such as mammography and ultrasound, factors in favor of this imaging tool are:

- magnetic waves used in MRI have no undesirable effects on patients unlike of ionizing radiation (x-rays), themselves cause on the long term, of tumor (due to the high energy that can permanently damage ties of DNA cells) [11];
- MRI with the help of contrast (DCE-MRI) allows to expand the age range of the screening program standards favoring the diagnosis in patients known as "under forty" who, before the age of 40, show a particularly dense glandular tissue of the breast and then that mammography does not discriminate decisively tissue lesions;
- the dynamic characteristic of the contrast medium allows to discriminate effectively tumors that exhibit a strong factor in angiogenesis; this feature is the most important added value in the quality of the results obtained.

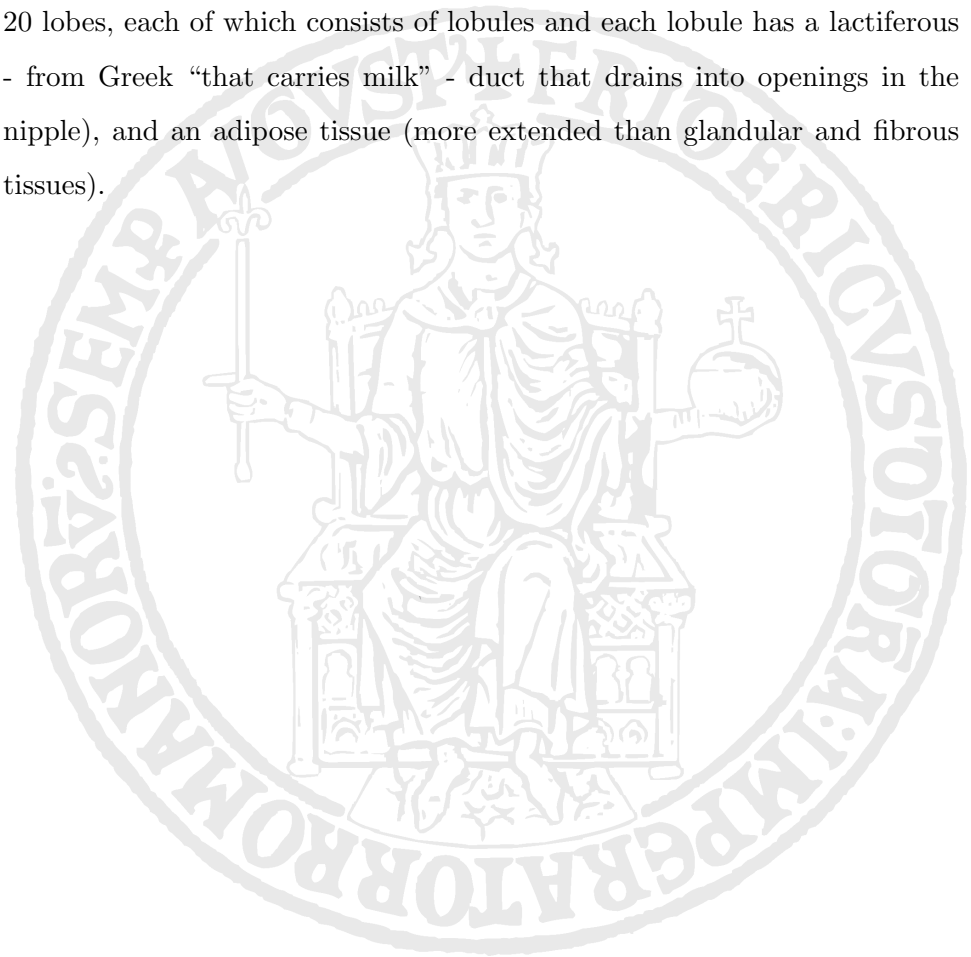
The analysis of a DCE-MRI is a complex and meticulous job due to the huge amount of data to be analyzed and the complexity of 3D representation of the entire scanned volume. Moreover, the time evolution of the intensity signal dynamic components is a decisive factor in the evaluation of the lesions.

The aim of this thesis it to develop an automatic, distributed, and comprehensive system of breast segmentation that can support the job of the radiologist providing a useful help service. These instruments are called CAD (Computer Aided Detection\Diagnosis); in the specific the proposed system makes use of pattern recognition techniques and supervised learning applied to a set of features extracted, after suitable pre-processing phase, from the acquired data. This instrument was developed in Matlab and interfaced via a special plug-in for OsiriX [12], the most used tools in the field of diagnosis and analysis of image data in DICOM format. Finally, to improve the performance of the entire process, the architecture was been distributed by sharing functionality between a client and a server and by centralizing and optimizing the more complex phases.

The thesis is divided into 5 chapters: the first briefly outlines the breast cancer and the principles of DCE-MRI, providing the basis for understanding operations of the proposed automatic segmentation; the second presents the CAD systems, analyzing the various stages, the data involved and the different possibilities of implementation known; the third describes the architecture of system by dividing it into two main components, the process of image segmentation DCE-MRI and architecture for secure distribution of content, focusing on implementations details and architectural choices of both components; the fourth presents the system and its components in terms of experimental results and performance; finally, the fifth chapter discusses the observations made based on the results obtained and any possible future developments.

1 Breast cancer

Mammary gland is an even and symmetrical organ, placed on the front wall of the chest, consists mainly of a fibrous tissue, a glandular component (15-20 lobes, each of which consists of lobules and each lobule has a lactiferous - from Greek "that carries milk" - duct that drains into openings in the nipple), and an adipose tissue (more extended than glandular and fibrous tissues).



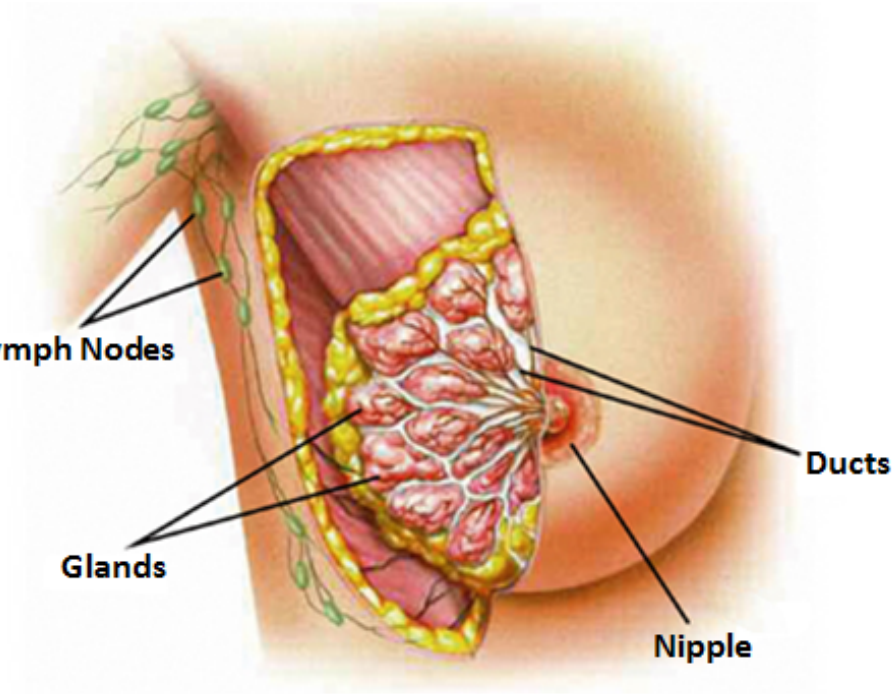


Figure 1.1: A female breast and its main parts

In order to understand the techniques, methods and approaches used in the whole work it is necessary to investigate the context in which they were developed, describing the main types, the related characteristics of breast cancer and the theory behind the DCE-MRI, focusing on the features that are particularly useful in highlighting the characteristics of neoplasms in exam.

1.1 Breast cancer

Our earliest written record regarding cancer (although the word cancer was not used) was discovered in Egypt and dates back to about 3000 BC. It is called the "Edwin Smith Papyrus" and is a copy of part of an ancient Egyptian textbook on trauma surgery. It describes 8 cases of tumors or ulcers of the breast that were treated by cauterization with a tool called the fire

drill. The writing says about the disease, "There is no treatment" [3, 13]. The cure of the disease over time has seen a shift in focus considering, at first surgical treatment as the only one therapy approachable and then, to the present day, considering the well-known dynamics of such injuries as important factors in planning curative or palliative therapy.

The number of cases of breast cancer worldwide has significantly increased since the 1970s, a phenomenon partly attributed to the modern lifestyles of Western world [14]; in fact recent studies have shown that tumors are, for the most part, environmental disease rather than genetic, with a ratio of 9:1 in cases attributable to environmental factors than cases attributable to genetic factors [15]. Among the environmental factors we have to consider any etiological factor isn't genetically inherited, such as pollution, smoking, nutrition, radiation, stress and trauma [16].

Neoplasms of the breast represent the most important lesions of this organ, although not the most frequent, they may assume macroscopic and histological aspects extremely variable according to the type of tissue from which they originate. Since that the tissue of the breast is same in the female and in the male, the breast cancer can also affect the men (even if it represents less of 1% of all male cancers) [17, 18]. In both sexes, the incidence is higher in the left breast and upper-outer quadrant breast as shown in fig.1.2 [19, 20].

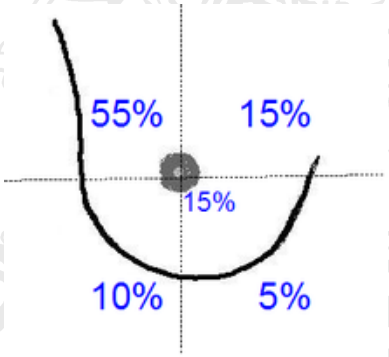


Figure 1.2: Spatial incidence of breast cancer

It is possible to rank breast tumors through a classification system introduced specifically to standardize and simplify the communication process of medical oncologists, the discriminating characteristics are: the involved tumor tissue, the histological characteristics highlighted and staging.
TNM (Tumor, Nodes, Metastases) Standardization is very used [21], it was developed in 1987 jointly according to the criteria of the UICC (International Union Against Cancer) and AJCC (American Joint Committee on Cancer); specifically TNM table make use of that codes:

- **T** distinguishes the primary tumor whose size and nature is expressed by the value that comes with it (X - 0 - IS - 1MIC, 1a, 1b, 1c, 1d - 2 - 3 - 4a, 4b, 4c, 4d). TX or T0 indicates a tumor cannot be definable or detectable; T1c tumors with size up to 2 cm to gradually evolve into T4b for tumor of any size but already adherents to the skin that has led to a hollowing or orange peel skin until T4d that indicates an inflammatory carcinoma;
- **N** indicates, if and to what extent, the lymph nodes are affected by breast cancer. The values that accompany N are different (X - 0 - 1a - 1b1, 1b2, 1b3, 1b4 - 2 - 3) and indicates by N0, absence of metastasis to N3 when there is metastatic involvement of the lymph nodes belonging to the ipsilateral internal mammary chain (N3);
- **M** refers to the presence of any metastases. MX indicates inability to ascertain the presence of distant metastases, excludes the M0, M1 locates in the distant organs.

In the case of breast cancer, table 1.1 shows the main TNM codes associating them with the staging code, it also standard, which indicates the extent of the disease allowing to compare patients and the results of different treatment protocols and to formulate a more accurate therapeutic and prognostic opinion.

Stage	TNM Code	Description	5-year survival
0	Tis N0 M0	Presence of in situ carcinomas: • Lobular carcinoma in situ: cancer is not aggressive but can present a risk factor for formation of a malignant lesion upcoming aggressive; • Ductal carcinoma in situ: affects the cells of the ducts. The possibility of having a more aggressive cancer increase.	93%
I	T1 N0 M0	Cancer is in the initial phase with a diameter less of 2 cm. There are no lymph nodes involved.	88%
IIA	T0 N1 M0	Cancer is localized in the lymph nodes of the breast.	81%
	T1 N1 M0	Cancer has a maximum diameter of 2cm, with infiltration in the lymph nodes of the breast.	
	T0 N0 M0	Cancer has a diameter of between 2cm and 5cm. There are no lymph nodes involved.	
IIB	T2 N1 M0	Cancer has a diameter of between 2cm and 5cm, with infiltration in the lymph nodes of the breast.	74%
	T3 N0 M0	Cancer has a maximum diameter of 2cm. There are no lymph nodes involved.	
IIIA	T0 N2 M0	Cancer has spread to lymph nodes of the breast.	67%
	T1 N2 M0	Cancer has a maximum diameter of 2cm, with considerable infiltration in the lymph nodes of the breast.	
	T2 N2 M0	Cancer has a maximum diameter of 5cm, with considerable infiltration in the lymph nodes of the breast.	
	T3 N1 M0 T3 N2 M0	Cancer has a diameter greater than 5cm, with considerable infiltration in the lymph nodes of the breast and/or other structures.	
IIIB	T4 N* M0	Cancer has spread in the tissue adjacent to the breast (skin, ribs, thoracic and pectoral muscles) with infiltration of the lymph nodes of breast	41%
	T* N3 M0	and/or of axillary cavity and/or of neck and/or of clavicle.	
IV	T* N* M1	Cancer has developed metastases, typically in the bones, lungs, liver and brain.	15%

Table 1.1: TNM table with stage and survival index at 5 years

1.2 Biomedical Imaging

With the terms "imaging" or "biomedical imaging" or "diagnostic by imaging" refers to the generic process through which it is possible to observe an area of a body not visible from the outside. The non-invasive diagnosis plays an important supporting role in prevention programs and, specifically, is the screening tools that are currently the most effective in the fight against cancer. Howevert the effectiveness of each imaging technique need to be evaluated in a risk/benefit balance.

1.2.1 Techniques for breast imaging

Imaging techniques applied in the research of breast diseases are:

Mammography (RX) uses ionizing radiation (X-rays) at low energy

(30 kVp) to impress images of the breast on planar X-ray films. This instrument is the "gold standard" for breast imaging because it is a simple test, fast running, highly specific and widely available on the territory; constituting, at present, the key exam for screening programs with the ability of recognition of the lesions when the tumor is not yet visible by palpation (preclinical stage). By contrast, as mentioned above, the ionizing radiations are, in the long term, themselves the cause of tumors [11] and the image quality is lower when compared to other diagnostic methods. Sensitivity: 85-90% for adipose breasts, 70% for dense (low fat content) breasts; specificity: 90-95%.

Ultrasound (ECT): based on the principles of the emission of echo and the transmission of the ultrasonic waves (between 2 and 20 Mhz choice taking into account that higher frequencies have a greater resolving power of the image, but penetrate less deeply into the subject); it is often used to complement other investigations such as mammography and clinical examination or for further diagnosis of lesions which persist dubious. In addition, the ECT replaces mammography in the study of hyper-dense breasts (rich of glandular tissue) as in women under 40 years (so called "under-forty"). In contrast ultrasound has a low resolving power for the breasts with a normal distribution of tissues and it is an operator-dependent procedure, since special skills are required (as manual dexterity and spirit of observation); at last it needs a coupling gel between the probe and the breast to eliminate the refraction effect of air. Sensitivity: 77%; specificity: 89% [22]. ETC, in association with mammography, improves the diagnostic accuracy by increasing the sensitivity (up to 90%) and specificity (up to 98%).

Magnetic Resonance (MRI): based on the physical principles of nuclear magnetic resonance, through the use of electromagnetic fields

and radio frequencies, allows to generate tomographic images (multi-layer) of tissues reconstructing a digital volume at three dimensions of the organ under examination. It is highly suitable for the investigation of lesions with strong neo-angiogenesis (the ability of a tissue to release VEGF growth factors that stimulate the proliferation of new blood vessels) and thanks to this prerogative MRI is widely used in the diagnosis of breast cancer even if it has contraindications related to long (about 40 minutes) capture process causing discomfort at patients for claustrophobic, or also other problems such as allergic to contrast media, pacemaker, ferromagnetic implants and extreme obesity. Sensitivity: 98%; specificity: 81% [22].

Computed Tomography (CT/TAC): as for the Mammography, uses ionizing radiations (X-rays) allowing to reproduce sections or layers (tomography) of the patient body and to perform three-dimensional elaborations [23]. It has few benefit in the diagnosis of breast cancer because, although it is in tomographic projection, it does not provide additional meaningful information to the common mammography (RX planar) increasing only the dose of ionizing radiations and the complexity of the acquisition procedure.

Nuclear Medicine (PET): in these investigation tools are collected all the diagnostic instruments for image that use a radiopharmaceutical formed by a radio-isotope tracer with short half-life, chemically bound to a biologically active molecule, called "vector", that indicates tissue metabolic activity. After a waiting period, during which the metabolically active molecule (usually a sugar) reaches a specified concentration inside the organic tissue to be analyzed, the isotope (with short average life) decays, emitting a positron. After a path that can reach at most a few millimeters, the positron annihilates with an electron, pro-

ducing a pair of gamma photons both of 511 KeV of energy emitted in opposite directions (back to back photons) [24]. PET provides physiological information, unlike all other diagnostic tools that provides morphological information of the anatomical region under examination. The survey instruments offered by nuclear medicine are often recommended for the purpose of diagnosing and staging metastases or discover the involved lymph nodes. The power of radio-tracer, in fact, differentiating the metabolic activity of the cells, allows to reach a sensitivity of almost 100% and a specificity of 98% in the evaluation and lymph node or in the search for metastases [25].

We can summarize the key features of (digital) imaging techniques, till now examined in the next table:

Technique	Sizes (px)	Numerical Resolution	Spatial Resolution	Sensitivity	Specificity
RX	2048	12–14 bit	0.1–0.2 mm^3	90% (70% hyper-dense)	95%
ECT	512	8 bit	(variable)	77% (90% con RX)	98%
MRI	256–1024	8–16 bit	1 mm^3	98%	81%
TAC	512	16 bit	2–4 mm^3	-	-
PET	64–256	32–64 bit	5 mm^3	-	-

Table 1.2: Numerical comparison of survey instruments by images

1.2.2 MRI e DCE-MRI

The principle of nuclear magnetic resonance is based on measurement of the precession of the spin of protons (or other nucleus with a magnetic moment as shown in fig.1.3) when they are subjected to a magnetic field. In conventional MRI the energy acceptor system is represented by hydrogen nucleus (H), the simplest and the most copious element in the human body, characterized by a not zero spin quantum number (spin 1/2).

The Instrument for MRI consists of a big main magnet capable of generating

a static homogeneous magnetic field (B) whose intensity varies, depending on the application cases and of quality requirements, from 0.1T in 10T reaching almost 20T in the experimental applications.

