Physica Medica 64 (2019) 293–303

Contents lists available at ScienceDirect

Physica Medica

journal homepage: www.elsevier.com/locate/ejmp

Original paper

Development of breast lesions models database

Kristina Bliznakova[a,*], Nikolay Dukov[a], Firgan Feradov[a], Galja Gospodinova[a], Zhivko Bliznakov[a], Paolo Russo[b], Giovanni Mettivier[b], Hilde Bosmans[c], Lesley Cockmartin[c], Antonio Sarno[b], Desislava Kostova-Lefterova[d], Elitsa Encheva[e], Virginia Tsapaki[f], Daniel Bulyashki[g], Ivan Buliev[a]

[a] Laboratory of Computer Simulations in Medicine, Technical University of Varna, Varna, Bulgaria
[b] Dipartimento di Fisica "Ettore Pancini", Universita' di Napoli Federico II and INFN Sezione di Napoli, Napoli, Italy
[c] Department of Radiology, Katholieke University of Leuven, Leuven, Belgium
[d] Aleksandrovska Hospital, Sofia, Bulgaria
[e] Radiotherapy Department, University Hospital "St. Marina", Medical University of Varna, Varna, Bulgaria
[f] Medical Physics Department, Konstantopoulio General Hospital, Nea Ionia, Attiki, Greece
[g] Surgery Department, University Hospital "St. Marina", Medical University of Varna, Varna, Bulgaria

ARTICLE INFO

Keywords:
Database
Breast lesions
Segmentation
Computational models
Breast imaging techniques

ABSTRACT

Purpose: We present the development and the current state of the *MaXIMA Breast Lesions Models Database,* which is intended to provide researchers with both segmented and mathematical computer-based breast lesion models with realistic shape.

Methods: The database contains various 3D images of breast lesions of irregular shapes, collected from routine patient examinations or dedicated scientific experiments. It also contains images of simulated tumour models. In order to extract the 3D shapes of the breast cancers from patient images, an in-house segmentation algorithm was developed for the analysis of 50 tomosynthesis sets from patients diagnosed with malignant and benign lesions. In addition, computed tomography (CT) scans of three breast mastectomy cases were added, as well as five whole-body CT scans. The segmentation algorithm includes a series of image processing operations and region-growing techniques with minimal interaction from the user, with the purpose of finding and segmenting the areas of the lesion. Mathematically modelled computational breast lesions, also stored in the database, are based on the 3D random walk approach.

Results: The *MaXIMA* Imaging Database currently contains 50 breast cancer models obtained by segmentation of 3D patient breast tomosynthesis images; 8 models obtained by segmentation of whole body and breast cadavers CT images; and 80 models based on a mathematical algorithm. Each record in the database is supported with relevant information. Two applications of the database are highlighted: inserting the lesions into computationally generated breast phantoms and an approach in generating mammography images with variously shaped breast lesion models from the database for evaluation purposes. Both cases demonstrate the implementation of multiple scenarios and of an unlimited number of cases, which can be used for further software modelling and investigation of breast imaging techniques. The created database interface is web-based, user friendly and is intended to be made freely accessible through internet after the completion of the *MaXIMA* project.

Conclusions: The developed database will serve as an imaging data source for researchers, working on breast diagnostic imaging and on improving early breast cancer detection techniques, using existing or newly developed imaging modalities.

1. Introduction

Nowadays, the research in novel imaging techniques as well as in testing and optimizing existing ones is inevitably related to the exploitation of medical patient images. These are used to evaluate the improved technologies in the x-ray techniques, to train and educate the medical specialists with the new technology, to extract features, which are basis for the development of advanced computer-aided detection systems or develop and train machine learning algorithms. For researchers, the use of images from databases may be the most flexible and time efficient approach, since data are summarized at one place and well documented. In case of breast imaging the demands are

* Corresponding author.
E-mail address: kristina.bliznakova@gmail.com (K. Bliznakova).

https://doi.org/10.1016/j.ejmp.2019.07.017
Received 5 December 2018; Received in revised form 1 July 2019; Accepted 22 July 2019
Available online 03 August 2019

similar. Presently, research and industry communities have focused efforts to optimize the newly established x-ray breast imaging techniques: contrast-enhanced mammography [1,2], novel mammography techniques such as spot mammography [3] and dual-energy mammography [4–6], digital breast tomosynthesis [7–9], breast CT [10], phase-contrast tomography and mammography [11,12].

All these imaging modalities are designed, optimised and validated because of the need for earlier breast screening, better visualization and diagnosis of breast lesions. In a preliminary stage, computational tools are used for the evaluation of these new imaging techniques. In this case, besides the computational model of the imaging system, the other important elements are both the anthropomorphic model of the breast and realistic in form and shape 3D lesion models. While the procedure and the know-how for the generation of computational breast models are already well-developed and established by several research groups [13–21], the corresponding ones for the creation of breast lesions are still to be developed. To have realistic breast lesion models, different approaches and sources of information may be exploited. One such approach is based on the mathematical description of the lesion. Simple forms like ellipsoids and spheres are frequently used as a rough approximation of irregular breast lesions [22]. Limited number of works addresses the mathematical modelling of irregular and spiculated breast lesions [14,23–26]. In this development process, main components of the developed algorithms are the use of Brownian motion, diffusion limited aggregation, and modified stochastic Gaussian random sphere model combined with an iterative fractal branching algorithm. Another approach is the use of algorithms for segmentation of breast lesions from 3D medical images: Computed Tomography (CT), Digital Breast Tomosynthesis (DBT), breast CT, Magnetic Resonance Imaging. With the introduction of the breast CT into clinical use, this modality may become a very popular approach for extracting 3D breast lesion shapes. Such an approach for segmenting 3D breast lesions was recently studied by Dukov et al. [27]. Their algorithm was applied successfully on patient DBT images as well as on breast cadavers and whole-body CT scans. Such segmentation approaches, when validated well, are popular in computer-aided detection systems [28,29] and may represent a good approach in creation of computational models for the needs of breast imaging optimization. Another possibility is to use different view 2D images from contrast-enhanced mammography and extrapolate the 2D breast lesion shapes to 3D form. In addition to these techniques, deep learning frameworks (for instance based on convolutional neural networks), were reported to successfully extract features (for instance masses) from both mammography and breast tomosynthesis images [30–32]. Very recently, Caballo et al. [33] reported on the use of machine learning technique for generating 3D super-resolution ($\sim$60 µm voxel size dimensions) breast models based on breast CT data which are characterized with lower resolution compared to data from breast tomosynthesis. This approach may be successfully adopted as well as to segment breast lesions from similar data and generate super-resolution computational breast lesion models, which is an approach to be further explored and studied for its feasibility by researchers. The main constraint of this approach at this stage is that breast CT has still limited availability in clinical practice and scientists' access to data is not such easy.