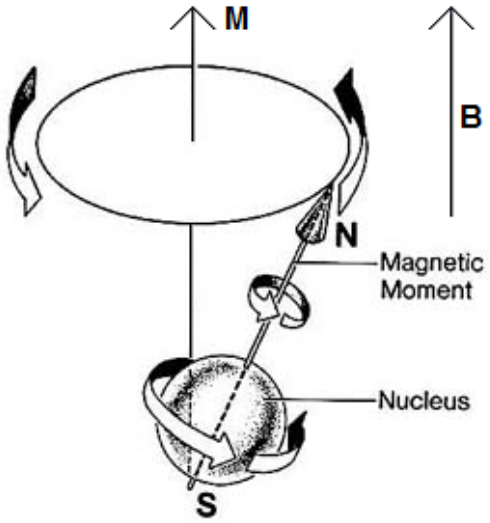


Figure 1.3: magnetic moment and precession of a proton with spin

The magnetic fields is generated through use:

- **Permanent magnets:** have a limited intensity; they cannot be turned off, raising questions about safety and maintenance;
- **Resistive electromagnets:** made of copper solenoids with a low constructive cost but high cost of use;
- **Superconducting electromagnets:** are the most common and consist of a solenoid made of superconductive material which, cooled with helium at a temperature of 4K, cancels their electrical resistance, allowing a cost of using more applicable.

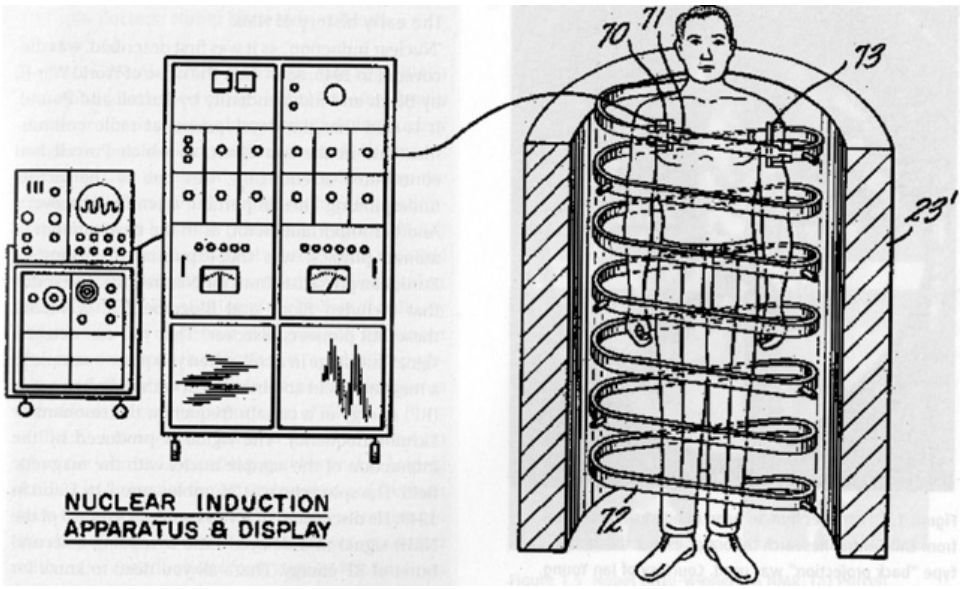


Figure 1.4: Raymond Damadian's: first "Apparatus and method for detecting cancer in tissue." [26]

Surrounded by the static magnetic field (B) generated by the main magnet, the spins of the protons inside the tissues tend to be aligned the lines of force (in parallel or anti-parallel way); because the spins that are aligned in parallel direction are in greater number, the tissue will possess a light total magnetization (M). This alignment is never total, but rather the spins of the various protons begin to show a precession (fig.1.3) around the direction of the magnetic field.

This precession shows a typical frequency called the Larmor frequency which is in the order of MHz, and then in the field of radio frequency (for a magnetic field of 1T and for the hydrogen atoms, the Larmor frequency is 42.6 MHz); then, if on the patient is applied a rotating magnetic field in this exact Larmor frequency and of sufficient energy, it is possible to rotate the magnetization of the protons of an arbitrary angle (said flip-angle) [27]. The transfer of energy required to rotate the magnetization occurs for the principle of resonance, physical principle that naming the imaging technique. After the impulse, the spins of the protons will gradually tend to return to

their initial alignment along the field (relaxation phenomenon); the trend of the magnetization is measured through a receiver coil in the plane perpendicular to the main magnetic field (this signal is called Free Induction Decay, or FID). That relaxation occurs with two separate time constants [28]:

> T_1 indicates the rate at which (how quickly) is reconstructed the magnetization vector along the direction of the main field, and it depends on the interaction between protons and the surrounding molecules (spin-lattice relaxation);
>
> T_2 indicates the rate at which (how quickly) is destroyed the component of transverse magnetization in ideal conditions, and it depends on the mutual interaction of protons neighbors (spin-spin relaxation).

The generation of images occurs through the repeated acquisition of signals coming from the body and the appropriate modulation using the gradient coils. Each voxel (Volumetric pixels, equivalent to a pixel in space in three dimensions) of the image has a frequency and/or a different phase respect to all the others, in this way it is possible to separate the signals coming from a single portion of tissue [28]. The final image is achieved through instruments and elaborations such as filters in frequency (to highlight one Larmor frequency and therefore a single slice) and Fourier Transformations (to decompose the signal in module and phase just coded to represent the remaining two sizes).

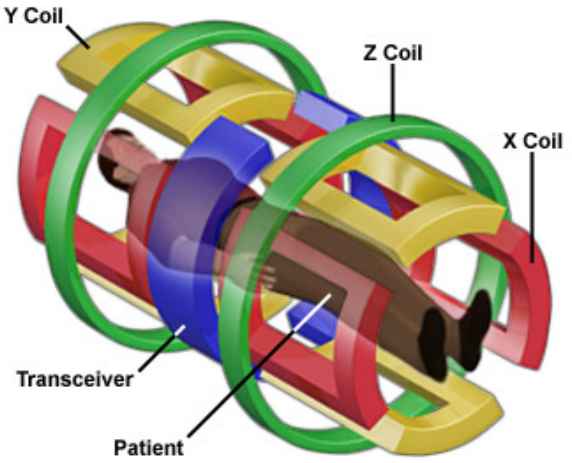


Figure 1.5: MRI Scanner Gradient Magnets

Peculiarity of magnetic resonance imaging, compared to other imaging techniques, is the ability to produce images that reflect different physical properties depending on the type of acquisition sequence used. The classic sequences are:

Proton Density (PD): estimates the number of H protons in resonance per unit of the tissue (voxel) directly calculated by the amplitude of the FID signal;

T_1 **weighed:** higher brightness of the voxel indicates a short T_1, typical of the tissue with a small molecular structure useful to highlights breast parenchyma or adipose tissue;

T_2 **weighed:** higher brightness of the voxel indicates a long T_2 and then highlights tissues containing water such as, for example, cyst.

Angiogenesis and Dynamic Contrast Enhanced

One of the cellular mutations generated by tumors is the ability to generate new blood vessels with the aim to bring oxygen and nutrient factors to the tumor cells themselves (fig.1.6); this mutation called neo-angiogenesis is

common to all types of cancer and it is supported by growth factors produced by the tumor cells themselves (VEGF growth factor).

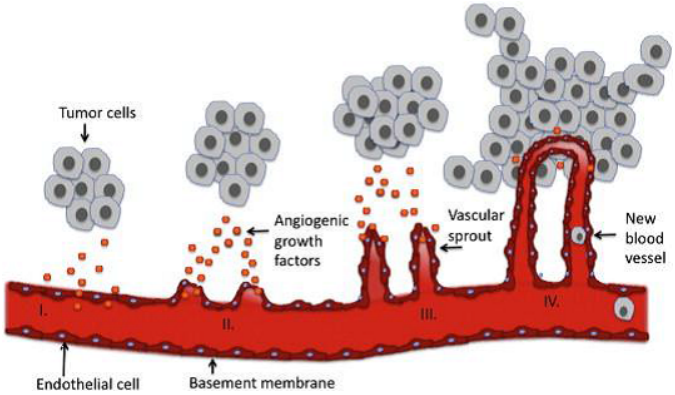


Figure 1.6: Process of tumor neo-angiogenesis

Magnetic resonance imaging involves the use of para-magnetic or super-para-magnetic contrast media which affects in an indirect manner on the information of the image, i.e. by altering the magnetic properties of the sensitive nucleus present in the tissue under examination (altering the relaxation times T_1 and T_2). The contrast agents are injected intravenously and the most used are:

- Gadolinium (Gd-DTPA), which has the largest number of unpaired electrons (equal to 7) and so the greater degree of para-magnetism;
- Iron Oxide particels (super-para-magnetic) whose effect is performed mainly on the T_2 relaxation time.

The contrast medium, circulating in the venous system, spreads with different speed in function of the vasculature of the tissue and, just because of the tumor properties of neo-angiogenesis, allows to highlight the damaged tissue respect to the surrounding healthy tissue.
Experimental studies have shown that the intensity of signal in time $SI(t)$ in the case of T_1 weighing well approximates the level of concentration of

contrast medium in time [29]. Studying the evolution of the time of the FID signal can be obtained a fourth dimension that, fixed a voxels (and then the 3 remaining dimensions), describes the trend, over time, of the effects of the contrast medium at that exact point in the tissue under examination (fig. 1.7). This curve is called TIC (Time Intensity Curve) or Enhancement Curve.

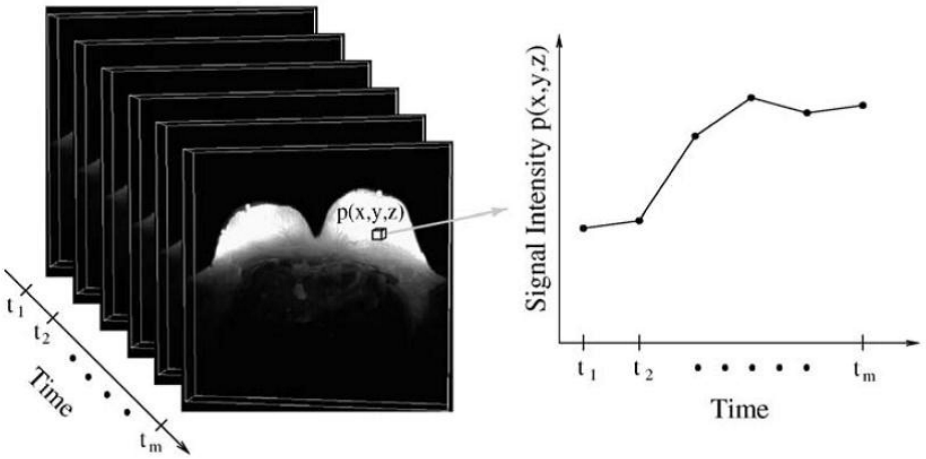


Figure 1.7: TIC of a voxel calculated from FID

A visual diagnosis can be obtained by analyzing the trend of the enhancement curve and comparing it to sample curves (fig.1.8) typical of particular tumor formation or that they approximate the dynamics of the blood flow typical of neoplastic lesions [30]:

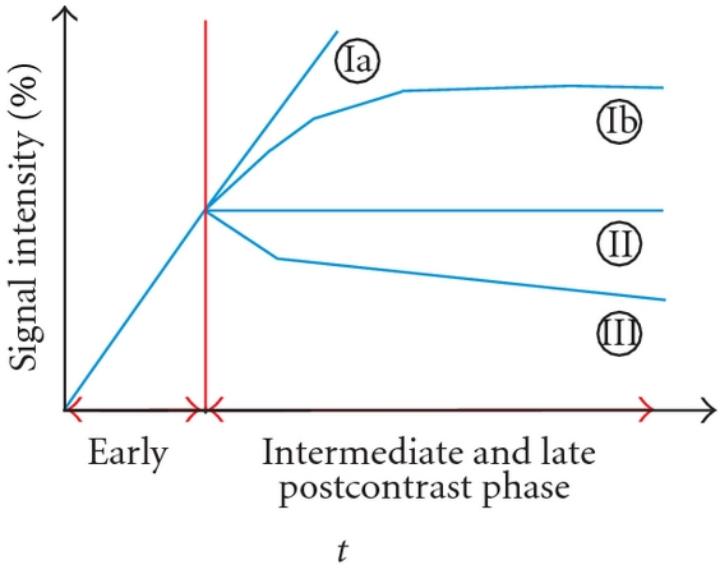


Figure 1.8: TIC sample curves

The sample curves ,shown in the graph, are considered guidelines for the visual interpretation of TIC and they can be divided [30]:

Type I: corresponds to a straight (Ia) or curved (Ib) line; enhancement (hence the contrast absorption) continues over the entire dynamic study [typical of healthy tissue or benign neoplasms];

Type II: a plateau curve with a sharp bend after the initial upstroke [typical of probably malignant lesions];

Type III: a washout time course [typical of malignant lesions].

2

Computer Aided Detection and Diagnosis

The common screening programs, that we said to be based on use of mammography as a primary method of investigation, strongly suggest the use of "double reading" [8] (repeated assessment several times by the same radiologist or by different radiologists) with possible discussion among radiologists or, in the case of divergent opinions, the opinion of a third radiologist. This recommendation helps us to understand the complexity and sensitivity of analysis of breast mammography. The radiologists often use of tools that assist in the detection of cancerous lesions till also to the evaluation of a complete diagnosis [31, 32]: these instruments are known as CAD (Computer Aided Detection\Diagnosis) and, supported by an appropriate and proved medical validity, are widely used in the analysis of complex medical investigations both for the extension of data to be taken into account (MRI\TAC\PET) and for an intrinsic uncertainty of the data due to the scanning process (ECO).

CAD systems analyze data through strict mathematical patterns, according to well-defined and deterministic algorithms. This feature allows to remove the difficulties due to intra-observability and inter-observability, represented by different valuations of the same region, under the same assumptions, by same doctor on different moments, and different evaluations of the same region by different doctors. Math feature behind the deductions (both in

detection phase and in diagnosis phase) let to evaluate sensitivity and specificity of such instruments in a precise and strict way showing objective improvement in these parameters [33, 34].

2.1 Need for CAD system: human error

The overall assessment of breast cancer using imaging tools is the result of a process that at first passes through a visual analysis and then through subjective opinion; so the human error can occur, both in the visual analysis and in the last stage of subjective opinion. Nowadays there are no automatic tools able to totally replace the radiologist in the diagnosis, but CAD systems help and support the doctor to reaching an opinion reducing the mistakes made during the assessment flow.
In 1987 Nodine and Kundel [35], studying the movement of the eyes of five radiologists during the study of chest x-ray searching for lung cancer (very similar in morphology to breast cancer), showed that the reading process is organized and selective, mainly focused on the regions of the image that the physician considers the most information-rich: eye movement was neither comprehensive nor systematic and, at the end of the inspection of the image, some areas have never crossed from the look of the doctor. At the end of the study, the authors proposed a system of classification of errors divided into three categories:

Sampling Errors: occur when the lesion does not fall within the field of view of the radiologist;

Recognition Errors: occur when the lesion is crossed by the look of the doctor, but is not recognized as such;

Decision-Making Errors: occur when a suspicious area is localized, but misclassified.

If we consider more complex diagnostic instruments such as DCE-MRI, object of this thesis, we should add that the search area is not a single two-dimensional image but a three-dimensional volume containing, usually, 256 pixels * 128 pixels * 80 pixels for a total of about 3 million of voxels; we should also add that the information contained in the dynamic evolution of the contrast medium is an important decision parameter in the evaluation of neoplastic lesions.

2.2 Steps of a CAD system

A CAD system, essentially, is composed of a set of waterfall performed independent stages; implemented steps are congruent to the purposes of the system that can range from the simple support to the doctor by providing useful instruments and information in a simple way to automatisms for the detection of suspicious lesions or the final classification of the same lesions into benign or malignant.
Literature [36] proposes the following (fig. 2.1) steps division:

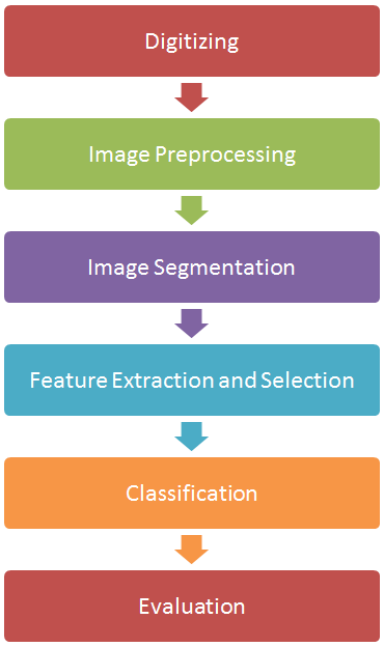


Figure 2.1: CAD mass classication [36]

Of course any additional step does not affect the achieved level of automation but it can only improve the quality or the numerousness of results. There is no need that all the stages are implemented in order to classify a system as a CAD; in fact there are very simple tools that implement only the early stages of classification flow, but that provide important and often crucial support in the final diagnosis. These support instruments, in the case of the study of a DCE-MRI breast, can implement: representations into different projection, three-dimensional reproductions of the scanned volume, temporal representation of the evolution of the dynamic signal (TIC) of a single Voxel or of a set of Voxel (average TIC), subtractive images, segmentation by thresholding, contrast adjustment and interpolation.

2.2.1 Digitizing

Instruments of modern diagnostic investigation are mainly electronic devices that produce a raw data that can be directly subject of the CAD elabora-

tion. One of the few diagnostic method still on analog support is mammography. It is gradually shifting toward digitization with indications of modern screening programs [37]. The digital imaging systems require the use of the DICOM format (Digital Imaging and Communications in Medicine): this is a standard for handling, storing, printing, and transmitting information in medical imaging [38]. DICOM differs from some, but not all, data formats because it groups information into data sets. That means that a file of a chest x-ray image, for example, contains the patient ID within the file, so that the image can never be separated from this information by mistake. Generally a DICOM data object consists of a number of attributes, including items such as name, sex, ID, etc., and also one special attribute containing the image pixel data [39, 40].

PatientID	MagneticFieldStrength
PatientName	ContrastBolusAgent
PatientSex	EchoTime
PatientBirthDate	RepetitionTime
PatientAge	SliceThickness
PatientWeight	PixelSpacing
PatientReligiousPreference	BitDepth
AcquisitionDateTime	BitsAllocated
ProtocolName	BitsStored
Modality	HighBit
Manufacturer	Columns
SeriesDescription	Rows

Table 2.1: Some of the most important fields of DICOM standard

We see, in tab. 2.1 that a large part of the data are related to the patient and so there is the need to treat it as confidential and sensitive data. The

rest of the data are related to the technical characteristics of the acquisition process and it is important that a CAD system should consider them.

2.2.2 Image Preprocessing

This phase includes a set of preliminary elaborations of the image at low level with the purpose of improving the quality by reducing noise introduced into the acquisition step or correcting any artifacts due to patient motion; this last requirement is critical if the CAD system is developed for diagnostic tools with a long acquisition time such as the DCE-MRI. The average duration of an acquisition with the use of contrast media is 40–50 minutes and, even though the patient is immobilized in special breast-coils [41, 42], that gather the breast in a more or less rigid way, different artifacts may occur caused by breathing, by the different consistency of the breast tissue or even by the movements of the patient. So it is necessary to expand this stage by introducing a step of registration of the sequence of volumetric images in order to reduce the effect of these motion artifacts [43].
Image registration is a process which permits the transformation of the reference systems of two different images in order to compare them; for medical images the registration process permits to transform or align two images collected in two different time instants or produced by different instruments. There is a large variety of algorithms for achieving image registration [44, 45] and they can be classified in according to:

Rigid\Non-Rigid: a rigid transformation (or affine) provides a set of transformations that include rotation, scaling, translation, and other affine transforms. Affine transformations are not able to model all possible natural deformations, especially in cases such as movement artifacts of a soft tissue which is the breast: in these cases it is necessary an elastic (non-rigid) transformation. Recording techniques use some search algorithms with linear or cubic interpolation on models

of affine transformation or free-form-deformation (FFD) [46]. This transformations can be applied to two-dimensional surfaces or three-dimensional volumes [47, 48].

Intensity-Based\Feature-Based: methods based on intensity rely on similarity measures that take into account the brightness values of each pixel\voxel. Instead, the feature-based techniques, make use of target that use some fixed points, manually set or automatically searched, to make a recording basing the optimization on the Euclidean distance of these targets.

Mono-modal\Multi-modal: the "modal" refers to the kind of scanner\sensor with which the images are acquired and, in case of a multi-modal, the registration aligns and compares images obtained from different scanners. The multi-modal registration is widely used in the medical field of complementary surveys such as CT\MRI brain and PET\CT total body [45].

Spatial-domain\Frequency-domain: The spatial methods operate in the domain of the image, by comparing the characteristics and\or the intensity pattern. Instead, in the frequency domain, it is possible to apply the method of "phase correlation", which consists of rephrasing an image in relation to another. This phase-shift in the field of frequencies corresponds to an alignment that, unlike many algorithms in the spatial domain, is able to reduce the noise, the occlusion, and other defects typical of medical images.

2.2.3 Image Segmentation

Segmentation is the process of subdivision of an image in distinct Regions Of Interest (ROI — connected sets of voxels) that are homogeneous compared to a given characteristic [49] and this stage is the most critical because the

precision and quality of the segmentation results impact in a direct way on the performance of the next steps; the image segmentation can be achieved at different granularity levels:

- Air\Breast segmentation
- Parenchyma\Adipose\Glandular tissue segmentation
- Suspicious masses segmentation (fig. 2.2)

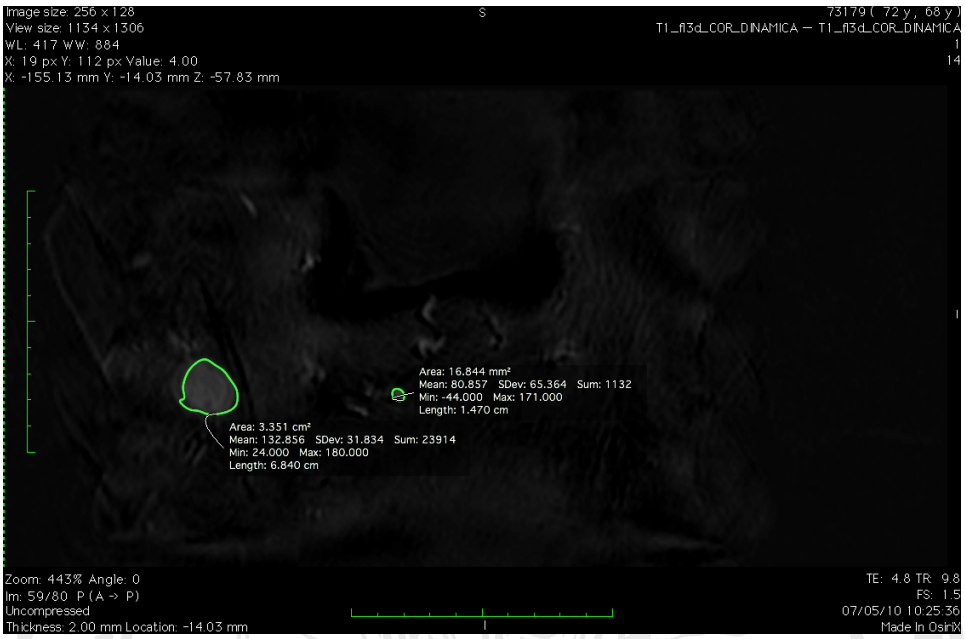


Figure 2.2: A MRI segmented in suspicious ROI in a CAD system view

The proposed segmentation techniques vary from the simplest thresholding segmentation based on signal intensity [30, 50] up to more advanced systems of segmentation by classification based on pattern recognition techniques and supervised learning, applied to a set of features [51, 52, 53].

2.2.4 Feature Extraction and Selection

With the techniques of feature extraction, the objects (regions identified in the previous phase and/or individual voxels) are represented by a vector of

parameters (feature) that provides a numerical projection of the features that are considered appropriate to describe the problem of classification [54, 55]. We can group the features according to the characteristics that they represent obtaining the five major classes of features for classification problems of neoplastic lesions by magnetic resonance imaging:

Class	**Characterization**
Dynamics Feature	Description of the temporal dynamics of the signal intensity through direct measures on TIC
Pharmacokinetic Feature	Description of some physiological parameters (Absorption, Distribution, Metabolism, and Excretion — ADME) of tissues, calculated on the basis of mathematical models
Morphological Feature	Description of the shape and structure of the regions obtained in the segmentation phase
Textural Feature	Description of the geometrical structure (texture) image through parameters that describe their statistical properties
Spatiotemporal Feature	Modeling of the signal in a four-dimensional space, representative of the temporal dynamics, the characteristics of the architectural and spatial variations of the voxels

Table 2.2: Classification of feature types

Depending on the classifiers used and on the purpose of the CAD system, the feature can be combined with each other to provide the information required to the classification. Then, among all the feature candidate, it is necessary to extract a small subset obtained through a phase of feature selec-

tion [56] in order to reduce the correlation between the feature and remove those proposed that do not provide decisive information in the classification step. Another object of feature selection is to improve the performance of the classifiers strongly influenced by the amount of data on input (Huges Phenomenon [57], over-fitting and computational complexity).

2.2.5 Classification

In this step, the CAD system collects the information provided by the previous steps expressing for each element (voxel or ROI and related features) a statistical classification (associating to each object a percentage that represents the statistical affinity of that object in any given class) and then expressing the final decision. CAD systems use different types of classifiers most of which, before being inserted in the operating flow of the CAD, require a training step that adapts the system to the specific problem trying to maximizing the ability of prediction and, especially, the capacity of generalization [58].
There are many different types of classifiers, each with its own peculiarities, that make it more or less suitable in specific matters. The theory of classification states that there is not the best classifier, that behaves better than the others (except for problems of very small or simple models and known perfectly). However, we can analyze the most used and eventually compare the statistical indexes obtained:

Multilayer perceptron (MLP) is a feedforward artificial neural network model that maps sets of input data onto a set of appropriate outputs. An MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function. [59]. MLP uses a supervised learning technique called backpropagation for training the network [60, 61]

Decision Tree (DT) is a flow-chart like structure in which internal node represents a test on an attribute\feature, each branch represents outcome of test and each leaf node represents class label (decision taken after computing all attributes). A path from root to leaf represents classification rule. This flow-chart is used as a predictive model in statistics, data mining and machine learning [62]. Training a Decision Tree is through greedy algorithms with a process of "top-down induction of decision trees" (TDIDT) [63].

Random Forest (RF) is an ensemble learning method for classification that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes output by individual trees [64].

Naive Bayes (NB) is a classifier based on the application of Bayes' theorem. It requires knowledge of a priori probabilities and the conditional probabilities related to the problem that, in case of medical diagnoses, are not available because of the variability of data (physiological heterogeneity of patients, morphological heterogeneity of patients and variety of pathologies).

Support vector machines (SVM) this classifier constructs a hyper-plane (or set of hyper-planes) in a high-dimensional space, which can be used for classification tasks. Intuitively, a good separation (binary thresholding) is achieved by the hyper-plane that has the largest distance to the nearest training data point of any class.

Adaptive Boosting (AdaBoost) is an ensemble learning meta-algorithm and it can be used in conjunction with many other learning algorithms to improve their performances [65]. AdaBoost is adaptive: it means that subsequent classifiers built are tweaked in favour of those

instances misclassified by previous classifiers. AdaBoost is sensitive to noisy data and outliers [58].

2.2.6 Evaluation

In this step the results of the classification step are taken into consideration with the aim of evaluating the final performance of whole diagnosis\detection system. For this purpose, in the design phase of the CAD it is taken into account the so-called "gold-standard" or "ground-truth" that, regardless of the aim of the system, represents the objective function of an optimization and reducing errors process. For segmentation and\or classification of tumor lesions, the gold standard is achieved by taking into account the medical report of some expert radiologists in the form of segmented ROIs, all histologically proved.
In the problem of binary classification, which could be a CAD system of separation of benign from malignant or even a CAD system highlights the suspicious tissue from non-suspicious, the performances can be evaluated through four key numerical parameters. Supposing to have two classes, respectively called "Positive" and "Negative", we define:

True Positive (TP) the number of elements belonging to the Positive class, classified by the system as belonging to the Positive class;

True Negative (TN) the number of elements belonging to the Negative class, classified by the system as belonging to the Negative class;

False Positive (FP) the number of elements belonging to the Negative class, classified by the system as belonging to the Positive class;

False Negative (FN) the number of elements belonging to the Positive class, classified by the system as belonging to the Negative class.

In the detection of neoplastic lesions, the most severe case is represented by the false negatives. This misclassification causes a delay in the diagnosis and so in the treatment of the disease and it can irremediably compromise the patient's health. Instead, the false positives are tolerated (if not excessive in numbers) because they make to start the diagnostic investigation in the same way as true positives do; further clinical examinations will ensure (or exclude) the pathology identified from CAD. A summary evaluation of the performance of the entire system can be expressed through statistical indexes (tab. 2.3) calculated from the validation parameters [66] set out above:

Indice	Formula	Descrizione
Positive predictive value (PPV - Precision)	$\frac{TP}{TP+FP}$	Expresses the probability that a positive result is actually connected to the disease (or a suspicious region)
Negative predictive value (NPV)	$\frac{TN}{TN+FN}$	Expresses the proportion of subjects with a negative test result that are correctly diagnosed
Sensitivity (SE)	$\frac{TP}{TP+FN}$	Provides an idea of the sub-segmentation, compared to ground truth
Specificity (SP)	$\frac{TN}{TN+FP}$	Provides an idea of the over-segmentation, compared to ground truth
Accuracy (ACC)	$\frac{TP+TN}{\mathrm{TP+TN+FP+FN}}$	Provides an idea of the overall system performance
Error (ERR)	1-ACC	Provides an estimate of the error (class independent) committed by the system

Table 2.3: Statistical indexes for evaluating the performance of classification systems [66]

Instead, if we want to estimate how much the chosen predictive model generalizes the problem in an independent way from the data, it is necessary to apply the statistical technique of cross-validation. Each round of cross-validation involves partitioning a sample of data into complementary subsets, performing the analysis on one subset (called the training set), and validating the analysis on the other subset (called the validation set or testing

set). To reduce variability, multiple rounds of cross-validation are performed using different partitions, and the validation results are averaged over all the rounds. There are different types of cross-validation: in the K-fold cross-validation, the original dataset is partitioned into K equal size sub-samples; the Leave-one-out cross-validation uses a single observation from the original dataset as the validation data, and the remaining observations as the training data; in medical field it is useful to use an approach leave-one-patient-out in which each step of the cross-validation uses an entire patient dataset as a validation set and the remaining data as the training data [67, 68].

2.3 Diagnostic tools and CAD architectures

To obtain the diagnosis of a patient using a DCE-MRI, the radiologist must analyze the three-dimensional volume slice by slice, detecting the regions that absorb more contrast media and then to deep the dynamics of these absorptions evaluating the curves voxel by voxel or with average values and finally, comparing them with the sample curves. To achieve this goal, even manually, there are DICOM files manipulation and processing softwares that help the doctors to write down a diagnose. CAD systems automate these tools implementing, in a more or less complete way, the steps so far described and they provide a final diagnosis which only needs to be investigated and contextualized by the radiologist.

2.3.1 OsiriX

OsiriX [12] is an image processing application for Mac dedicated to all DICOM images produced by a wide range of medical equipment such as MRI, CT, PET end others. OsiriX has been specifically designed for navigation and visualization of multimodality and multidimensional images providing 2D views, 3D views, and correlating the temporal dimension of dynamic

investigations by providing a 4D visualization. The latest versions allow a comparison on a fifth dimension of surveys from different instruments obtaining, in addition to temporal correlation, even mutual information in a single 5D view.

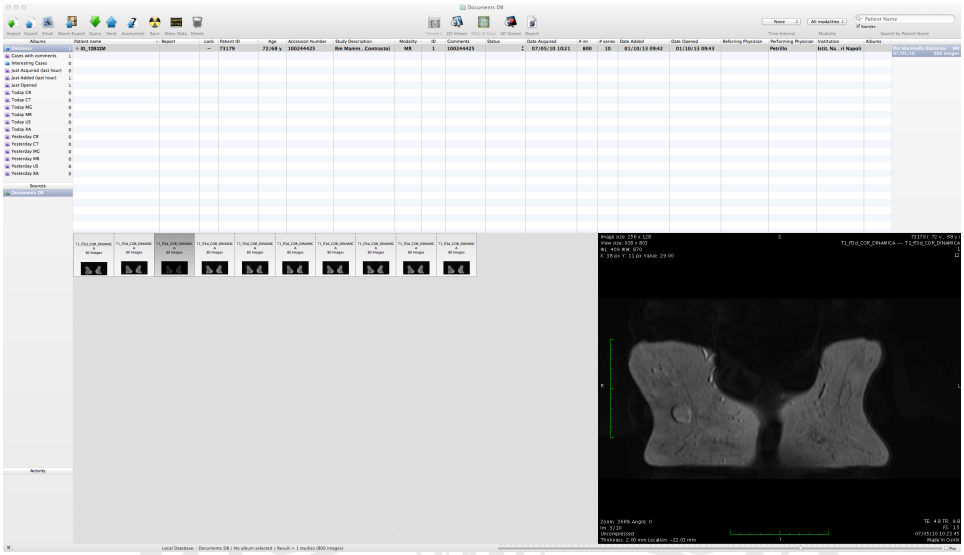


Figure 2.3: OsiriX application window screenshot

The capabilities are required to OsiriX as a support tool in the specific case of the diagnosis of breast cancer, although OsiriX does not directly implement decision-making phases such as classification or segmentation, are as follows:

Switch of projections: allows to rotate the volume respect to the different projections (frontal\coronal, sagittal\median, longitudinal\transverse);

Multiplanar reconstruction (MPR): allows to calculate and display of arbitrary planes to assess the tissues or lesions in their respective local symmetries;

Maximum intensity projection (MIP): provides a three dimensional representation fast and light from the computational point of

view by projecting the maximum value of the signal intensity on an imaginary plane;

Representation of the temporal dynamics (TIC, average TIC): considering a single voxel the software represents the curve of variation of the signal intensity over time (TIC); considering a ROI allows to show the average trend of enhancement curves;

Segmentation by thresholding: defining a variable threshold to segment the software allows in simply and directly an image in grayscale by providing a binary image;

Contrast adjustment: the software allows to stretch, shift and equalize the histogram of the image in an arbitrary way to highlight the areas with different contrast variation;

Subtractive image: is one of the most important and crucial, among functionality provided by OsiriX, because it allows to highlight at a glance the effect of the contrast medium by subtracting voxel by voxel the values of two different time series and obtaining a third volume in which the differences are shown as an image. The protocol considered in this work is the result of an acquisition made in ten temporal instants: one performed before the injection of the contrast medium (pre-contrast series) and the remaining nine acquired at a time interval of 56 seconds each other (post-contrast series). Generally the maximum absorption peak of the para-magnetic medium is obtained into the 4th instant post-contrast. Labeling the temporal instants with T_i with $i = 0$ for the pre-contrast serie and $t = [1, 9]$ for the post contrast series, the practice provides that a very useful subtractive image is obtained by subtracting the 4th instant post contrast (fig. 2.5) with the pre-contrast serie (fig. 2.4).

$$T_1 = T_4 - T_0$$

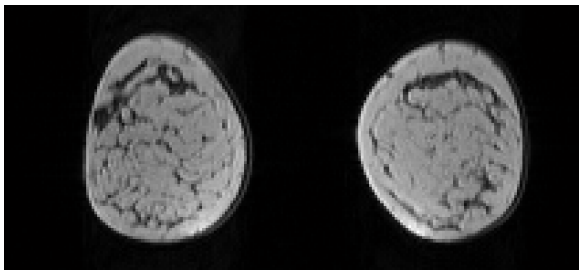


Figure 2.4: A precontrast DCE-MRI slice

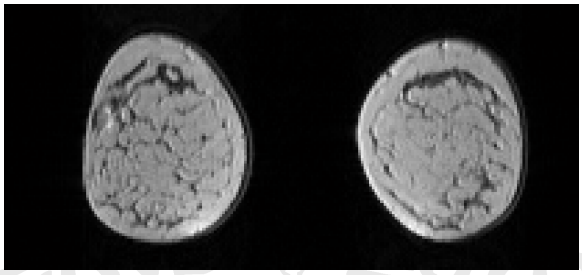


Figure 2.5: A postcontrast DCE-MRI slice

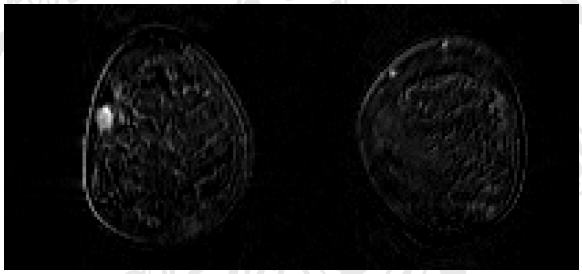


Figure 2.6: A subtractive DCE-MRI image

Another important feature of OsiriX is the existence of an Software Development Kit (SDK) available to developers and the scientific community allowing to design and implement plug-in and then extend the basic functionality of the software with procedures of utility or research applications in the medical and computing fields.
The plugins, that can be developed in OsiriX, are different types depending on the context in which they are used:

imageFilter: allows to apply filters to the data or to the current view;

roiTool: allows the development of tools to generate or manipulate the Region Of Interest;

fusionFilter: allows to operate among different representations of the same study considered or to combine different information of different diagnostic instruments;

database: allows to create plugins useful to manipulate the diagnostic studies in their entirety by providing storage and cataloguing capabilities;

other: allows to implement functionality not covered by the above types.