For research in tumour modelling, breast modelling, as well as modelling, optimising and testing of breast imaging techniques, a large number of simulated, patient and experimental images is needed, these last acquired from physical objects developed by the research teams. Large databases of images would be needed to observe, model, validate and evaluate breast lesions. Therefore, the creation and the technical and scientific support of a database, dedicated to such purposes is of high interest.

This work describes the creation of the MaXIMA Shared Database, which is an initiative that aims to consolidate the efforts of three universities (Technical University of Varna, University of Napoli Federico II and Katholieke University of Leuven) to initially collect, and then gradually add at a single site, datasets of medical x-ray imaging examinations (e.g. mammography, breast tomosynthesis, breast CT). In addition, the database will contain x-ray images or tomograms of different physical (anthropomorphic) phantoms, computational models of breast lesions and finally, a variety of synthesised whole-breast computational phantoms with or without breast lesions inclusions. In the development of the MaXIMA Shared Database, specific efforts have been dedicated to facing and finding a solution for two major challenges: *i*) organizing a common data repository, i.e. to provide storage space and establish an approach for storing the data in a structured way; *ii*) providing a convenient access to the repository data, i.e. development of a user interface for browsing the database content, defining different search criteria, and obtaining the datasets of interest.

2. Materials and methods

2.1. Clinical data

Fifty eight sets of clinical images from patients diagnosed with lesions and images from breast cadavers were collected in the database. Images from tomosynthesis were acquired from Alexandrovska University Hospital, Sofia, Bulgaria, where a Giotto Tomo, IMS (http://www.tomosynthesis-giotto.com/) is in exploitation, and from the University Hospital of Leuven, Belgium, using Siemens Mammomat Inspiration (https://www.healthcare.siemens.com/). The breast cadavers and the whole-body CT scans were implemented at the University Hospital "Saint Marina" in Varna, using SOMATOM (Siemens) CT system. The description of the data is summarised in Table 1.

Two basic approaches for computational models of breast lesions have been elaborated: *i*) segmenting breast x-ray images obtained in a 3D mode (breast CT, tomosynthesis) and *ii*) mathematical modelling.

2.2. Computer models of tumours based on segmentation of 3D patient breast tomosynthesis images

Three-dimensional x-ray breast images may be obtained from breast tomosynthesis and breast cone-beam CT modalities, as well as, whole-body CT scans and breast cadaver samples scanned at CT or micro CT. From these, the most widely available sources are patient data from DBT. Fig. 1 outlines the general approach for segmenting lesions from 3D breast imaging.

In this paper, we will use the algorithm of Dukov et al. [27]. The core of the algorithm is the use of a region-growing method initially adopted by the team for liver segmentation [34], and recently adopted to segment lesions in low contrast DBT. The patient clinical data in the form of a 3D image is the input for the algorithm. Before processing data, a proper anonymizing of patient data is performed. Image filtering is necessary to decrease the artefacts due to the reconstruction algorithm in DBT. A region, containing the breast lesion is selected. Then the segmented lesion is subjected to a set of morphological operations which aim to remove unbounded segmented areas. A region growing method is applied to initially segment the breast lesion. Finally, post-processing is used to correct wrongly segmented tissue parts. Image

Table 1
Clinical patient cases used for the database.

Cases	Clinical system	Image resolution, pixels per mm
10 malignant 4 benign	Giotto tomosynthesis	11.10
36 malignant	Siemens tomosynthesis	11.86
5 malignant	Siemens Somatom Definition AS CT system	3
3 malignant (breast cadaver)	Siemens Somatom Definition AS CT system	3

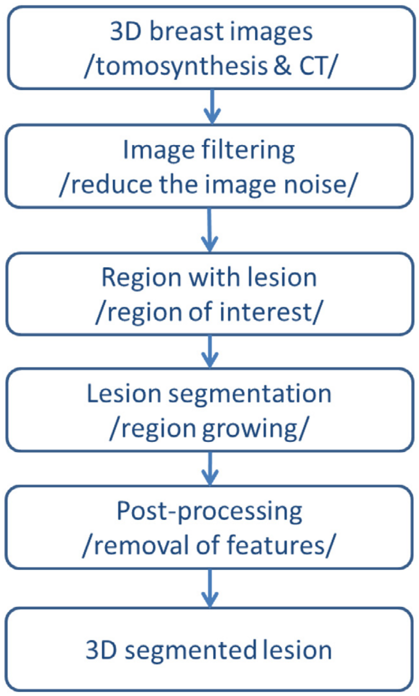

Fig. 1. Steps in creation of the breast lesion volume.

processing operations include the following morphological operations: erosion, area opening, and dilation.

2.2.1. Use of DBT datasets

DBT sets were the major of data source for models of breast lesions in the databases. An example is shown in Fig. 2 and extensively discussed in [27]. In this particular case, data from mediolateral oblique DBT view were exploited from a patient diagnosed with a malignant ILA speculated mass. Resulting images obtained after implementing some of the main steps of the algorithm applied on these DBT images are shown in Fig. 2. The number of images is 83, while the number of segmented slices are 25, the slice thickness is 1 mm and pixel size is 0.086 mm × 0.086 mm.

Initially, a region of interest is defined large enough to cover the breast lesion in all tomosynthesis slices, as shown in Fig. 2a. Then a procedure aiming to transform the images into binary is applied (Fig. 2b). Further, a set of morphological operations processes the initial segmented volume (Fig. 2c), followed by a dedicated region-growing algorithm (Fig. 2d). The final 3D segmented model is shown in Fig. 2e (lesion dimensions: 36.8 mm × 20.6 mm × 25 mm).

2.2.2. Use of low resolution CT datasets

Low resolution full body CT scans may also be used to segment breast lesions. An example is given in Fig. 3 for CT images obtained at the University Hospital "Saint Marina" in Varna, using a SOMATOM unit (Siemens). The acquisition was made by utilizing a standard protocol that provides images of size 512 × 512 pixels, 16 bits grey level resolution, pixel was size 0.98 mm × 0.98 mm. The slice thickness is 3 mm, the number of all slices is 177, while the number of images with the breast lesion is 22.