2.3.2 Local Architecture

A CAD system, that develops the whole diagnostic process, can be provided as an independent service and directly accessible from the radiologist workstation. The software will implement a diagnostic algorithm that includes some of the steps described above: starting from the DICOM file, it will produce a result that the doctor can use to support a comprehensive and well-structured diagnosis; the final response will still result of medical assumptions and based on professionalism, competence and experience of an experienced radiologist.

2.3.3 Remote Architecture

The medical and scientific progress is increasingly based on complex and constantly evolving algorithms which often require a significant computational load not always available on the conventional workstations; in addition to the problem of computational load, there is the necessity of updating the

workstations of each radiologist to comply to the new diagnostic algorithms validated by the scientific community in a regular way; the best solution, to reduce the problems described above, is to distribute the architecture of the diagnostic process.

Scheinine et al. [69] propose an Object-Oriented Client-Server system for interactive segmentation of medical images based on JAVA for the client and CORBA for the distributed system, connected by a TCP/IP socket protocol. Mayer et al. [70] implemented a "processing on demand" client-server architecture for 3D image processing in which the computation load is all on the server side, while the client requests the desired images one slice (2D) at a time. At last, Sherif et al. [71] present an evolution of the open standard DICOM to support communication between DICOM entities over a TCP/IP network. The above cited papers were mostly designed for computing systems belonging to several generations ago, unable to handle even simple tasks like 3D visualization. Moreover, their focus are on the architecture, omitting an evaluation about the severe and complex security and privacy issues associated with the cloud computing [72].

Information Security

Information security is the protection of information and minimises the risk of exposing information to unauthorised parties [73]; moreover its the protection of information and information systems from unauthorized access, use, disclosure, disruption, modification, or destruction in order to provide confidentiality, integrity, and availability [74]. When we face the problem of the transmission of medical data over the network, we encounter the processing of sensitive data related to personal information and health conditions of the patients, moreover it is required that the system presents characteristics of the service availability and reliability [75, 76]. To achieve the aim of securing the use of the service, we use encryption techniques, secure protocols and

techniques for dependability assessment to increase the confidence about the robustness of the designed solution before putting it into operation and to detect and remove possible critical leaks [77]

Networks

Distributed systems heavily rely on the state and quality of network interconnections. In professional systems, such as hospitals and scientific centers of research, often are available an infrastructure of interconnect of an high quality. In Italy, for example, the GARR Consortium (Gruppo per l'Armonizzazione delle Reti della Ricerca) interconnects 90 Italian universities, 145 research laboratories and 54 hospitals active in research and it provides a connection bandwidth ranging from 1Gbps up to 20 Gbps. In professional environments which do not adhere to this type of infrastructure it often uses of MPLS connections that reach the order of 10-100Mbps uplink.

Considering the state of the art infrastructure to connect, we must ensure that the overhead added to the transmission system does not weigh significantly (in relation to the processing time and the speed-up achieved on a server running adequate computational capabilities). To ensure certain safety criteria, encrypted connections are widely used with the TLS\SSL standard protocol (more specifically discussed in next chapters) and eventually a data compression that can reduce the size of the payload and thus the transmission time.

The tables 2.4, 2.5 and 2.6 below compare the transmission times and throughput on three networks with different bandwidths by varying the parameters of the transmission protocol with the presence or absence of encryption of the session using SSL, and with a data compression (performed with a standard compression level and detailed discussed in next chapters).

SSL	Zip	Byte	Sec	Mbps
No	No	111275606	5,101	21,814
No	Yes	32028169	1,234	25,955
Yes	No	111275606	6,345	17,538
Yes	Yes	32028169	1,847	17,341

Table 2.4: Transmission rate on LAN 10/100 Mbps

SSL	Zip	Byte	Sec	Mbps
No	No	111275606	11,615	9,580
No	Yes	32028169	3,462	9,251
Yes	No	111275606	12,927	8,608
Yes	Yes	32028169	3,756	8,527

Table 2.5: Transmission rate on British Telecom MPLS 10 Mbps

SSL	Zip	Byte	Sec	Mbps
No	No	111275606	72,521	1,534
No	Yes	32028169	21,792	1,470
Yes	No	111275606	102,159	1,089
Yes	Yes	32028169	31,742	1,009

Table 2.6: Transmission rate on SHDSL 4 Mbps

3

Proposed architecture

In this chapter it is shown the proposed architecture for final implementation of the CAD system. The main requirement, taken into account at design time, was that to preserve the hardware and software tools currently in use at medical facilities to help the integration and fruition of the offered service. The large part of radiologists uses OsiriX as a tool for handling, processing and evaluation of magnetic resonance imaging (and various other medical investigations that are not subject to this work). So it is required, by using the potential offered by OsiriX SDK extension, to keep this tool as a user's primary interaction. The architecture requires the use of a plugin specifically developed to make available the functionality provided by the software. For distribution of services, a client-server module, implemented in Java, developed using the file system provided by the operating system and it communicates with the Plug-in collecting medical data to be analyzed and returning the final response of the system.A whole segmentation flow implemented, tested and validated on server side (named BLADeS - Breast Lesions Automatic Detection System); using the numerical computing environment MATLAB 2012a and based on the criteria analyzed in Chapter 2, it provides a detection of breast lesions by recognizing suspicious ROI in the volume of DCE-MRI.

Next image shows a system overview grouped into its main components.

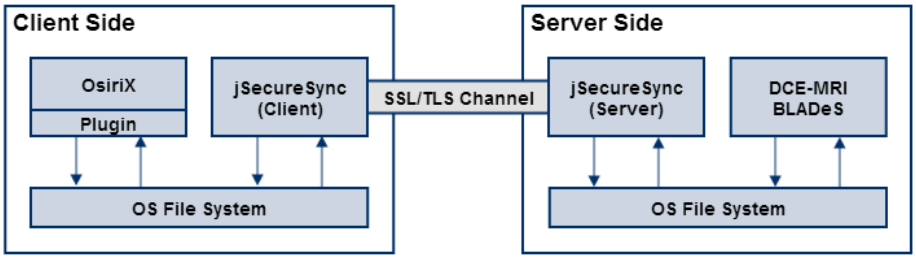


Figure 3.1: Proposed system overview

In order to simplify understanding of the contents, the architecture can be separated into two services by dividing the problem of segmentation (shown in §3.1) from the problem of the transmission and distribution (shown in §3.2).

3.1 Segmentation flow

The segmentation of suspicious lesions in DCE-MRI is the core of the CAD system. Working on volumes and multidimensional arrays, it was natural to choose a vector computing environment as MATLAB achieving a consequent optimization in time and flexibility by application of mathematical, morphological and array-level operations.
Starting from the concepts discussed in Chapter 2, it was designed a complete operation flow to achieve a segmentation of the volume through successive refinements until to have a set of voxels to consider suspicious; so the set of suspicious ROI to provide to the radiologist is obtained from the closure of all the voxels that are considered suspicious in compact and disjoint volumes. Therefore, the CAD system implemented in MATLAB is a Computer-Aided Detection system that implements all the steps (in fig. 2.1) till the detection of neoplastic masses leaving to radiologist the last interpretation of the aggressiveness of each lesion. Due to the configuration as a "cascade of operations" chosen for the detection system, it is easy to add

a next additional stage to achieve an automated diagnosis (object of future research, see §5) to transform the system by adding the capabilities of a CAD diagnosis tool.

3.1.1 Segmentation steps

In the proposed architecture, the general steps, provided by a CAD system, are complemented and structured to make a better gradual segmentation and basing choices made on well-validated criteria.

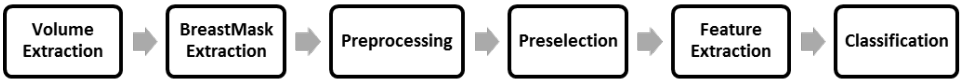


Figure 3.2: Proposed system flow

In the specific:

Volume Extraction: the first operation performed by the system is to extrapolate the breast scanned volume from the DICOM file, storing the whole information in a 4-dimensional Matlab matrix (128 pixels x 256 pixels x 80 slices x 10 acquisitions). The files that constitute the DICOM format occupy about 110 MB and they are collected in a single Matlab variable that occupies only 27MB. The excellent memory management justifies in part the choice of the computing environment.

BreastMask Extraction: the first approximation of the segmentation is performed dividing the breast parenchyma (whatever it is) from the air that surrounds it and from the tissues of the pectoral muscle. The aim is to simplify the next stages by reducing the computation due to the large amount of voxels with a ratio of about 8:1. The voxels that represent the volume of the breast, and then the only need by the next step, decreased from about 3 million to approximately 350,000. The result of this step is a binary mask that represents the only breast voxel (Breast-Mask - BM);

Preprocessing: this phase collect all the operations need to prepare the volume to the next stages. In particular, an image registration over the time instants is made in order to reduce noise effects introduced by any voluntary and especially involuntary movements of the patient;

Preselection: a second approximation of the results is obtained by evaluating voxel by voxel signal intensity and considering (in an indirect way) the maximum value of absorption of the contrast medium. This voxel-based technique provides a set of suspicious voxel and it is often used for a segmentation of the first approximation;

Feature Extraction: this step is necessary to complete the classification and it is only performed on the set of dynamic features selected at the design stage. This reduced subset decreases the processing time of the classification phase and the time need to feature extraction itself;

Classification: the most of the work involved the selection and optimization of the used classifier: a Support Vector Machine (SVM) trained to work on dynamic features. The output of the classification step is the union of all voxel labeled as suspect that represent suspect ROIs that will be subject to advice of an expert radiologist.

The validation of the results was performed considering some changes in some modules (specifically: preprocessing, preselection and classification modules) and evaluating the final results of the algorithm considering it as a closed box. The changes taken into consideration will be adequately treated in all their details in the corresponding sections; the table 3.1 shows for each module the changes taken into account.

Preprocessing	Preselection	Classification
NoReg	NoPres	MLP
MEDx3	SiPres	RF
MlReg		SVM

Table 3.1: Changes in some modules of the segmentation flow

3.1.2 BreastMask Extraction

The problem of segmentation of the breast tissue faced in this phase, besides reducing the number of voxels as load to the later stages, allows to obtain a first approximation of the result by selecting the only voxels relative to the breast tissue; other voxels, related to the background and other tissues adjacent to the breast, are not of interest for this study and they present problematic results and noise caused by the different physical nature of the material (air or other organic tissue). In the table 3.2 we see that the breast, identified by a BreastMask (a binary mask that label only the voxel bellowing to the breast tissue), takes on about 12% of the volume of extracted data.

Patient ID	**# BM Voxel**	**% of volume**
b2	127095	4,8%
b3	115648	4,4%
b4	460360	17,6%
b5	336288	12,8%
b6	374831	14,3%
b7	379723	14,5%
b8	401638	15,3%
b9	375400	14,3%
b10	115472	4,4%
b11	571868	21,8%
b12	311180	11,9%
b13	296293	11,3%
b14	231588	8,8%
b15	269521	10,3%
b16	235380	9,0%
b17	170448	6,5%
m1	229647	8,8%
m2	139038	5,3%
m3	222426	8,5%
m5	352572	13,4%
m7	281248	10,7%
m8	264958	10,1%
m9	458960	17,5%
m10	350253	13,4%
m11	541752	20,7%
m12	693395	26,5%
m13	364325	13,9%
m14	362404	13,8%
m15	297452	11,3%
m16	351252	13,4%
m17	370548	14,1%
m18	168715	6,4%
m19	630574	24,1%
m20	312787	11,9%
m21	140557	5,4%
average	323017	12,3%

Table 3.2: Percentage of the breastmask

In literature, several BreastMask segmentation algorithms have been proposed [78, 79], some of them [80, 81] based on Otsu thresholding [82]. The procedure of breast-mask extraction method applied in this flow, it is based on already done work [83] and it is divided into three main steps:

- Preprocessing (fig. 3.3),
- Otsu thresholding (fig. 3.4),
- Morphological Refinements (fig. 3.5).

The starting volume is the 3D data obtained as the minimum signal intensity (SI_{min}), along time, of the original 4D volume value. The pre-processing step uses min-max normalization in order to avoid sub-segmentation in slices that have small range of values. As last step, in order to smooth mask borders and fill internal holes, a cascade of morphological operators is applied in this order: "Closing" and "Holes filling" applied on global volume; "Erosion" applied slice by slice in coronal view direction; "Dilatation" applied on the global volume.

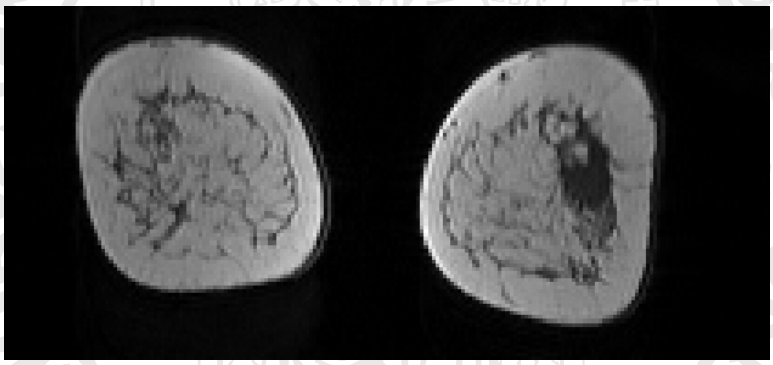


Figure 3.3: A slice of a DCE-MRI

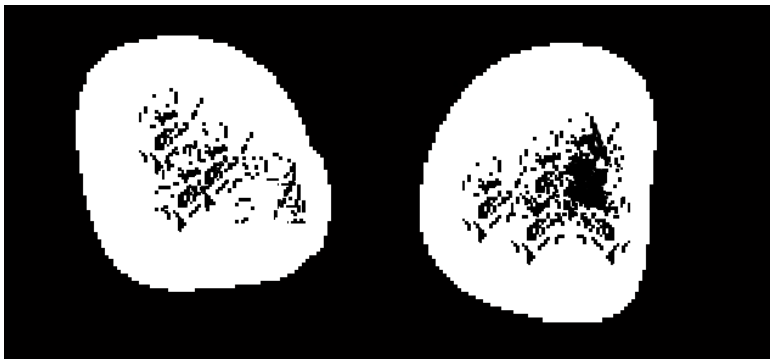


Figure 3.4: A slice of DCE-MRI after Otsu thresholding

Figure 3.5: A slice of a Breast Mask of DCE-MRI

3.1.3 Preprocessing (Image Registration)

The DCE-MRI is a scan that has a duration of several minutes (about 40 minutes) and therefore it is affected by so-called motion artifacts that are ghost images due to accidental disturbances occur during scanning:

- Cardiac Pulsation
- Anatomical structures in constant motion (heart, lungs, blood vessel walls, throat, eyes)
- Breath
- Other involuntary movements

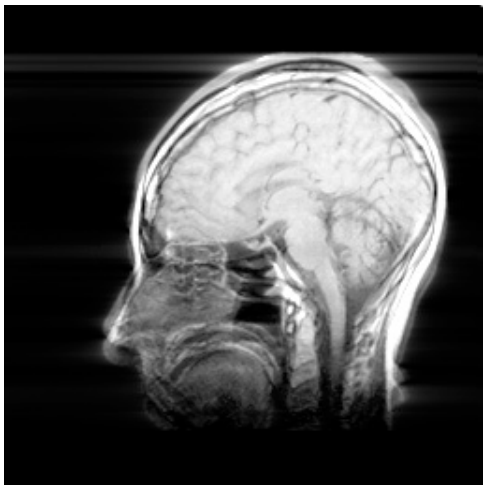


Figure 3.6: Brain DCE-MRI showing some motion artifacts

In the case of breast DCE-MRI motion artifacts are limited to the soft component of the breast tissue itself. In compensation the supine position on special brest-coils (which also improve the detail obtained from normal detection coils - fig. 3.7) restricts the movements of the patient, reducing the motion artifacts.

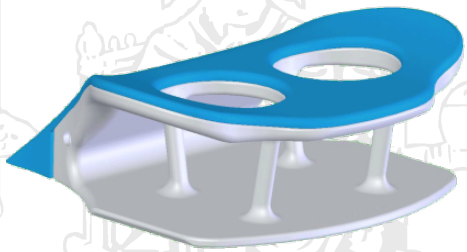


Figure 3.7: A model of breast-coil used in DCE-MRI

In any case, since the deductions made in the later stages are based on the dynamic component of each voxel it is necessary that these movements are eliminated or, at most, attenuated. Without a correct registration, the TIC curve constructed into a precise spatial coordinate of the volume may refer to different voxel and show a lack of correlation due to the displacement of the tissues.

There are several approaches to the recording of images. In the first approximation it is possible to consider the effect of movements of the patient as an additive noise and then it is possible to apply a median filter to reduce such noise, this filter introduces a loss of information due to the weighing. More advanced techniques, as already mentioned in section 2, provide morphological transformations obtained by successive refinements.
The effects of registration can be estimated with appropriate similarity measures such as:

Measure	Formula	Description
Root Mean Square Error (R-MSE)	$\frac{1}{n}\sqrt{\sum(I(t_0) - T(I(t)))^2}$	Proportional to the local spatial distance of the two images.
Normalized Cross Correlation (N-CC)	$\frac{\sum(I(t_0) - \bar{I}(t_0))\sum\left(T(I(t)) - T(\bar{I}(t))\right)}{\sigma_I \sigma_{T(I)}}$	Estimate the distance between two images comparing compared to the standard deviation, and then attenuating the effect of a different distribution of the histograms due to the effect of the contrast medium

Table 3.3: Statistical indexes for evaluating the performance of registration systems [66]

We compared different techniques of image registration:

NoReg: no image registration made;

MEDx3: median filter applied slice by slice on a window of 3 pixels x 3 pixels;

MLReg: rigid registration obtained by the image processing toolkit of Matlab 2012a [84]

RUReg: non-rigid registration (elastic) obtained by applying the Rueckert algorithm [46]

Applying the image registration on ten patients and calculating average indexes, we obtained preliminary results (tab 3.4) useful to choose the most suitable technique to include in the architecture.