This type of breast lesion was classified by the medical experts as invasive ductal adenocarcinoma. In this case we applied the algorithm presented in Dukov et al. [27] including the main steps in Fig. 1:

- contrast improvement by using median filter;
- selection of the region of interest where the lesion is present (in order to improve the speed and efficiency of the algorithm) (Fig. 3a, b);
- binarization, leading to lesion segmentation (Fig. 3c);
- corrections for over-segmented regions (Fig. 3d);
- production of slices with segmented breast lesion used for final verification by the medical doctors;
- segmentation of 3D breast lesions (Fig. 3e).

The case shown in Fig. 3 is characterised by low resolution and increased image noise in tomographic images, as well as low contrast between the glandular and cancerous breast tissue due to their similar linear attenuation coefficients [35–38]. In this case the algorithm segmented a slightly larger area than the actual one (Fig. 3c) and a physician was required to evaluate the segmentation area subjectively and to perform small corrections (Fig. 3d). The lesion was then placed into the database.

A similar example is shown in Fig. 4, where we illustrate segmentation of an invasive ductal carcinoma. In this particular case, using the same slice thickness of 3 mm and pixel size of 1.27 mm × 1.27 mm, the breast lesion was generated on the basis of 31 tomograms out of 134 available for this patient.

The result of this segmentation was validated subjectively by our radiologists and oncologists. For this purpose, they inspected first the CT volume and then the segmented breast lesion superimposed on the original tomograms. Overall, our medical experts were very satisfied with the segmentation results.

2.2.3. Use of CT datasets of breast cadavers

Similarly to the previous cases, CT sets from breast cadavers may also be used as a source of models for breast lesions. An example is shown in Fig. 5, first column for a cadaver breast obtained at the University Hospital "Saint Marina" in Varna, and scanned immediately at the CT facility. The acquisition protocol includes 264 images of size 910 pixels × 512 pixels, slice thickness 1.5 mm. The images have square pixels of size 0.98 mm × 0.98 mm. Sixteen images contained an invasive ductal carcinoma. Following the segmentation procedure (highlighted in Fig. 5a–c, first column), the breast lesion was segmented as shown in Fig. 5d, first column.

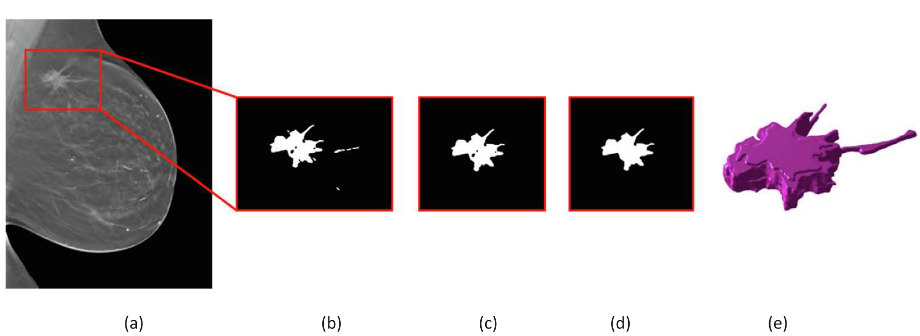

Fig. 2. Main steps in the creation of a breast lesion from DBT images: (a) selection of region of interest, (b) obtaining binary image, (c) artefact reduction, (d) solid lesion (e) segmented tumour.

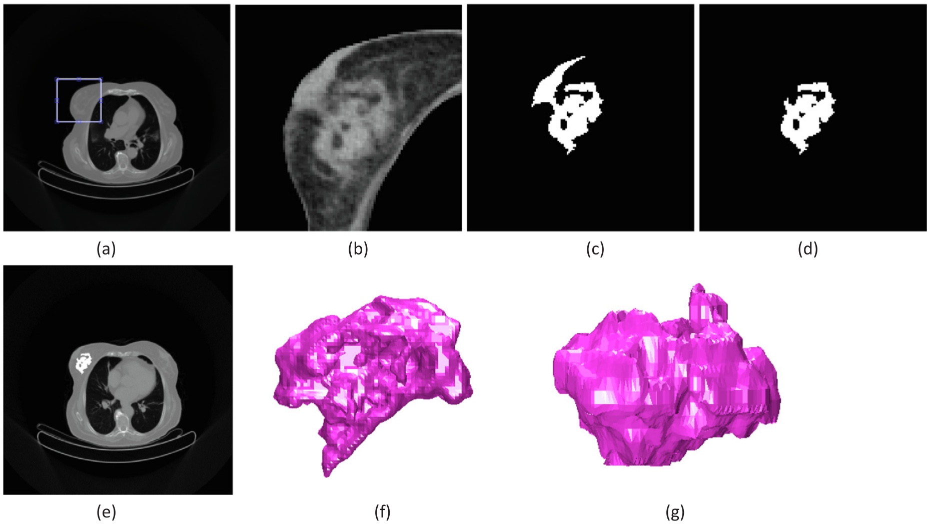

Fig. 3. Segmenting a difficult case in the breast from CT images of a patient diagnosed with invasive ductal adenocarcinoma: (a) CT tomogram, (b) region of interest where the breast lesion is located, (c, d) segmenting the lesion, (e) segmented area overlaid on the original tomogram, (f, g) segmented breast lesions (lesion dimensions: 66.4 mm × 54.7 mm × 42.0 mm).

Similarly, the second column in Fig. 5 shows the procedure applied on a CT set derived from another cadaver, scanned under the same conditions. The acquisition protocol included 100 breast tomography images of size 733 × 512 pixels (slice thickness 0.6 mm, image pixel size is 0.32 mm × 0.32 mm). Twenty seven images contained a breast lesion.

2.3. Mathematical breast lesions

The approach for generation of mathematical breast lesions models is based on a method reported in refs. [14] and further refined in [25,39]. Its basis is the 3D random walk algorithm (similarly used also by other researchers [19,26,40]), generated in a predefined 3D volume, which in fact is an empty three-dimensional matrix. Since in our case the random walk algorithm is applied on voxel objects, this algorithm is modified to the so called nearest neighbour random walk algorithm. The main steps in generating breast lesion models are shown in Fig. 6a.

The tumour size is a function of the size of the voxel matrix denoted as M × M × M and the voxel resolution, defined by the user. Initially, the voxel values of the lesion matrix are set to zero. The user assigns a number of random walks: this number is marked as $\mathbf{N_{brownian\_runs}}$. The random walk starts from the centre of the matrix and each step moves randomly to one of the neighbouring voxels, assigning to it the lesion's elemental composition. Each random walking process stops either at the matrix boundaries or when the assigned maximum number of discrete steps, $\mathbf{N_{run\_length}}$ (number of voxels per walk), is reached. The received "dusty" structure is converted into a lesion with a solid geometry by applying further processing: averaging, dilation, erosion morphologic operations. In all operations, the structuring element is the cube. For instance, for the dilation operation, the repeated dilation process is obtained with a larger cube size (5 × 5 pixels) followed by dilation using a smaller cube size (3 × 3 pixels), while averaging is reached with uniform averaging filter. Other morphological image processing methods – like closing and morphological smoothing (opening followed by closing) – have also cube as a structuring element. The size of the structure element as well as its shape can be changed by the user. Closing and dilations were used also for binary image processing.

Fig. 6b shows the shape of the created breast lesion after the implementation of each step from Fig. 6a for the case of M = 100, $\mathbf{N_{brownian\_runs}}$ = 100, $\mathbf{N_{run\_length}}$ = 2000. By changing the model parameters (number of walks $\mathbf{N_{brownian\_runs}}$, and the number of discrete steps per walks $\mathbf{N_{run\_length}}$, degree of averaging, dilation and erosion) the user can change the shape of the modelled lesion as well as the degree of malignancy. Examples of generated irregularly shaped lesions are shown in Fig. 11.

2.4. Database creation

The creation of the MaXIMA Shared Database has been conducted in four phases: initial collection of imaging data; analysis of the variability of the data (actual and potential) and taking decision for the database implementation approach; creation of database repository; user interface development.

2.4.1. Initial data collection

One of the major challenges in the creation of the MaXIMA Shared Database has been the large variety of medical images that would be used as sources for extracting real lesion shapes. Therefore, as a first step, we collected samples of different medical images with the basic objective to create an overview of the potential variety of the medical data to be collected.

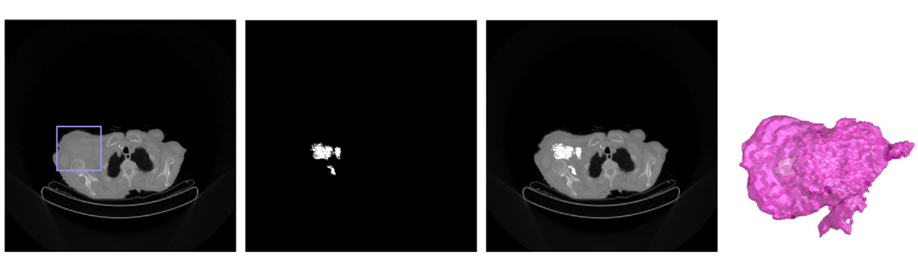

Fig. 4. Segmenting invasive ductal carcinoma from patient CT images: (a) CT tomogram with the region of interest where the breast lesion is located, (b) segmented breast lesion, (c) segmented area superimposed on the original tomogram, (d) segmented breast lesion (lesion dimensions: 129.5 mm × 138.4 mm × 93 mm).

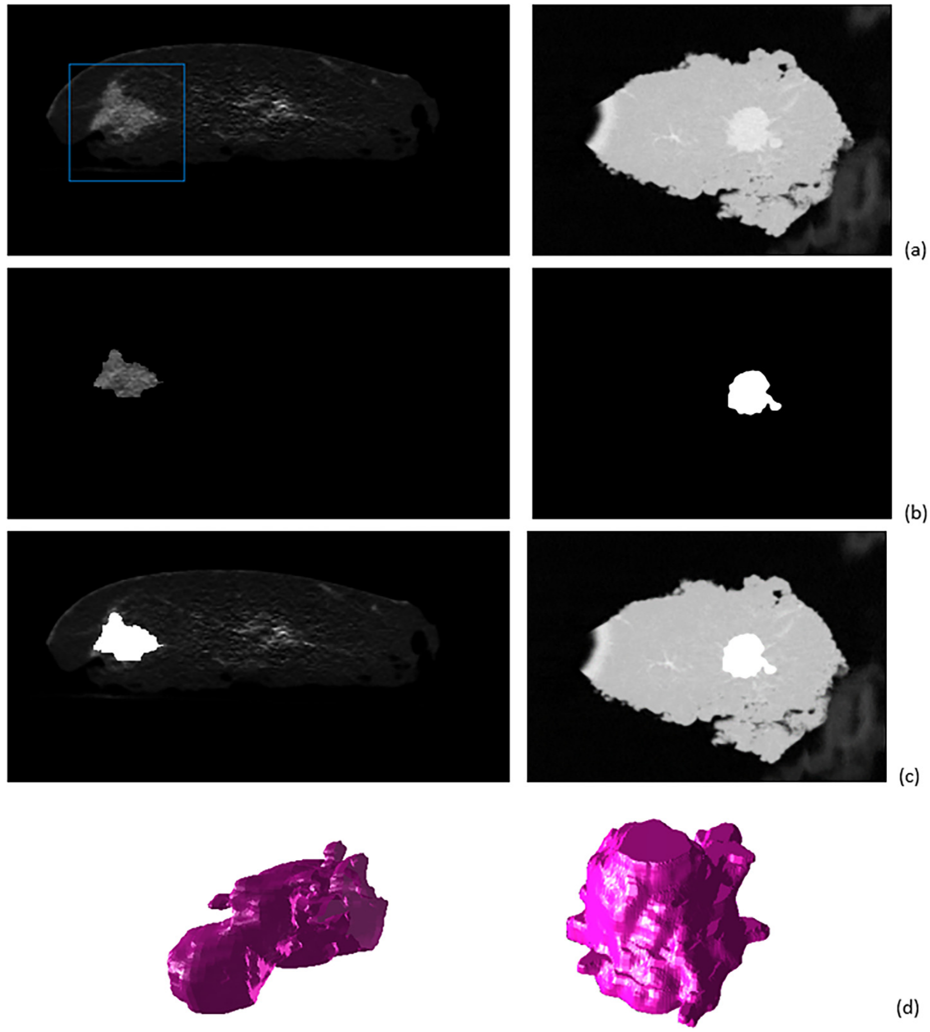

Fig. 5. Segmenting two malignant breast lesions from breast CT of two breast cadavers (the two columns): (a) selected original tomogram, (b) segmented lesion area, (c) segmented area superimposed on the original tomography slice and (d) segmented 3D breast lesions (size of the left lesion: 34.0 × 24.5 × 21 mm, size of the right lesion: 37.6 mm × 42.7 mm × 16.2 mm).