Techniques	**Exec. Time (sec)**	**R-MSE**	**N-CC**
NoReg	-	0,08479	93,78%
MEDx3	6,235	0,09565	91,66%
MLReg	402,73	0,0803	94,56%
RUReg	11047,64	0,07679	95,16%

Table 3.4: Preliminary registration results

We can demonstrate the usefulness of this registration process and to assess the quality of the results obtained by visually comparing the result of a subtractive image (fig 3.8). If we display the trend of TIC in the proximity of a lesion (in the red point) before (left of pic. 3.9 and after (right of pic. 3.9) the image recording, we can see that the registration fixes the TIC's points that now is similar to II sample curve shown in fig. 1.8.

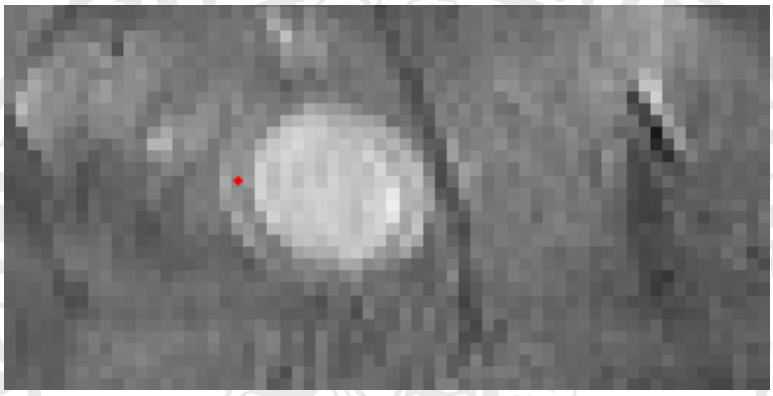


Figure 3.8: Motion artifact on a breast subtractive image; in red a critical voxel

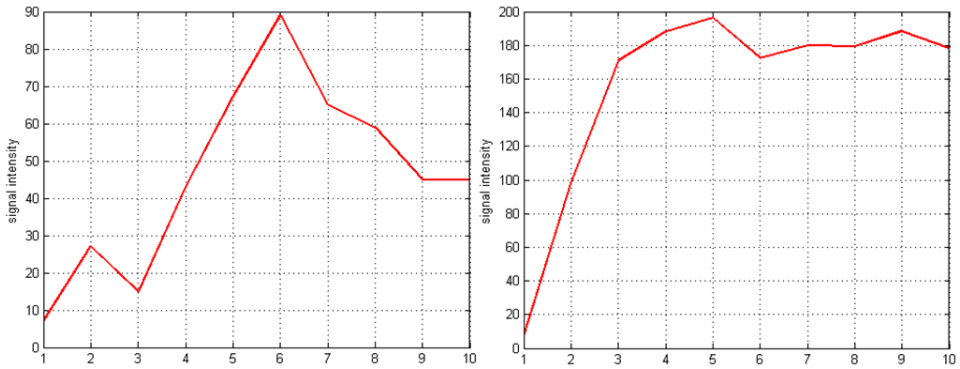


Figure 3.9: Effect of a registration shown on a voxel tic

Although the numerical (Tab. 3.4) and visual results (fig. 3.10) of the elastic registration are the best achievable with the instruments provided by the scientific literature, the execution time (about 3 hours) of such approach constitutes a blocking implementative limit and, to proceed to realization of a CAD, we examined the only two approachable techniques. Surely it is important to integrate the "non-rigid" registration technique and it is currently interest of study (see §5).

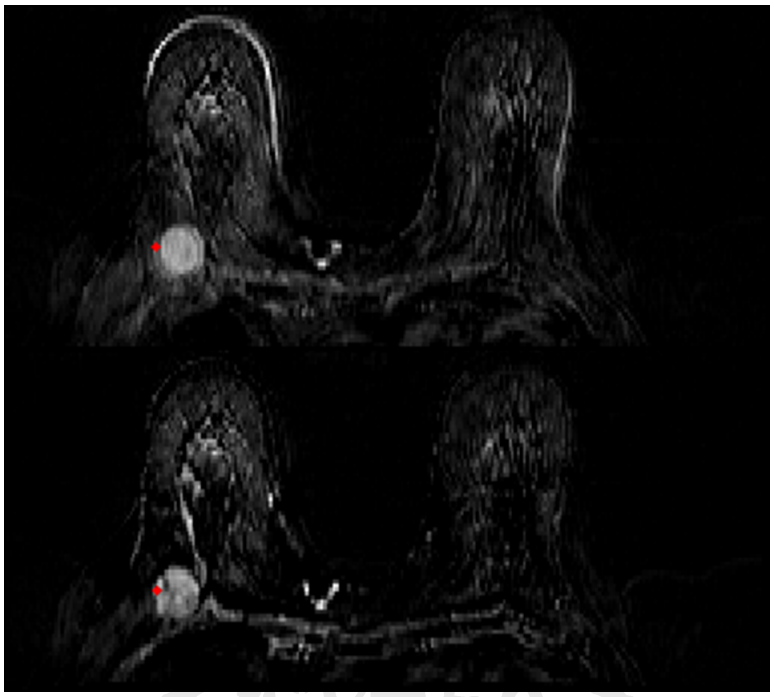


Figure 3.10: Visual result of elastic images registration

3.1.4 Preselection

Many segmentation works based their results on a simple thresholding carried out voxel by voxel and obtained by evaluating in a direct or indirectly way the values of the signal intensity [85, 86]. In order to get a better estimate of the value of absorption of the contrast medium, it is necessary to evaluate the enhancement curve in terms of Relative Enhancement (RE) that is the temporal evolution of the signal relatively to the value of signal intensity of the voxel in the pre-contrast serie; in this way the curve still models the absorption of the contrast agent along time, but it refers (in a first approximation order) to mmol/kg.

$$RE(k) = \frac{TIC_{post}(k) - TIC_{pre}}{TIC_{pre}}$$

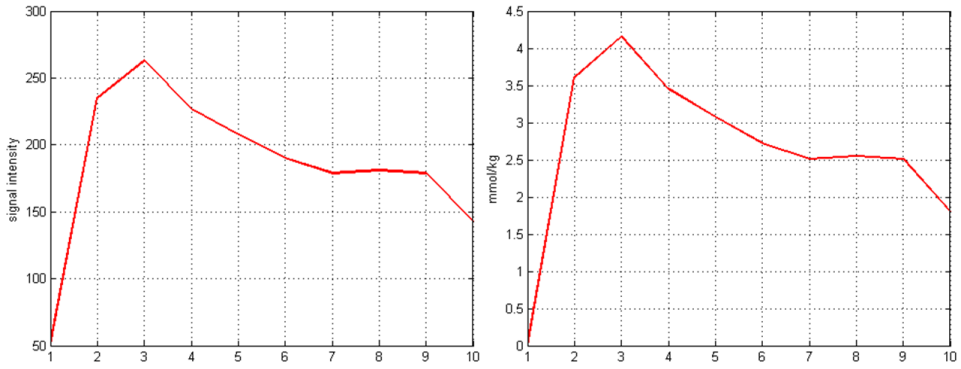


Figure 3.11: Comparison of a Signal Intensity TIC (on the left) and a Relative Enhanced TIC (on the right)

The results of a segmentation based on the cited work show [86] a very high sensitivity in some cases very close to 100% [85]. A so high sensitivity is indicative of a reduced fraction of false negatives and it is obtained by thresholding a very low 30% [85] of enhancement of precontrast image. By contrast we have an over-segmentation quite marked that makes the results an excellent first level approximation if compared to a precise segmentation of the lesions. We decide to use the technique with a more conservative thresholding on enanchement, about 30% [85] (fig. 3.13) applying the simplest version of the proposed method, that is to evaluate only the thresholding and ignore all the conditions for suspiciousness added from the work; this last precaution allows to achieve a sensitivity of 100% in both the proposed methods (fig. 3.12 and 3.13)

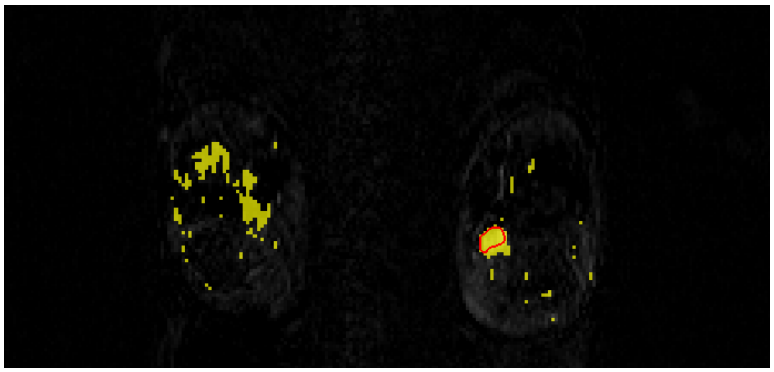


Figure 3.12: Preselection obtained by implementing part of [86]

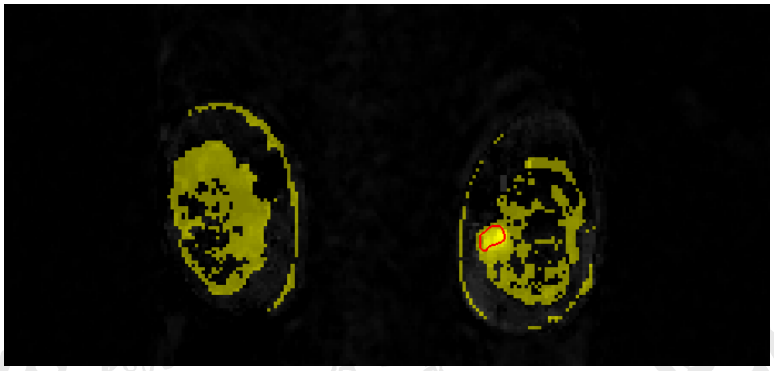


Figure 3.13: Preselection obtained by implementing part of [85]

Looking at the above images we see that the tumor indicated by the gold-standard is always in the preselection mask. We can also see the added value obtained through the pre-selection step by comparing the average over the time of the TIC curves contained in some specific regions (fig. 3.14).

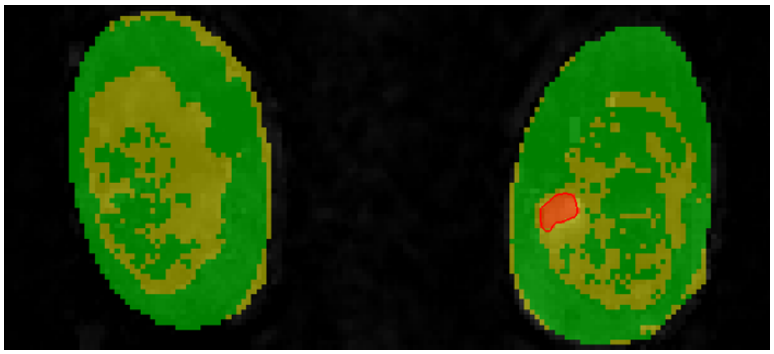


Figure 3.14: Result of preselection (in green), manual segmented roi (in red) and all other voxel of BreastMask (in green) overlapped on a MRI

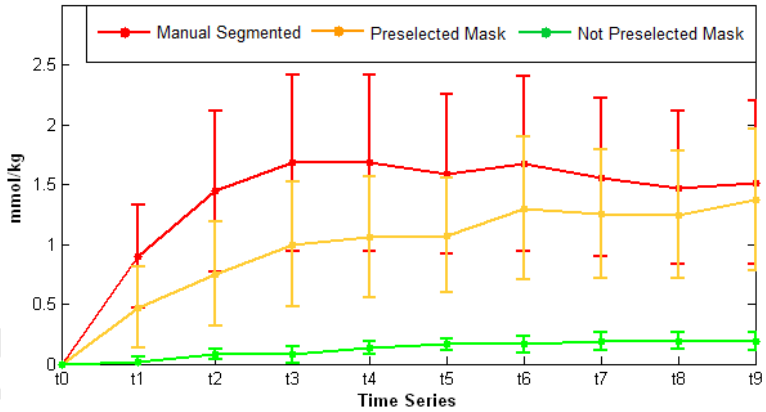


Figure 3.15: Some average tics in different regions: manual segmented, preselected and whole breastmask (shown in fig. 3.14)

In the graph 3.15 we have represented the mean value and the standard deviation of the following curves:

- In red the curves contained in the ROI manually segmented by the doctor and that, therefore, represent our gold-standard;
- In yellow the curves preselected using the designed module;
- In green the rest of the curves contained in breastmask.

It is interesting to note that the curves of the ROI manually segmented by the doctor, as well as the preselected curves, show a high variance and thus they

could lead to an overlap. If we try to train a classifier to distinguish between the red lines and the green lines, it is very probable that the selected features, in a feature reduction, and the training would produce a binary classification very close (in the results) to a segmentation with thresholding dispersing the discriminant power of the classifier. Essentially the preselection step reduces the field of action of the classifier to a much more complex problem, which is to distinguish the curves in red from those in yellow.

3.1.5 Dynamics Feature in DCE-MRI

The proposed system uses a set of model-free features that describe the problem by analyzing the dynamic evolution with some direct measurements on the enhancement curve (fig. 3.16).

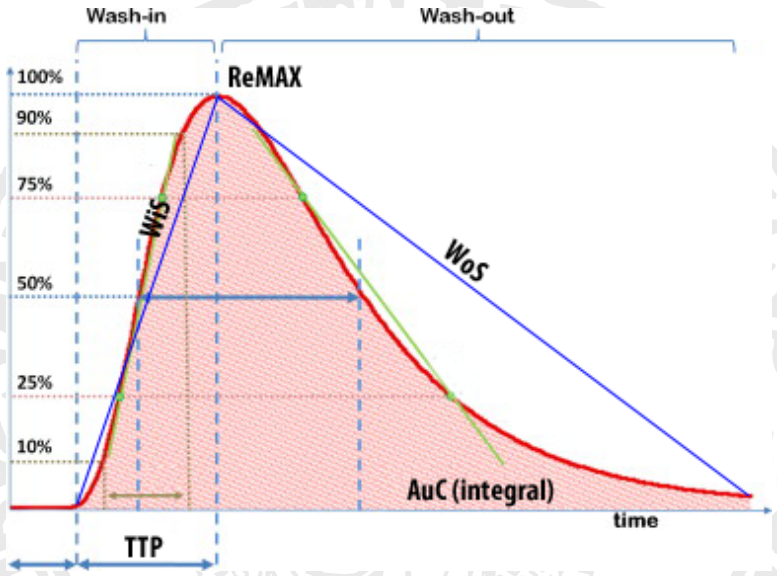


Figure 3.16: Some dynamics feature on a TIC

The next table contains all the feature considered in the problem of detection and\or diagnosis of breast cancer.

Feature	Description	Formula
AUC	Measuring the area under the intensity curve	$\int_0^{+\infty} SI(t)dt$
AUC_{TTP}	Measuring the area under the intensity up to the time of peak intensity	
CEI	Contrast enhancement image - Measuring the maximum absorption of the contrast agent during the first minutes of acquisition τ (τ=2-3 min)	$\max_{0 \le t_i \le \tau} (SI(t_i) - BS, 0)$
CT	Curve Type - Describes the kind of curve	Persistent (I); Plateau (II); Washout (III)
$\frac{dSI(t_i)}{dt_i}$	Derivative of signal intensity	$\frac{SI(t_{i+1}) - SI(t_i)}{t_{i+1} - t_i}$
$E(t_i)$	Measuring the Enhancement of the signal intensity	$SI(t_i) - BS$
ES_{var}	Enhancement Slope Variance - Measuring the contrast agent activity	$Var\left(\frac{E(t_i)}{t_i - t_{inj}}\right) \quad i = 1, \dots, N$
ES_{sd}	Enhancement Standard Deviation	σ_{SI}
MITR	Maximum Intensity Time Ratio - Measuring the temporal behaviour of the changes in signal intensity	$\frac{E_{max}}{TTP - t_{inj}}$
APR-MITR	Adjacent-Peripheral Region - Measuring the ratio between the MITR of different sections of a lesion (central, middle, and adjacent peripheral)	$\frac{MITR(adjacent)}{MITR(peripheral)}$
PCR-MITR	Peripheral -Central Region - Measuring the ratio between the MITR of different sections of a lesion (central, middle, and adjacent peripheral)	$\frac{MITR(peripheral)}{MITR(central)}$
NMITR	Measure the temporal behavior of the changes in signal intensity	$\frac{RE_{max}}{TTP - t_{inj}}$
POE	Pattern of enhancement - Describes the kind of enhancement	No enhancement; Centrifugal; Centripetal; Heterogeneous; Homogeneous; Ring like; Ductal; Flat shape; Rim
PI	Perfusion index - Summarized in a single index the temporal trend of the whole curve	$\frac{SI_{max}}{SI_{max-index}} + \sum_{i=1}^{N} RE(t_i)E(t_i) + \sum_{i=1}^{N} SER(t_i)\big(SI(t_i) - SI(t_N)\big)$
PSC	Post initial signal course - Evaluate the type of curve ICT according to the value obtained defines three types of curves: rising (increase>10%); plateau (between -10% and +10%), washout (decrease>10%)	$\frac{SI(t_N) - \max_{t_i=\{t_a,t_b\}} SI(t_i)}{\max_{t_i=\{t_a,t_b\}} SI(t_i)}$ $t_a = 110s; \; t_b = 220s$

Feature	Description	Formula
$RC(t_i)$	Rate of change - Measuring the change in signal intensity between consecutive instants	$\frac{SI(t_i) - SI(t_{i-1})}{SI(t_{i-1})}$
RC_{max}	Rate of change at the maximum point	$\max_t RC(t)$
$RE(t_i)$	Relative enhancement	$\frac{E(t_i)}{BS}$
RE_{max}	Relative enhancement at the maximum point	$\max_t RE(t)$
N-$RE(t_i)$	Relative enhancement normalized	$\frac{RE(t_i) - \mu_{RE(t_i)}}{\sigma_{RE(t_i)}}$
$RES(t_a, t_b)$	Relative enhancement slope	$\frac{RE(t_b) - RE(t_a)}{t_b - t_a}$
$RSI(t_i)$	Relative signal intensity	$\frac{SI(t_i)}{BS}$
RSI_{max}	Relative signal intensity at the maximum point	$\max_t RSI(t)$
$RSI_{dec}(t_a, t_b)$	Relative signal intensity decrease	$\frac{SI(t_a) - SI(t_b)}{SI(t_a)}$
$RSI_{decttp}(t_a)$	Relative signal intensity decrease from the peak	$\frac{SI_{max} - SI(t_a)}{SI_{max}}$
$SER(t_i)$	Signal enhancement ratio	$\frac{E(t_i)}{E(t_N)}$
BS	Signal intensity at baseline - Measuring the signal intensity before dosing of the contrast agent; in the case multiple precontrast measurements it makes the average	$SI(t_0)$
$SI_R(t_i)$	Signal intensity with R normalization	$\frac{BS}{SI(t_i)}$
$SI_\Phi(t_i)$	Signal intensity with φ normalization	$\frac{2}{\pi}\tan^{-1}\left(\frac{SI(t_{i+1} - SI(t_i))}{t_{sampling}}\right)$
SOD	Sum Of local Differences - Measure the separation from the background	$BS + \sum_{i=1}^{N} \lvert SI(t_i) - SI(t_{i-1}) \rvert$
TTP	Time to peak - l'istante di tempo al quale corrisponde il massimo segnale	$SI_{max-index}$
WIS	Wash-In Slope - Measure the slope in the area of absorption (by the dosing of the contrast agent to the point of maximum)	$\frac{RE_{max}}{TTP}$
WOS	Wash-Out Slope - Measure the slope in the area of washout (from the point of maximum until the end of measurement)	$\frac{RE_{max} - RE(t_N)}{(t_N - TTP)}$

In order to reduce the overfitting and shorten training time is necessary to extract a subset of the features listed above. For the choice of this subset we relied on some experienced radiologists advice who recommended as five main features that appear to be the most descriptive of the problem to be segmented:

Area under TIC (AUC): total amount of contrast agent absorbed;

Relative Enhancement at Maximum Point (ReMax): contrast agent peak;

Time To Peak (TTP): the time in which ReMax is gained;

Wash-In Slope (WIS): angular coefficient of linearized approximation of TIC curve from time 0 to TTP;

Wash-Out Slope (WOS): angular coefficient of linearized approximation of TIC curve from time TPP to last time.