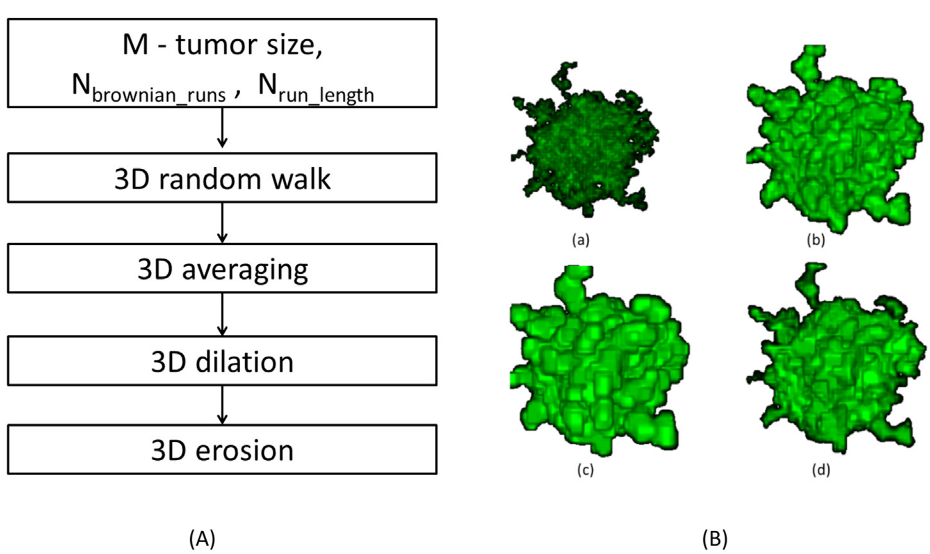

Fig. 6. Creation of breast lesions: (A) Steps in the creation of breast lesions, and (B) breast lesion models created for M = 100; $N_{brownian\_runs}$ = 100; N_{run_length} = 2000, obtained after implementing (a) initial random walks, (b) 3 × 3 × 3 averaging of the volume, (c) dilation with a 3 × 3 × 3 and (d) 5 × 5 × 5 erosion.

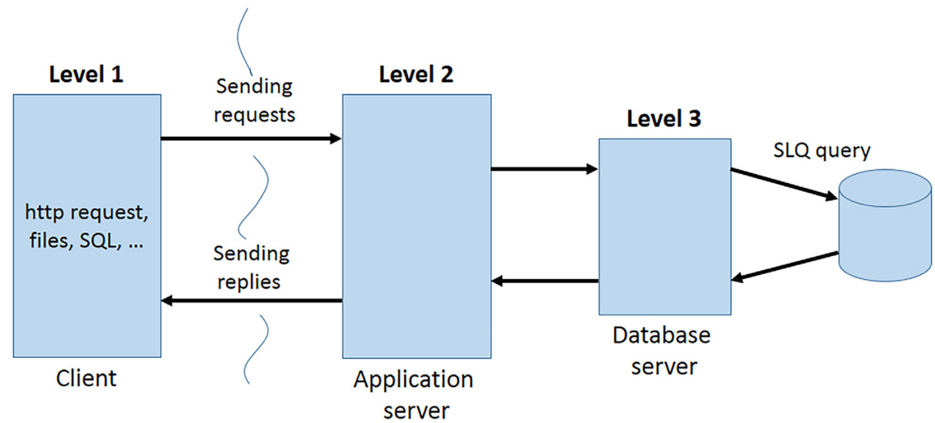

Fig. 7. Client-server architecture of the Maxima database application.

2.4.2. Selected approach

The preliminary analysis of the variety of the initially collected imaging data determined the way of organizing the storage and its size: *i) Storage place* – a remotely accessible dedicated server is used for the storage of the computational breast lesion models, *ii) Database architecture* – the structure of the storage space is additionally indexed, allowing filtered searching and execution of queries, *iii) User interface* – a web-based user interface was chosen for connecting and querying databases remotely.

2.5. Implementation

The MaXIMA Imaging Database Application uses client-server architecture (Fig. 7), where the data from the Server application is extracted, computed and delivered to the Users in an appropriate format via client applications. The advantages of the client-server architecture are: *i*) centralized resources, common for all users; *ii*) improved security; *iii*) server-level administration; *iv*) scalability.

The organization of the Maxima database is relational – the information is stored in tables interconnected by fields containing connection information (Fig. 8). The tables in the database have primary and secondary indexes which control the operation of the database. The primary key is a unique record identifier in the table that defines each record, while the secondary key is a pointer to a primary from another table.

The Maxima database contains 4 tables (shown in Fig. 8a): (a) *patients* – containing a list of records for 3D images and their attributes; (b) *rstype* – containing a list of imaging modalities with an additional information; (c) *medorg* – containing a list of data sources with corresponding information; (d) *users* – containing a list with an information for the users who can access the application. All four tables are connected to each other by means of primary and secondary indexes, creating the relations in the database.

2.6. User application

Participating researchers can use the application to manage the imaging database. Only registered users can access the database, using one of the two possible access options – User and Administrator. Regular Users can view and search records in database, use file browser to view and download images, view and search data sources and imaging modalities. Administrators can add, edit and remove Users, change User types, add, edit and remove records and images from/to the database and add, edit and remove data sources and imaging modalities.

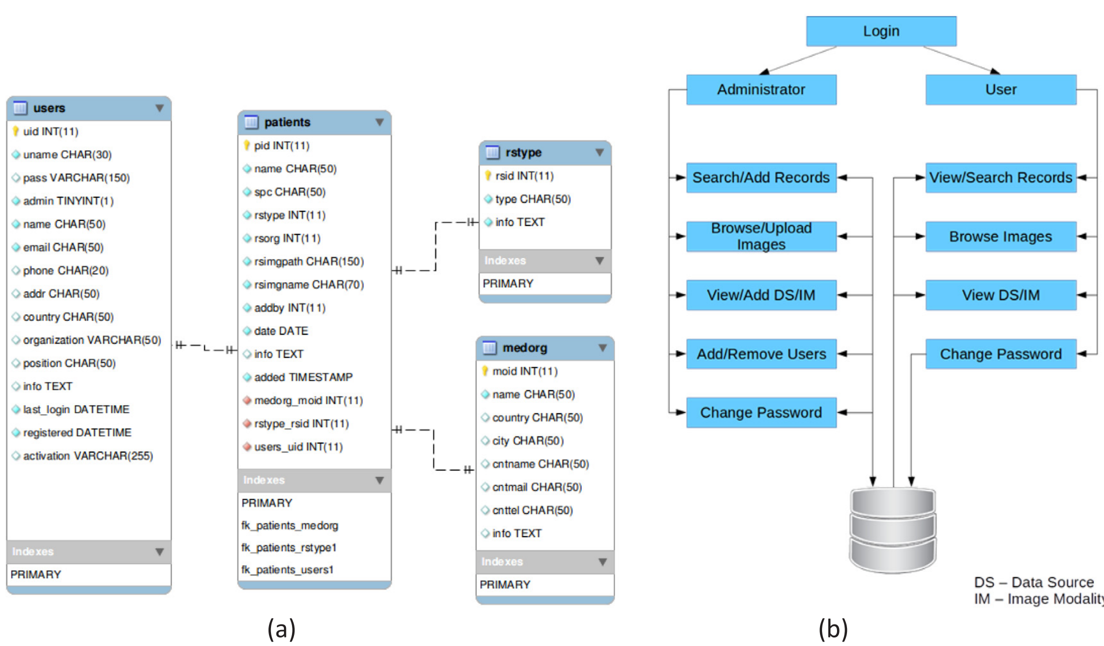

Fig. 8. Application architecture: (a) relational model of the database application and (b) structure of the application.

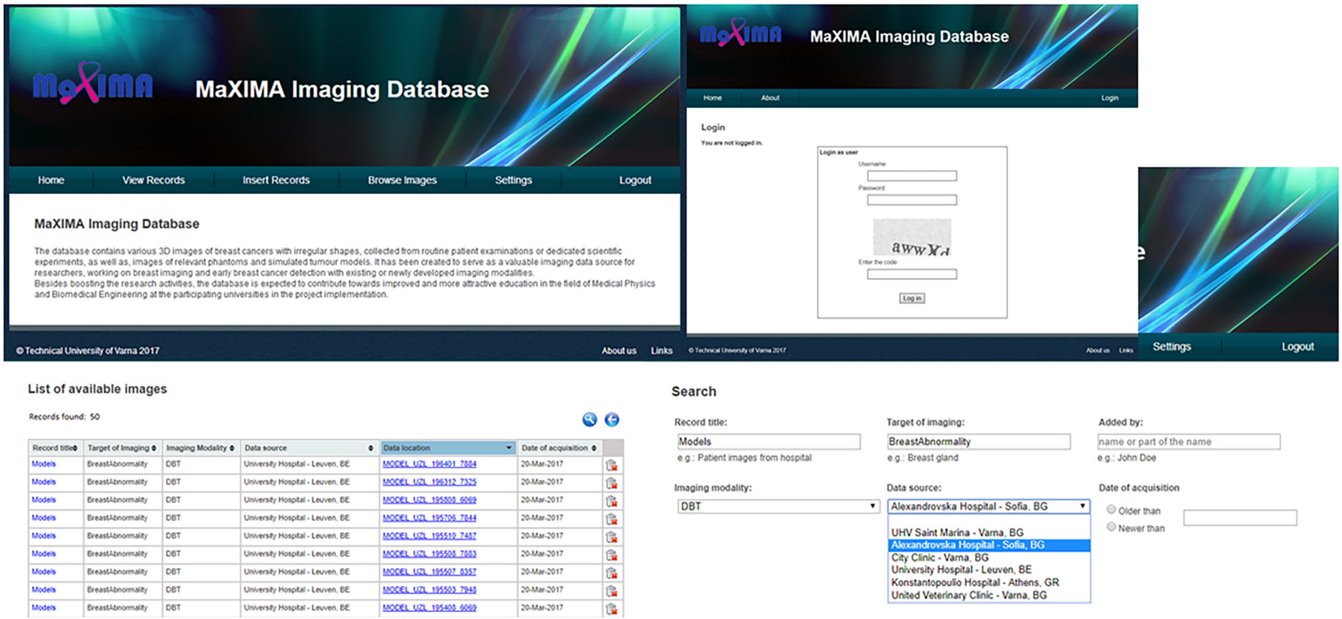

Fig. 9. A screenshot from the Maxima database web application (http://maxima-tuv.eu/database/).

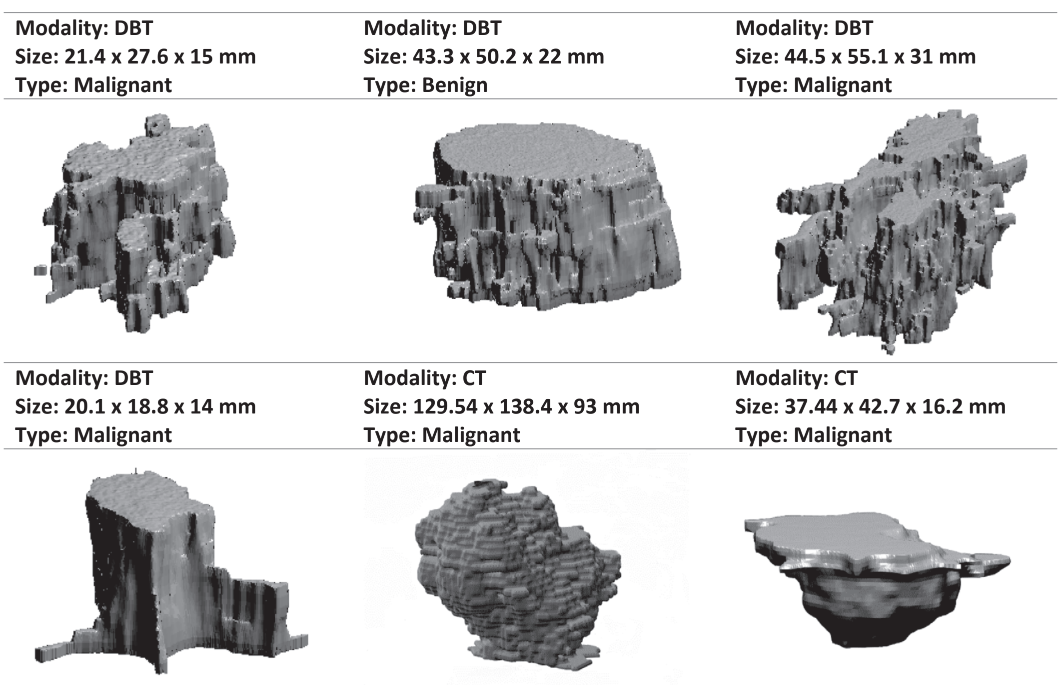

Fig. 10. Segmented lesion models, based on patient CT and breast tomosynthesis data.

The relational model for this database is shown in Fig. 8b.