3.1.6 Training and Classification

In this section we enumerate the used classifiers and we describe the training techniques and the results validation methods used.
Specifically the following classifiers are used:

- Multilayer perceptron (MLP) methods of validation of results
- Random Forest (RF)
- Support vector machines (SVM)

All classifiers are trained considering the subset of the dynamic features consisting of the five feature (shown in fig 3.16) described in the previous section.

The training phase, being at the end of the whole process, was used for the tuning of the full flow by varying all possible combinations in the modules; All the combinations taken into account are in the table 3.1 and include the possibility of varying three options of image registration, the absence or the presence of the pre-selection step and three different classifiers for a total of 18 possible combinations.
Using a cross-validation technique such as leave-one-patient-out in order to ensure a good statistical independence of the results obtained, we were able to test every single combination getting for each one of them statistical indicators of complete segmentation system.

3.2 Distributed architecture

The service distribution architecture is designed to meet the need to provide the developed protocol, designed in the previous section (§3.1), in an operational and practical way and especially it manages privacy and security; then the system requirements are the follows:

- Ensure privacy of radiologists;
- Ensure privacy of patients;
- Be flexible to new protocols;
- Adhere to the current network standards;
- Be independent to the operating system;
- Have an user-friendly GUI;
- Must be quite fast (to be used in a clinical environment) and let the end user to continue his work (even on different patients) during server side operation.

Then the architecture is a client-server system developed in Java (called jSecureSync), that allows secure synchronization between multiple clients and a server meeting the requirements imposed.
Usually, to ensure part of the requirements, we can use Access Control techniques such as:

- Confidentiality
- Authentication
- Authorization

3.2.1 Confidentiality

In information security, confidentiality refers to preventing the disclosure of information to unauthorized individuals or systems. That protection must be realized regardless of the security of the communication system used: indeed it assumes particular interest if the problem of ensuring the confidentiality of communication is used when the system is inherently insecure (such as the Internet). In particular we chose Transport Layer Security (TLS) and its predecessor, Secure Sockets Layer (SSL), that are cryptographic protocols designed to provide communication security over the Internet [87]. This protocol use X.509 certificates (an ITU-T standard [88] for a public key infrastructure - PKI [89]) and hence asymmetric cryptography to assure the counterparty whom they are talking with, and to exchange a symmetric key. Then this symmetric session key is used to encrypt data flowing between the parties.
To ensure the safety requirements we had to implement the version Client-authenticated of TLS Handshake Protocol (basic and most common implementation provides only authentication of the server side - eg. HTTPS)

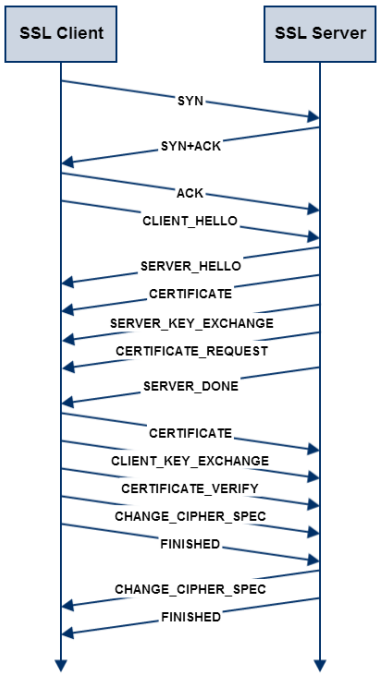


Figure 3.17: A complete TLS\SSL channel handshake

Certificates, that are generated at users registration time, adhere to the X.509 standard and they are issued with a 3072bit RSA Key Pair, according to the 2012 NIST recommendations [90].
TLS\SSL permits the secure channel setup, agreeing by both parties, based on the following parameters:

- a shared session key agreement based on the Elliptic curve Diffie-Hellman (ECDH) [91, 92] algorithm;
- a Cipher-Block Chaining (CBC) is adopeted;
- data is encrypted with a 128bit shared key AES algorithm (RFC 3268 [90, 93]).

These settings result, in agreement with the NIST nomenclature, in an internal protocol configuration string: "TLS ECDHE RSA WITH AES 128 CBC SHA" (RFC 4492 [94]).

3.2.2 Authentication and Authorization

Access to the functionality of the system is carried out without adding a new login level or particular authentication steps but using data collected by the SSL session to obtain the certificate and verifying their validity. This security requirement is achieved by using Java Authentication and Authorization Service, or JAAS, the Java implementation of the standard Pluggable Authentication Module (PAM) information security framework [95].

JAAS provides and manages three main resources:

- a representation of identity (Principal) and a set of credentials (Subject);
- a login service that will invoke the application callbacks to ask the user things like username and password. It returns a new Subject;
- a service that tests if a Subject was granted a permission by an administrator.

To obtain a login module that responds to the authentication of users it is required to implement the *javax.security.auth.spi.LoginModule* interface.

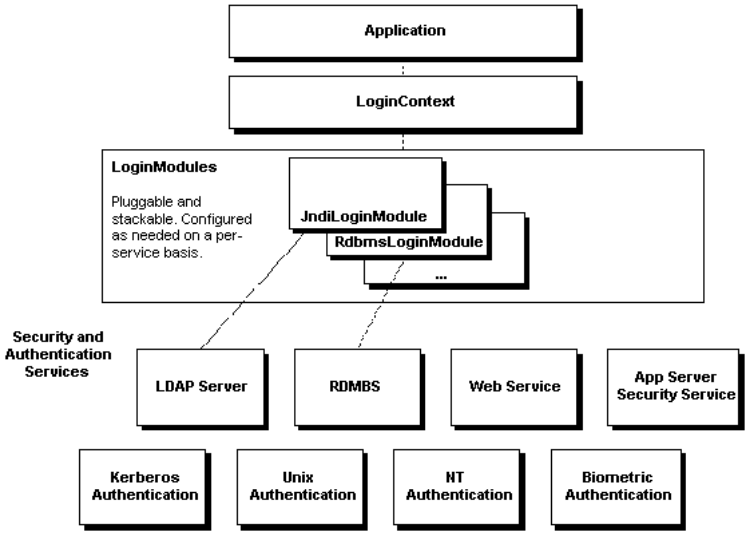


Figure 3.18: JAAS: the Java implementation of the standard Pluggable Authentication Module (PAM)

The LoginContext (*javax.security.auth.login.LoginContext*) is the core of the JAAS framework which kicks off the authentication process by creating a Subject. As the authentication process proceeds, the subject is populated with various principals and credentials for further processing; in particular it will need to apply and test the authorization policies, that will be created to handle each user requests.

3.2.3 OsiriX Plug-in

In OsiriX we developed a plugin that displays and manages the requests made to the server (fig. 3.19). The plugin was been written in Objective-C using the Cocoa framework.

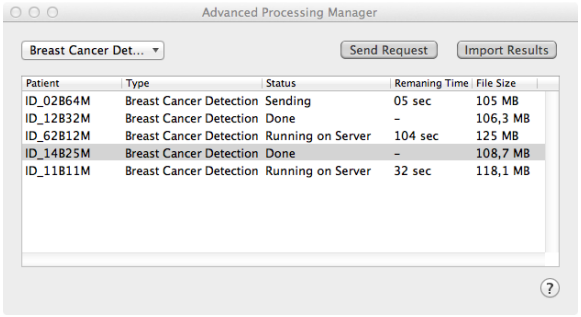


Figure 3.19: OsriX plugin GUI

For each job required the plugin shows the punctual status, the estimate remaining time and the size transferred on the network channel; when a job is completed the user can request the results and gather the segmented ROIs. Some GUI functionality is left for further research purpose.

3.2.4 Protocol

Clients and server communicate over the TLS\SSL channel through a very simple (and with low overhead) protocol (fig. 3.20, whose commands are:

- $[C \rightarrow S]$ HELO: client has to open a synchronization process with the server
- $[S \rightarrow C]$ OK v1.0: the server acknowledges and, at the same time, it informs the client about the protocol version
- $[C \rightarrow S]$ C_SND_RQ: the client requests to synchronize a file
- $[S \rightarrow C]$ S_SND_ACK: the server allows the client to send over the channel
- $[S \rightarrow C]$ S_SND_RQ: the server requests to synchronize a file
- $[C \rightarrow S]$ C_SND_ACK: the client allows the server to send over the channel

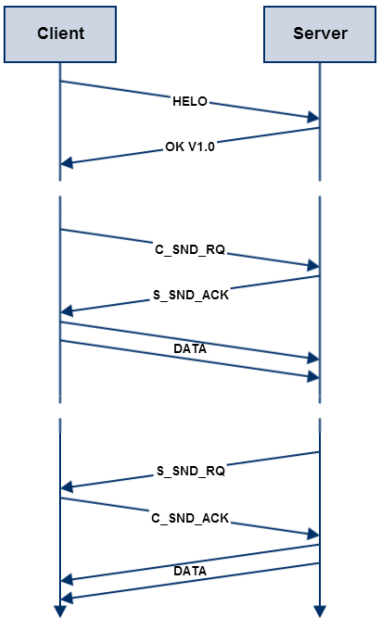


Figure 3.20: Internal protocol time diagram

3.2.5 Compression

In order to reduce the transmission overhead, JSecureSync always performs a lossless compression before starting a transmission. We use Zip4j [96] a java open-source library to handle Zip files that offers the following features:

- supports Zip format with 10 compression level [0-9];
- create, Add, Extract, Update, Remove files from a Zip file;
- read/Write password protected Zip files;
- supports AES 128/256 Encryption;
- supports Standard Zip Encryption;
- supports Zip64 format;
- supports Store (No Compression) and Deflate compression method;
- create or extract files from Split Zip files (Ex: z01, z02,...zip);

- supports Unicode file names;
- progress Monitor.

DICOM files take up about 110MB and maintain the information without providing a built-compression, but relying on external standards such as JPEG2000 and others. We decide to implement compression looking for a trade-off between the best compression factor and the additional time used in the compression. In table 3.5 we highlight the standard compression levels, the compression ratio and finally we compare them to the transmission time of the packet (over a LAN 10/100 Mbps network); all data are obtained on an average of 10 patients.
Due to the relevant correlation of medical images data, the best results are obtained by using the fastest compression (level 1). Beyond this level a better compression is not achieved but the time taken to reach it increases significantly.

Compr. Level	**Original Size**	**Compr. Size**	**Compr. Ratio**	**Compr. Time**	**Trasm. Time**
0 (Store)	111 MB	111 MB	0.00%	0.75 s	12.38 s
1 (Fastest)	111 MB	34 MB	69.36%	**3.47 s**	**3.82 s**
3 (Fast)	111 MB	34 MB	69.36%	3.55 s	3,76 s
5 (Normal)	111 MB	32 MB	70.85%	16.88 s	3.68 s
7 (Maximun)	111 MB	31 MB	71.66%	45.54 s	3.61
9 (Ultra)	111 MB	31 MB	72.12%	116.96 s	3.50 s

Table 3.5: Analysis of the compression phase. Times are averaged over ten patients.

4 Results

In this chapter the performances of the architecture are evaluated, both in a global view and analyzing in detail the experimental results of the neoplastic lesions segmentation subsystem.

4.1 Materials

All patients underwent imaging with a 1.5T scanner (Magnetom Symphony, Siemens Medical System, Erlangen, Germany) equipped with breast coils. DCE T1-weighted FLASH 3D coronal images were acquired (TR/TE: 9,8/4,76 ms; flip angle: 25 degrees; field of view 370x185 mm^2; matrix: 256x128 px^2; thickness: 2 mm; PixelSpaceing:1,4453 mm; gap: 0; acquisition time: 56s; 80 slices spanning entire breast volume). One series (t0) was acquired before and 9 series (t1 t9) after intravenous injection of 0.1 mmol/kg of a positive paramagnetic contrast agent (Gd-DOTA, Dotarem, Guerbet, Roissy CdG Cedex, France). An automatic injection system was used (Spectris Solaris EP MR, MEDRAD, Inc.,Indianola, PA) and injection flow rate was set to 2 ml/s followed by a flush of 10 ml saline solution at the same rate.
The dataset is constituted of 35 women breast DCE-MRI 4D data, (average age 40 years, in range 16-69) with benign or malignant lesions, all histopathologically proven: 16 lesions were benign and 19 were malignant

(tab. 4.1).

ID	Age	Side	Diagnosed (histopathologically proven)
b2	36	Right	Benign neoplastic lesion
b3	40	Left	Benign neoplastic lesion
b4	36	Right	Benign neoplastic lesion
b5	32	Left	Benign neoplastic lesion
b6	69	Right	Benign fibroadenoma
b7	47	Left	Benign fibrocystic dysplasia
b8	22	Left	Benign fibroadenoma
b9	34	Right	Adenosis associated to benign hyperplasia fibroadenomatosa
b10	32	Right	Benign fibrocystic dysplasia
b11	39	Left	Benign fibroadenoma
b12	42	Right	Benign fibrocystic dysplasia
b13	16	Right	Benign fibroadenoma
b14	26	Both	Benign neoplastic lesion (two)
b15	38	Right	Benign fibroadenoma
b16	38	Left	Benign fibroadenoma (two)
b17	26	Right	Benign neoplastic lesion (two)
m1	35	Left	Malignant neoplastic lesion (two)
m2	45	Left	Malignant neoplastic lesion
m3	46	Left	Malignant neoplastic lesion
m5	40	Right	Malignant neoplastic lesion
m7	45	Left	Infiltrating ductal carcinoma of intermediate grade
m8	40	Right	Infiltrating ductal carcinoma
m9	42	Both	Infiltrating ductal carcinoma
m10	45	Left	Malignant neoplastic lesion
m11	52	Left	Infiltrating lobular carcinoma
m12	53	Left	Ductal carcinoma in situ
m13	41	Right	Infiltrating ductal carcinoma
m14	38	Left	Infiltrating ductal carcinoma
m15	41	Right	Infiltrating ductal carcinoma
m16	45	Right	Infiltrating ductal carcinoma
m17	53	Left	Dysplasia associated with ductal carcinoma in situ
m18	44	Right	Infiltrating ductal carcinoma
m19	29	Left	Infiltrating ducttal-lobular carcinoma
m20	44	Right	Ductal carcinoma in situ (two)
m21	27	Right	Malignant neoplastic lesion

Table 4.1: Patients DB.

4.2 Performance evaluation: Segmentation

In this section we show the performance of the only segmentation subsystem, using statistical indicators already defined in the table 2.3. The table 4.2 reports the results of the evaluation study by varying the pre-processing method, with and without pre-selection phase, and by varying the classifier (tab. 3.1). For each combination the table reports the mean value (evaluated on a leave-one-patient-out basis) of sensitivity, specificity and accuracy, in decreasing order of accuracy.

PreProc.	**PreSel.**	**Classifier**	**Accuracy**	**Sensitivity**	**Specificity**
Median	Yes	SVM	98.70%	71.56%	98.94%
Median	Yes	RF	98.65%	67.78%	98.95%
Median	Yes	MLP	98.63%	72.82%	98.85%
Rigid Reg.	Yes	SVM	98.57%	67.13%	98.85%
Rigid Reg.	Yes	MLP	98.48%	68.79%	98.74%
Rigid Reg.	Yes	RF	98.40%	64.19%	98.70%
None	Yes	SVM	98.39%	69.82%	98.66%
None	Yes	RF	98.31%	65.92%	98.62%
None	Yes	MLP	97.87%	72.61%	98.13%
Median	No	RF	95.07%	90.81%	95.10%
Median	No	SVM	94.75%	92.74%	94.75%
Rigid Reg.	No	RF	94.28%	90.19%	94.30%
None	No	RF	94.19%	90.74%	94.21%
Rigid Reg.	No	SVM	94.00%	91.20%	94.00%
Rigid Reg.	No	MLP	93.97%	92.18%	93.98%
None	No	MLP	93.75%	90.65%	93.76%
Median	No	MLP	93.73%	92.82%	93.71%
None	No	SVM	93.61%	91.08%	93.61%

Table 4.2: ROI detection performance.

These results suggest that our approach with a SVM classification can give simultaneously a great accuracy and specificity, and a good value of sensitivity. It is worth noting, by considering the accuracy, that the results obtained by using the median registration outperform those obtainable with

a rigid registration, which, in turn, are better with respect to those achieved without registration. The obtained results also confirmed that the use of a Relative-Enhancement-based voxel pre-selection gave always rise to better results with respect to the case when no pre-selection is used. When pre-selection was used, the SVM classifier performed consistently better than MLP and RF (differences in accuracy are always statistically significant — $p < 0.05$).

The intermediate results and the final ones are shown in the following figures; the process has been applied to a malignant and benign tumor.

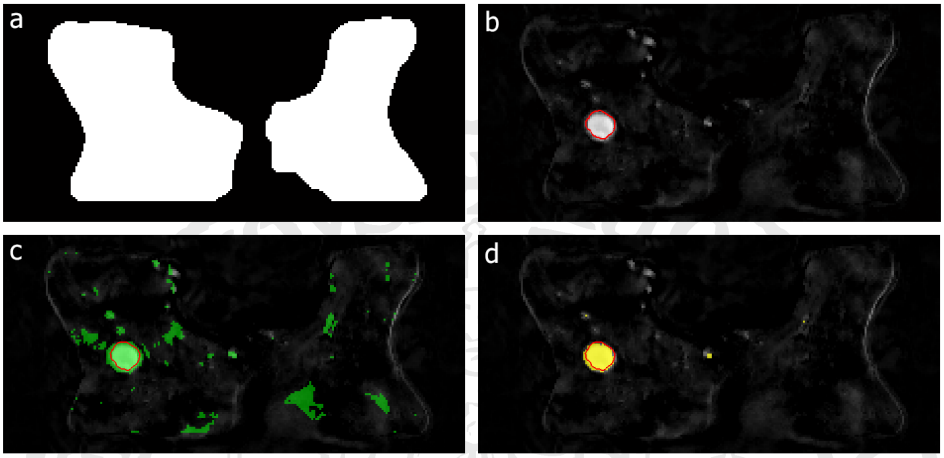


Figure 4.1: A benign lesion: a) Brest-mask; b) Manual ROI lesion segmentation (red perimeter); c) Pre-selection mask (green area); d) automatic detected ROI (yellow area)

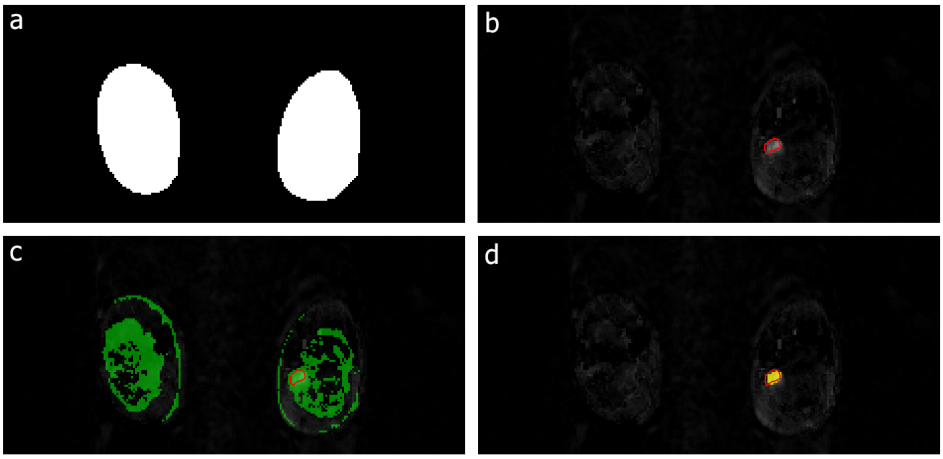


Figure 4.2: A malignant lesion: a) Brest-mask; b) Manual ROI lesion segmentation (red perimeter); c) Pre-selection mask (green area); d) automatic detected ROI (yellow area)

In order to further validate the results obtained, we have implemented some works currently available in the literature:

Fusco et al. (MLP based) [97]: Implements, among other techniques facing in the paper, an approach based on dynamic features and MLP classifier;

Fusco et al. (Pixel-Based) [85]: Implements a pixel-based technique that applies a binary Thresholding on some dynamic features such as the maximum relative enhancement (ReMax) and the instant of time when the peak occurs (Time To Peak - TTP);

Torricelli et al. (Pixel-Based) [86]: It also implements a pixel-based technique as the previous one but with different choices of feature and Thresholding. It is developed for segmentation of Colon/Rectum Cancer.

Table 4.3 reports our best result (from tab. 4.2) compared with other methodologies. It is evident that our method demonstrates the best accuracy, with a sensitivity which is significantly higher than the second best.

As regards the other two approaches, the difference among their accuracies and the one obtained by our approach is statistically significant ($p < 0.05$).

Methodology	**Accuracy**	**Sensitivity**	**Specificity**
Proposed Methodology	98.70%	71.56%	98.94%
Torricelli et al. [86]	98.69%	25.80%	99.49%
Fusco et al. [85]	86.99%	90.97%	86.99%
Pixel-Based on RE [97]	86.59%	75.44%	86.64%

Table 4.3: Performance comparison of the proposed method with other approaches.

In order to better show our results, figures 4.3 and 4.4 compare the automatic segmented region obtained for a benign and a malignant lesion, respectively, by using the approaches reported in table 4.3. In this case the advantage of using our approach appears even more evident.

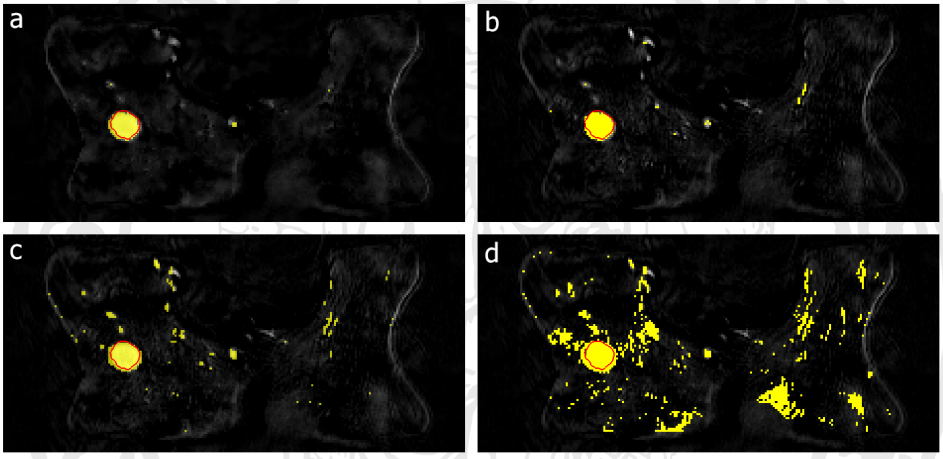


Figure 4.3: Comparing results for a benign lesion: a) our proposed approach; b) the pixel-based approach proposed by Torricelli et al; c) the MLP-based approach proposed by Fusco et al; d) the Pixel-Based on RE proposed by Fusco et al.

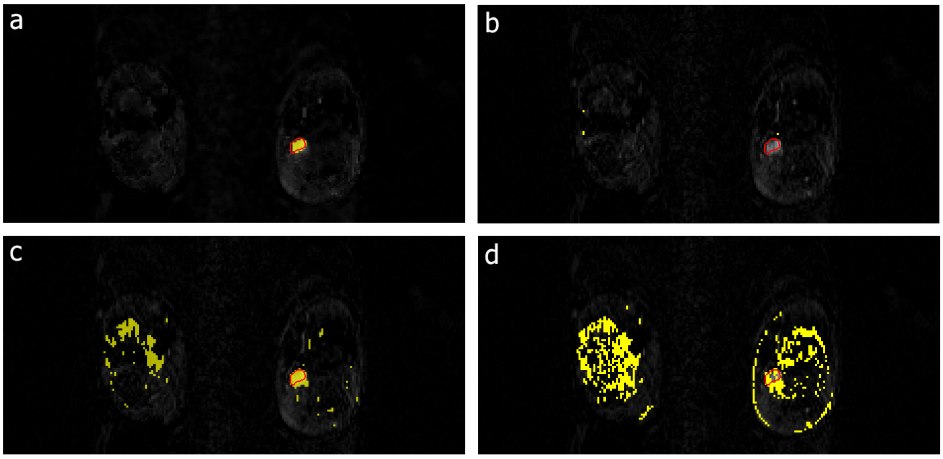


Figure 4.4: Comparing results for a malignant lesion: a) our proposed approach; b) the pixel-based approach proposed by Torricelli et al; c) the MLP-based approach proposed by Fusco et al.; d) the Pixel-Based on RE proposed by Fusco et al. Note that in this case b) was not able to detect the lesion.

4.3 Performance evaluation: Architecture

To prove the validity of the client-server model we compare the execution time required to perform the entire operation into two distinct conditions: the first condition (tab. 4.4 on the right) requires that the segmentation is performed totally on a typical OsiriX workstation (Apple iMac with Intel Core 2 Duo 2.0 GHz with 3GB RAM) and the second one (tab. 4.4 on the left) shows the performance improvement which is obtained in the developed client-server architecture using the speed-up achieved on a server dedicated to the segmentation of neoplastic lesions (2x Quad Core Xeon 3.0Ghz 32GB RAM).

	Remote					Local
ID	$C \rightarrow S$	**Segment. time (s)**	$S \rightarrow C$	**Total time (s)**	**Trasm. time (s)**	**Segment. time (s)**
b2	29.73	134.94	1.24	165.91	18.67%	1142.13
b3	28.34	158.48	2.04	188.85	16.09%	1322.21
b4	30.14	149.36	1.80	181.30	17.62%	1211.98
b5	30.27	115.48	1.55	147.30	21.60%	998.30
p5	31.32	148.99	1.65	181.96	18.11%	1223.94
m1	28.84	147.32	1.35	177.50	17.00%	1523.54
m2	29.71	152.82	1.81	184.35	17.10%	1342.35
m3	30.12	121.63	1.23	152.98	20.49%	1083.15
m5	30.42	143.27	1.72	175.42	18.32%	1263.65
m7	28.73	168.47	1.93	199.14	15.39%	1526.23
Avg.	29.76	144.08	1.63	175.47	17.89%	1263.74

Table 4.4: Performance comparation of remote computing time (including Client to Server transmission, image segmentation on the Server and Server to Client transmission) versus local computing time over ten patients.

It is also important to note that the overhead added by the distributed architecture is just 18% of the total processing time.

5

Conclusion and future works

The results shown demonstrate the validity and quality of an automatic detection tool as the one developed in this thesis. As many times pointed out the ultimate goal is not to replace the radiologist in the difficult task of identifying or diagnosing neoplastic diseases, but to simplify the difficult flow that leads to a good diagnosis by providing tools as reliable as possible. As first step, taken at the end of this long work, we presented the results, just discussed in this thesis, to the scientific community by publishing it [98, 99].

The road to a complete, reliable and medically valid tool through some improvements that cannot be ignored; The whole segmentation system offers several margins improvements including:

- a better implementation of the brestmask extraction module in order to improve the segmentation of the breast tissue;
- the integration of other feature than dynamics one. Of special interest are the pharmacokinetic feature that deduce some physiological parameters of the examined tissue basing on the kinetics of the contrast medium measured in DCE-MRI. These features represent a new direction in segmentation and classification of neoplastic tissues in general;

- a better Implementation of elastic image registration.

Other interesting ideas for future developments mainly concern to consider different acquisition protocols, ensuring that the segmentation techniques support at least the most common protocols. At last, before to release the service to end users, especially in a so sensitive sector as the medical environment, it is necessary to deep in a critical manner the problem of the dependability; this aspect is essential to ensure meaning of "trust" about the system and especially about the results. This list of possible improvements does not represent a reminder but it is a detailed list of activities which today are the object of study of the working group in which has been developed this thesis.

Ringraziamenti

≪Dobbiamo dunque considerare lo stato presente dell'universo come effetto del suo stato anteriore e come causa del suo stato futuro. Un'intelligenza che, per un dato istante, conoscesse tutte le forze da cui la natura è animata e la situazione rispettiva degli esseri che la compongono, se fosse abbastanza vasta da sottoporre questi dati ad analisi abbraccerebbe nella stessa formula i moti dei corpi più grandi dell'universo e quelli dell'atomo più leggero: per essa non ci sarebbe nulla d'incerto, ed il futuro come il passato sarebbe presente ai suoi occhi.≫
Pierre-Simon Laplace, 1812

Sono tante le persone che mi hanno accompagnato fino a questo stato "presente" ma prima di cominciare ad annoiarvi col lungo elenco di ringraziamenti (neanche tanto lungo) vorrei ringraziare me stesso, non per narcisismo ne tanto meno per egoismo ma perché credo che, per quanto ci possano indicare, accompagnare o addirittura spingere in un percorso, a noi spetta l'ultima scelta, consapevole od inconsapevole che sia, per il mondo saremo gli unici artefici del nostro "destino"!

Le persone che ho avuto o che ho scelto accanto a me nella mia vita sono poche ma ognuna di queste ha contribuito nel bene e nel male a fare di me quel che sono oggi. Voglio ringraziarvi per ciò che mi avete regalato:

Ringrazio mia Madre, per avermi regalato la convinzione e la determinazione che mi contraddistingue, convinzione che forse devo ancora imparare a dosare ma che mi ha sempre permesso di affrontare a testa alta ogni problema.

Ringrazio mio Padre, per avermi regalato un modello a cui potermi ispirare e, se oggi c'è un ingegnere informatico in più è "colpa" sua.

Ringrazio mia Sorella per avermi regalato un supporto ed un "ti sopporto" nell'intero lavoro di tesi (e non solo).

Ringrazio mio Fratello per avermi regalato ogni mattina una "dolce" undicesima sveglia umana dove ben due cellulari avevano fallito.

Ringrazio la mia Famiglia per avermi regalato una strada da percorrere. Approfitto per dirvi quello che difficilmente riuscirei a dirvi altrimenti: vi voglio bene.

Ringrazio Linda, non per avermi regalato l'amore, ad amare siamo bravi tutti, ma per avermi regalato la voglia di amare. Quella che si perde facilmente e che con difficoltà ci si accorge di aver perso.

Ringrazio Massimo ed Enzo per avermi regalato il "migliore amico" quello che tutti i bimbi hanno ma che forse nessun adulto sa di avere più. Tra le tante ringrazio Massimo per avermi regalato una birra quando avevo sete.

Ringrazio Nicola e Roberta per avermi regalato un'amicizia Vera in un modo che di Vero ha ben poco.

Ringrazio Luca per avermi regalato un amico dalle mille sfaccettature ed avermi insegnato ad affrontare i problemi più seri (ed i viaggi più lunghi...) con la giusta simpatia

Ringrazio Stefano per avermi regalato un degno "compagno di avventura" che spero possa proseguire con ulteriori e sempre migliori successi.

Ringrazio i prof.s Carlo Sansone e Mario Sansone per avermi regalato più di una possibilità ed avermi trasmesso la passione che spesso non arriva allo studente che siede l'altro lato di una cattedra.

Ringrazio i miei Amici (TUTTI) per avermi regalato un sorriso e soprattutto per avermene strappati altrettanti.

Nella speranza che non prendiate queste parole come un gesto dovuto ma piuttosto come un gesto sentito, vi ringazio di tutto...
Gabriele

Bibliography

[1] Ruper A. Willis. *The spread of tumors in the human body.* London: Butterworths, 197, 3 edition, November 1973.

[2] Ahmedin Jemal, Freddie Bray, Melissa M Center, Jacques Ferlay, Elizabeth Ward, and David Forman. Global cancer statistics. *CA: a cancer journal for clinicians*, 61(2):69–90, 2011.

[3] Steven I Hajdu. A note from history: landmarks in history of cancer, part 1. *Cancer*, 117(5):1097–1102, 2011.

[4] Flavia Foca, Silvia Mancini, Lauro Bucchi, Donella Puliti, Marco Zappa, Carlo Naldoni, Fabio Falcini, Maria L Gambino, Silvano Piffer, Maria E Sanoja Gonzalez, et al. Decreasing incidence of late-stage breast cancer after the introduction of organized mammography screening in italy. *Cancer*, 119(11):2022–2030, 2013.

[5] Jacques Ferlay, Hai-Rim Shin, Freddie Bray, David Forman, Colin Mathers, and Donald Maxwell Parkin. Estimates of worldwide burden of cancer in 2008: Globocan 2008. *International journal of cancer*, 127(12):2893–2917, 2010.

[6] Jacques Ferlay, Hai-Rim Shin, Freddie Bray, David Forman, Colin

Mathers, and Donald Maxwell Parkin. Globocan 2008 v2.0, cancer incidence and mortality worldwide: Iarc cancerbase no. 10. lyon, france: International agency for research on cancer, 2013. http://globocan.iarc.fr/.

[7] Peter C Gøtzsche and Margrethe Nielsen. Screening for breast cancer with mammography. *Cochrane Database Syst Rev*, 4(1), 2009.

[8] Ministero della salute. Screening oncologici - raccomandazioni per la pianificazione e lesecuzione degli screening di popolazione per la prevenzione del cancro della mammella, del cancro della cervice uterina e del cancro del colon retto, November 2006. http://www.ccm-network.it/documenti_Ccm/pubblicazioni/Screening_oncologici.pdf.

[9] Heidi D Nelson, Kari Tyne, Arpana Naik, Christina Bougatsos, Benjamin K Chan, and Linda Humphrey. Screening for breast cancer: an update for the us preventive services task force. *Annals of internal medicine*, 151(10):727–737, 2009.

[10] Paul S Tofts. T1-weighted dce imaging concepts: Modelling, acquisition and analysis. *signal*, 500(450):400, 2010.

[11] Amy Berrington de González and Sarah Darby. Risk of cancer from diagnostic x-rays: estimates for the uk and 14 other countries. *The lancet*, 363(9406):345–351, 2004.

[12] Antoine Rosset and Luca Spadola. Osirix: an open-source software for navigating in multidimensional dicom images. *Journal of Digital Imaging*, 17(3):205–216, 2004.

[13] American Cancer Society. The history of cancer, 2012. http://www.cancer.org/acs/groups/cid/documents/webcontent/002048-pdf.pdf.

[14] Jeremy Laurance et al. Breast cancer cases rise 80% since seventies. *The Independent*, 29, 2006.