3. Results and discussion

3.1. Maxima database

The database interface is web-based (Fig. 9) and hence platform independent, user friendly and it can be accessed from the official project website (http://maxima-tuv.eu/database/) or through a direct link (http://maxima.tu-varna.bg/). In order to obtain access to the database, a user needs to request access by submitting a short justification e-mail.

The MaXIMA Imaging Database currently contains the following datasets: 138 breast cancer models obtained by segmentation of 50 three-dimensional patient breast tomosynthesis images; 8 models obtained by segmentation of 3D micro CT images of biopsy specimens; and 80 models based on mathematical algorithm. The last group of lesion models is expected to be the largest in the database. The parametrical description of the breast lesions allows to apply a combinatorial approach for the generation of a large number similar though different tumour models. The same is valid for the computational breast phantoms without abnormalities. The variety of both types of models would result in a potentially huge number of realistic breast models with included abnormalities. Such a possibility and approach provide the virtual breast imaging studies with two important object sources: *i)* Completely new computational models that the researches may

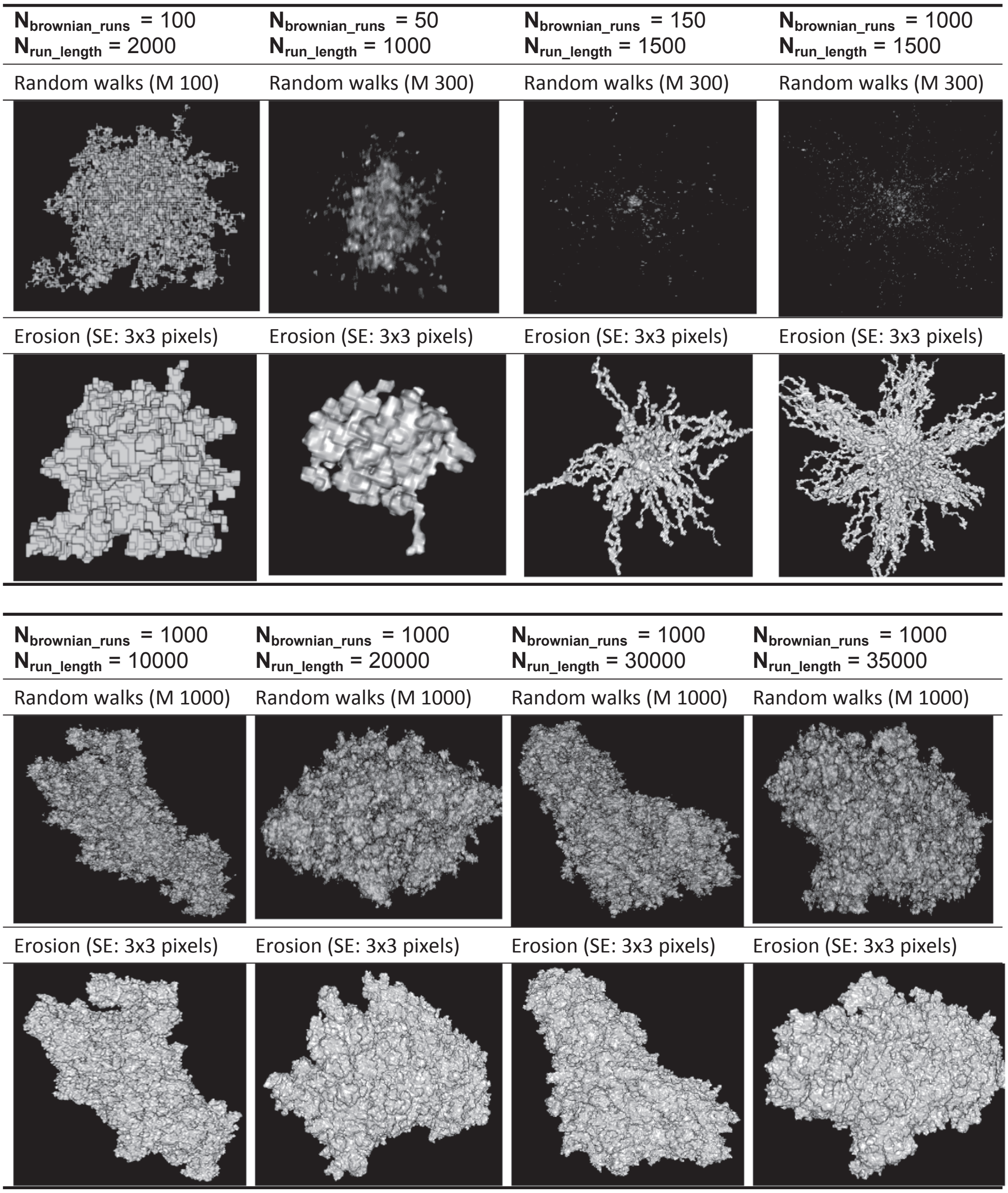

Fig. 11. Irregular breast lesions obtained with unmodified random walk algorithm using different model describing parameters. SE – structure element used in morphological operations.

generate following the guidelines for breast and lesion modelling and using their own sets of model parameters; *ii*) A large set of different types of breast and/or lesion models or both, provided in groups corresponding to clinically observed cases or cases of interest. The groups have similar properties but contain different cases due to the statistical variance of the generated models; *iii*) As all properties of the generated models will be known, this type of sets may serve as a unique annotated and reference set for a large number of virtual studies. Examples for computational models created from segmentation and by applying the mathematical algorithm are shown in Figs. 10 and 11, respectively, accompanied by a short description: a detailed description is given in a readme file.

Segmented 3D volumes, containing the segmented shape and related relevant information about it is created and saved in a separate MATLAB data file. To facilitate further processing, analysis and computations, two types of files are available. The first one contains the segmented lesion (in gray values) and the necessary information related to the file structure named *info structure* (available also as a text file). The second file type contains two sets of images – the original slices and the binary images of the segmented lesion shape, as well as a description related to these file structures. The file structure always contains the lesion model's name, the volume and voxel dimensions, the geometrical center of the lesion model, displayed range and the original pixel values.

The segmented breast lesion images were then stacked by considering the aspect ratio of the pixels in the tomographic images and the spatial distribution of tomographic slices. Further, they are placed on a database repository (www.maxima-tuv.eu). The lesion shapes are also saved as raw data in binary file, accompanied with general information, which will permit the data handling.