[15] Preetha Anand, Ajaikumar B Kunnumakara, Chitra Sundaram, Kuzhuvelil B Harikumar, Sheeja T Tharakan, Oiki S Lai, Bokyung Sung, and Bharat B Aggarwal. Cancer is a preventable disease that requires major lifestyle changes. *Pharmaceutical research*, 25(9):2097–2116, 2008.

[16] Melissa Bondy. Cancer mortality and morbidity patterns in the us population: An interdisciplinary approach. *JAMA: The Journal of the American Medical Association*, 306(6):657, 2011.

[17] Larissa A Korde, Jo Anne Zujewski, Leah Kamin, Sharon Giordano, Susan Domchek, William F Anderson, John MS Bartlett, Karen Gelmon, Zeina Nahleh, Jonas Bergh, et al. Multidisciplinary meeting on male breast cancer: summary and research recommendations. *Journal of clinical oncology*, 28(12):2114–2122, 2010.

[18] Marc T. Goodman, Ko-Hui Tung, and Lynne R Wilkens. Comparative epidemiology of breast cancer among men and women in the us, 1996 to 2000. *Cancer Causes & Control*, 17(2):127–136, 2006.

[19] Carin I Perkins, Joellyn Hotes, Betsy A Kohler, and Holly L Howe. Association between breast cancer laterality and tumor location, united states, 1994–1998. *Cancer Causes & Control*, 15(7):637–645, 2004.

[20] Anders Ekbom, Hans-Olov Adami, Dimitrios Trichopoulos, Mats Lambe, Chung-cheng Hsieh, and Jan Pontén. Epidemiologic correlates of breast cancer laterality (sweden). *Cancer Causes & Control*, 5(6):510–516, 1994.

[21] Leslie H Sobin and Irvin D Fleming. Tnm classification of malignant tumors. *Cancer*, 80(9):1803–1804, 1997.

[22] Ansgar Malich, Thomas Böhm, Mirjam Facius, Martin G. Freesmeyer, Marlies Fleck, Roselle Anderson, and Werner A. Kaiser. Differentiation of mammographically suspicious lesions: evaluation of breast ultrasound, mri mammography and electrical impedance scanning as adjunctive technologies in breast cancer detection. *Clinical radiology*, 56(4):278–283, 2001.

[23] Gabor T. Herman. *Fundamentals of computerized tomography: image reconstruction from projections*. Springer, 2009.

[24] Simon R Cherry and Magnus Dahlbom. Pet: physics, instrumentation, and scanners. In *PET*, pages 1–117. Springer, 2006.

[25] Helen Young, Richard Baum, Ursula Cremerius, Karl Herholz, Otto Hoekstra, Adriaan A. Lammertsma, Jan Pruim, and Pat Price. Measurement of clinical and subclinical tumour response using [^{18}f]-fluorodeoxyglucose and positron emission tomography: review and 1999 eortc recommendations. *European Journal of Cancer*, 35(13):1773–1782, 1999.

[26] Raymond Damadian. Tumor detection by nuclear magnetic resonance. *Science*, 171(3976):1151–1153, 1971.

[27] Wikipedia. Nuclear magnetic resonance, 2013. http://en.wikipedia.org/wiki/Nuclear_magnetic_resonance.

[28] Wikipedia. Magnetic resonance imaging, 2013. http://en.wikipedia.org/wiki/Magnetic_resonance_imaging.

[29] D Shahbazi-Gahrouei, M Williams, and BJ Allen. In vitro study of relationship between signal intensity and gadolinium-dtpa concentration at high magnetic field strength. *Australasian radiology*, 45(3):298–304, 2001.

[30] Christiane Katharina Kuhl, Peter Mielcareck, Sven Klaschik, Claudia Leutner, Eva Wardelmann, Jürgen Gieseke, and Hans H Schild. Dynamic breast mr imaging: Are signal intensity time course data useful for differential diagnosis of enhancing lesions? *Radiology*, 211(1):101–110, 1999.

[31] Fiona J Gilbert, Susan M Astley, Magnus A McGee, Maureen GC Gillan, Caroline RM Boggis, Pamela M Griffiths, and Stephen W Duffy. Single reading with computer-aided detection and double reading of screening mammograms in the united kingdom national breast screening program1. *Radiology*, 241(1):47–53, 2006.

[32] Matthew Gromet. Comparison of computer-aided detection to double reading of screening mammograms: review of 231,221 mammograms. *American Journal of Roentgenology*, 190(4):854–859, 2008.

[33] Rachel F Brem, Janet Baum, Mary Lechner, Stuart Kaplan, Stuart Souders, L Gill Naul, and Jeff Hoffmeister. Improvement in sensitivity of screening mammography with computer-aided detection: a multiinstitutional trial. *American Journal of Roentgenology*, 181(3):687–693, 2003.

[34] Joshua J Fenton, Stephen H Taplin, Patricia A Carney, Linn Abraham, Edward A Sickles, Carl D'Orsi, Eric A Berns, Gary Cutter, R Edward Hendrick, William E Barlow, et al. Influence of computer-aided detection on performance of screening mammography. *New England Journal of Medicine*, 356(14):1399–1409, 2007.

[35] Calvin F Nodine, Harold L Kundel, et al. Using eye movements to study visual search and to improve tumor detection. *Radiographics*, 7(6):1241–1250, 1987.

[36] HD Cheng, XJ Shi, Rui Min, LM Hu, XP Cai, and HN Du. Approaches

for automated detection and classification of masses in mammograms. *Pattern recognition*, 39(4):646–668, 2006.

[37] U.S. Preventive Services Task Force. Uspstf recommendation: Screening for breast cancer, 2009. http://www.uspreventiveservicestaskforce.org/uspstf/uspsbrca.htm.

[38] National Electrical Manufacturers Association NEMA. Digital imaging and communications in medicine (dicom) part 1: Introduction and overview, 2004. http://medical.nema.org/dicom/2004/04_01PU.PDF.

[39] Peter Mildenberger, Marco Eichelberg, and Eric Martin. Introduction to the dicom standard. *European radiology*, 12(4):920–927, 2002.

[40] Dean W. Bidgood, Steven C. Horii, Fred W. Prior, and Donald E. Van Syckle. Understanding and using dicom, the data interchange standard for biomedical imaging. *Journal of the American Medical Informatics Association*, 4(3):199–212, 1997.

[41] Sruba Chowdhury and Eddy B Boskamp. Breast coil for magnetic resonance imaging, November 15 1994. US Patent 5,363,845.

[42] Werner A. Kaiser. Dynamic magnetic resonance breast imaging using a double breast coil: an important step towards routine examination of the breast. In *Frontiers in European radiology*, pages 39–68. Springer, 1990.

[43] Christine Tanner, David J. Hawkes, Michael Khazen, Preminda Kessar, and Martin O. Leach. Does registration improve the performance of a computer aided diagnosis system for dynamic contrast-enhanced mr mammography? In *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*, pages 466–469. IEEE, 2006.

[44] Lisa Gottesfeld Brown. A survey of image registration techniques. *ACM computing surveys (CSUR)*, 24(4):325–376, 1992.

[45] Antoine J. B. Maintz and Max A. Viergever. A survey of medical image registration. *Medical image analysis*, 2(1):1–36, 1998.

[46] Daniel Rueckert, Luke I Sonoda, Carmel Hayes, Derek LG Hill, Martin O Leach, and David J Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *Medical Imaging, IEEE Transactions on*, 18(8):712–721, 1999.

[47] Ardeshir A. Goshtasby. *2-D and 3-D image registration: for medical, remote sensing, and industrial applications*. Wiley Online Library, 2005.

[48] Graeme P. Penney, Jürgen Weese, John A. Little, Paul Desmedt, Derek L. G. Hill, et al. A comparison of similarity measures for use in 2d-3d medical image registration. *Medical Imaging, IEEE Transactions on*, 17(4):586–595, 1998.

[49] Lina Arbach, Alan Stolpen, and Joseph M Reinhardt. Classification of breast mri lesions using a backpropagation neural network (bnn). In *Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on*, pages 253–256. IEEE, 2004.

[50] Lina A. Meinel, Thomas Buelow, Dezheng Huo, Akiko Shimauchi, Ursula Kose, Johannes Buurman, and Gillian Newstead. Robust segmentation of mass-lesions in contrast-enhanced dynamic breast mr images. *Journal of Magnetic Resonance Imaging*, 32(1):110–119, 2010.

[51] Xi Liang, Kotagiri Ramamohanara, Helen Frazer, and Qing Yang. Lesion segmentation in dynamic contrast enhanced mri of breast. In *Digital Image Computing Techniques and Applications (DICTA), 2012 International Conference on*, pages 1–8. IEEE, 2012.

[52] Weijie Chen, Maryellen L. Giger, Ulrich Bick, and Gillian M. Newstead. Automatic identification and classification of characteristic kinetic curves of breast lesions on dce-mri. *Medical Physics*, 33:2878, 2006.

[53] Weijie Chen, Maryellen L. Giger, and Ulrich Bick. A fuzzy c-means (fcm)-based approach for computerized segmentation of breast lesions in dynamic contrast-enhanced mr images. *Academic radiology*, 13(1):63–72, 2006.

[54] Geir Torheim, Fred Godtliebsen, David Axelson, Kjell Arne Kvistad, A. Haraldseth, and Peter A. Rinck. Feature extraction and classification of dynamic contrast-enhanced t2*-weighted breast image data. *Medical Imaging, IEEE Transactions on*, 20(12):1293–1301, 2001.

[55] Yaniv Gal, Andrew Mehnert, Andrew Bradley, Dominic Kennedy, and Stuart Crozier. Feature and classifier selection for automatic classification of lesions in dynamic contrast-enhanced mri of the breast. In *Digital Image Computing: Techniques and Applications, 2009. DICTA'09.*, pages 132–139. IEEE, 2009.

[56] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3:1157–1182, 2003.

[57] María C. Alonso, José A. Malpica, and A. M. Agirre. Consequences of the hughes phenomenon on some classification techniques. In *ASPRS 2011 Annual Conference, Milwaukee, Wisconsin May*, pages 1–5, 2011.

[58] Stuart Jonathan Russell, Peter Norvig, John F. Canny, Jitendra M. Malik, and Douglas D. Edwards. *Artificial intelligence: a modern approach*, volume 74. Prentice hall Englewood Cliffs, 1995.

[59] Frank Rosenblatt. Principles of neurodynamics. perceptrons and the theory of brain mechanisms. Technical report, DTIC Document, 1961.

[60] David E. Rumelhart, Geoffrey E. Hintont, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

[61] David E. Rumelhart and James L. McClelland. Parallel distributed processing: explorations in the microstructure of cognition. volume 1. foundations. *Nature*, 1986.

[62] Lior Rokach. *Data mining with decision trees: theory and applications*, volume 69. World Scientific, 2007.

[63] Ross J. Quinlan. Induction of decision trees. *Machine learning*, 1(1):81–106, 1986.

[64] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[65] Yoav Freund and Robert E. Schapire. A desicion-theoretic generalization of on-line learning and an application to boosting. In *Computational learning theory*, pages 23–37. Springer, 1995.

[66] Pierre A. Devijver and Josef Kittler. *Pattern recognition: A statistical approach*. Prentice/Hall International Englewood Cliffs, NJ, 1982.

[67] Wikipedia. Cross-validation (statistics), 2013. http://en.wikipedia.org/wiki/Cross-validation_(statistics).

[68] Murat Dundar, Glenn Fung, Luca Bogoni, Michael Macari, A. Megibow, and Bharat Rao. A methodology for training and validating a cad system and potential pitfalls. In *International Congress Series*, volume 1268, pages 1010–1014. Elsevier, 2004.

[69] Alan L. Scheinine, Marco Donizelli, and Marco Pescosolido. An object-oriented client-server system for interactive segmentation of medical images using the method of active contours. In *Bildverarbeitung für die Medizin 1998*, pages 308–312. Springer, 1998.

[70] Achim Mayer and Hans-Peter Meinzer. High performance medical image processing in client/server-environments. *Computer methods and programs in biomedicine*, 58(3):207–217, 1999.

[71] Sherif M. Yacoub and Hany H. Ammar. The development of a client/server architecture for standardized medical application network services. In *Application-Specific Systems and Software Engineering and Technology, 1999. ASSET'99. Proceedings. 1999 IEEE Symposium on*, pages 2–9. IEEE, 1999.

[72] Dan Svantesson and Roger Clarke. Privacy and consumer risks in cloud computing. *Computer Law & Security Review*, 26(4):391–397, 2010.

[73] Hein S. Venter and Jan Eloff. A taxonomy for information security technologies. *Computers & Security*, 22(4):299–307, 2003.

[74] Committee on National Security Systems CNSS. Cnss instruction: National information assurance (ia) glossary, 2010. https://www.cnss.gov/Assets/pdf/cnssi_4009.pdf.

[75] Algirdas Avizienis, J-C Laprie, Brian Randell, and Carl Landwehr. Basic concepts and taxonomy of dependable and secure computing. *Dependable and Secure Computing, IEEE Transactions on*, 1(1):11–33, 2004.

[76] Jean-Claude Laprie. Dependable computing and fault-tolerance. *Digest of Papers FTCS-15*, pages 2–11, 1985.

[77] Bev Littlewood. Dependability assessment of software-based systems: state of the art. In *Software Engineering, 2005. ICSE 2005. Proceedings. 27th International Conference on*, pages 6–7. IEEE, 2005.

[78] Shandong Wu, Susan P. Weinstein, Emily F. Conant, Mitchell D. Schnall, and Despina Kontos. Automated chest wall line detection for whole-breast segmentation in sagittal breast mr images. *Medical physics*, 40:042301, 2013.

[79] Maria Lorenzo-Valdés, Gerardo I Sanchez-Ortiz, Raad Mohiaddin, and Daniel Rueckert. Atlas-based segmentation and tracking of 3d cardiac mr images using non-rigid registration. In *Medical Image Computing and Computer-Assisted InterventionMICCAI 2002*, pages 642–650. Springer, 2002.

[80] Thorsten Twellmann, Axel Saalbach, C. Muller, T. W. Nattkemper, and A. Wismuller. Detection of suspicious lesions in dynamic contrast enhanced mri data. In *Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE*, volume 1, pages 454–457. IEEE, 2004.

[81] Thorsten Twellmann, Axel Saalbach, Olaf Gerstung, Martin O. Leach, Tim W. Nattkemper, et al. Image fusion for dynamic contrast enhanced magnetic resonance imaging. *Biomedical engineering online*, 3(1):35, 2004.

[82] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.

[83] Stefano Marrone. Tecniche di intelligenza artificiale per la classificazione di immagini biomedicali nella detezione di lesioni alla mammella. *Università degli Studi di Napoli "Federico II"*, 2012.

[84] MathWorks. Matlab image processing toolkit: Image registration techniques, 2012. http://www.mathworks.it/it/help/images/registering-an-image.html.

[85] Roberta Fusco, Mario Sansone, Carlo Sansone, and Antonella Petrillo. Selection of suspicious rois in breast dce-mri. In *Image Analysis and Processing–ICIAP 2011*, pages 48–57. Springer, 2011.

[86] Pietro Torricelli, Annarita Pecchi, Gabriele Luppi, and Renato Romagnoli. Gadolinium-enhanced mri with dynamic evaluation in diagnosing the local recurrence of rectal cancer. *Abdominal imaging*, 28(1):0019–0027, 2003.

[87] Tim Dierks. The transport layer security (tls) protocol version 1.2. *RFC 5246*, 2008. http://www.ietf.org/rfc/rfc5246.txt.

[88] ITU-T. X.509 : Information technology, 2008. http://www.itu.int/rec/T-REC-X.509/.

[89] David Solo, Russell Housley, and Warwick Ford. Internet x. 509 public key infrastructure certificate and crl profile. *RFC 2459*, 1999. http://www.ietf.org/rfc/rfc2459.txt.

[90] Elaine Barker, William Barker, William Burr, William Polk, and Miles Smid. Recommendation for key management–part 1: General (revision 3). *NIST special publication*, 800:57, 2011.

[91] Elaine Barker, Don Johnson, and Miles Smid. Recommendation for pair-wise key establishment schemes using discrete logarithm cryptography. *NIST Special Publication*, pages 800–56A, 2007.

[92] Certicom Research. Standards for efcient cryptography (sec): Elliptic curve cryptography, 2000. http://www.secg.org/collateral/sec1_final.pdf.

[93] Pete Chown. Advanced encryption standard (aes) ciphersuites for transport layer security (tls). *RFC 3268*, 2002. http://www.ietf.org/rfc/rfc3268.txt.

[94] Simon Blake-Wilson, Bodo Moeller, Vipul Gupta, Chris Hawk, and Nelson Bolyard. Elliptic curve cryptography (ecc) cipher suites for transport layer security (tls). *RFC 4492*, 2006. http://www.ietf.org/rfc/rfc4492.txt.

[95] Java Authentication. Authorization service (jaas). reference guide for the j2se development kit 5.0, 2001.

[96] Srikanth Reddy Lingala. Zip4j, 2012. http://www.lingala.net/zip4j.

[97] Roberta Fusco, Mario Sansone, Antonella Petrillo, and Carlo Sansone. A multiple classifier system for classification of breast lesions using dynamic and morphological features in dce-mri. In *Structural, Syntactic, and Statistical Pattern Recognition*, pages 684–692. Springer, 2012.

[98] Stefano Marrone, Gabriele Piantadosi, Roberta Fusco, Antonella Petrillo, Mario Sansone, and Carlo Sansone. Automatic lesion detection in breast dce-mri. In *Image Analysis and Processing–ICIAP 2013*, pages 359–368. Springer, 2013.

[99] Gabriele Piantadosi, Stefano Marrone, Mario Sansone, and Carlo Sansone. A secure osirix plug-in for detecting suspicious lesions in breast dce-mri. In *International Conference on Algorithms and Architectures for Parallel Processing–ICA3PP 2013*. Springer, 2013.

[100] National Cancer Institute. Cancer staging, 2013. http://www.cancer.gov/cancertopics/factsheet/detection/staging.

[101] Oleg S. Pianykh. *Digital imaging and communications in medicine (DICOM): a practical introduction and survival guide*. Springer, 2012.

[102] Essam A. El-Kwae, Joel E. Fishman, Maria J. Bianchi, Pradip M. Pattany, and Mansur R. Kabuka. Detection of suspected malignant patterns in three-dimensional magnetic resonance breast images. *Journal of Digital Imaging*, 11(2):83–93, 1998.

List of Figures

List of Tables