3.2. Applications

Description of a breast lesion is an important part of the realistic

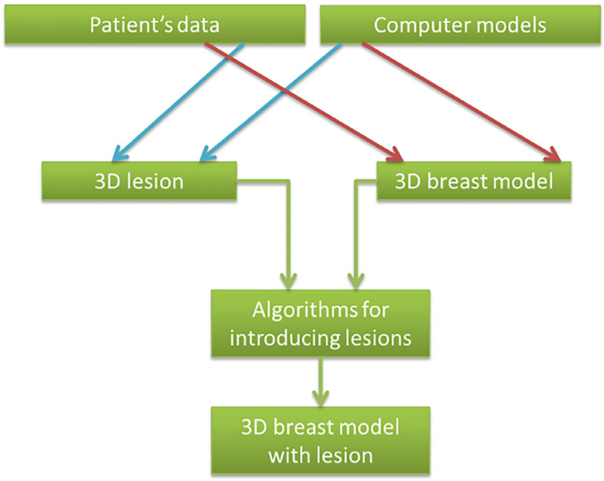

Fig. 12. Block-diagram of the procedure for introducing breast lesions in healthy breast tissue phantoms.

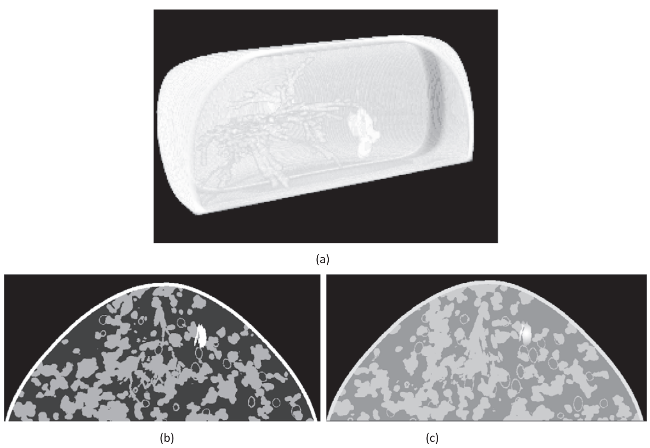

Fig. 13. Breast lesion integration: (a) breast model with the lesion shown in Fig. 4d, and (B, C) a slice from the combined breast model with two integration methods. The different contrast between the slices shown in (b) and (c) is due to the fact that in (c) the values in the slice represent attenuation coefficients, while in (b) the values are numbers that correspond to a given elemental composition (for instance: adipose tissue – 203, glandular tissue – 216, skin – 234, lesion – 235). The latter is needed when Monte Carlo simulations is used with custom Monte Carlo codes for modelling and simulation of radiation transport.

breast modelling. Breast lesion models are of interest and are built to be incorporated into breast models either existing or under development, to foster the realisation of reliable virtual studies in the field of breast imaging and cancer detectability and diagnosis. Two applications selected from the research field of the Maxima group are demonstrated bellow.

3.2.1. Procedure for introducing the breast lesions into the breast

Computational models of breast lesion models are usually inserted into computationally generated breast phantoms Fig. 12. The resulted combined computational breast phantoms are used to study the visibility of breast lesions in 3D breast imaging techniques. The introduction of the breast lesions into the generated breast models is a crucial task, which defines the realism of the projected anatomic details onto the 2D image (and subsequently tomosynthesis image) model breast lesion. For this important aspect of the whole breast model procedure, methods for integration of created abnormalities need to be developed and implemented. For the current and future research and educational activities, the existence of user-selected methods for inserting lesions in generated breast phantoms is a prerequisite for a good work. Therefore, suitable algorithms used to insert the modelled breast lesions are of high importance in order to obtain realistic 3D breast phantoms with these lesions. This approach allows the implementation of multiple scenarios and unlimited number of cases, which can be used for further software modelling and investigation of breast imaging techniques.

The computational models of the breast lesions are available as three dimensional voxel based matrices with values 1 for the lesion and 0 for the background. The user selects the position in the breast volume where the breast lesion will be integrated. Two approaches for insertion of breast masses are shown in Fig. 13. In the first approach, lesion models are matched to the properties of the glandular tissue or water [37,38]. All voxels of the breast lesion have the same elemental composition and are assigned attenuation coefficients equal to those of glandular tissue or water (as shown in Fig. 13b). The second approach uses smooth integration of lesion models within the breast. The user selects the place in the breast volume (phantom) where the breast abnormality is integrated. Then the properties of the abnormality are changed in such a way that the abnormality is smoothly placed into the surrounding tissues within the breast volume (as shown in Fig. 13c). The degree of smoothing of each voxel from the abnormality matrix depends on its distance from the centre. The more distant the voxels are, the higher degree of smoothing is achieved (as described in Appendix B3 in [14]).

The breast lesion is the irregular breast lesion model presented in Fig. 4d. This breast lesion was introduced in the computational breast phantom, generated by the in-house developed and validated software tool *BreastSimulator* [41]. The later tool can create models of healthy breasts, by varying parameters like shape, size, duct tree features, Cooper ligaments, skin, etc. The generated, in this case, 3D breast phantom is of a medium in size with a 25% glandular tissue without skin (and 33% with skin, respectively). The number of Cooper ligaments (presented as ellipsoids) is approximately 1000, while the glandular tree is composed of about 8000 cylinders. The model is then transformed into a 3D breast voxel matrix, which contained glandular, adipose and skin tissues. Consequently, the model is computationally subjected to a compression procedure in order to simulate the compression procedure during the mammography imaging [42], as the main parameters include the desired breast compressed thickness and the elastic modulus of the breast tissues. The compression simulation resulted in a breast phantom with a thickness of 4 cm. The selected breast lesion model is then inserted into the computational breast model and the resulting combined breast model is shown in Fig. 13a, while Fig. 13b,c shows a selected slice that contains the lesion inserted into the computational breast model by using both approaches, respectively.

3.2.2. Procedure for evaluation of modelled 3D breast lesions

The procedure for the generation of a large number of projection images with breast lesions for evaluation purposes is also of great importance. The approach, demonstrated in Fig. 14, is the one of the leading research groups [43,44]: projections of targets (the breast lesions) are simulated and then inserted at different positions in the clinical mammograms.

The main steps are illustrated in Fig. 14 and include: *i)* generation of x-ray projection images of 3D breast lesions by using an in-house developed software application [45], capable to simulate the x-ray transport through the computational lesions. For this purpose, the incident spectra, and geometry: distances from the source to the object and detector need to be modeled. Usually, analytical relationship between the source intensity of the x-rays and the intensity registered at the detector surface is exploited. The transmitted intensity reaching the detector is calculated using the Beer's law. Ideal images are then convolved with the detector features. *ii)* The generated image is then convolved with an anonymized planar and free of breast lesions patient mammography image and used in evaluation assessments (research)

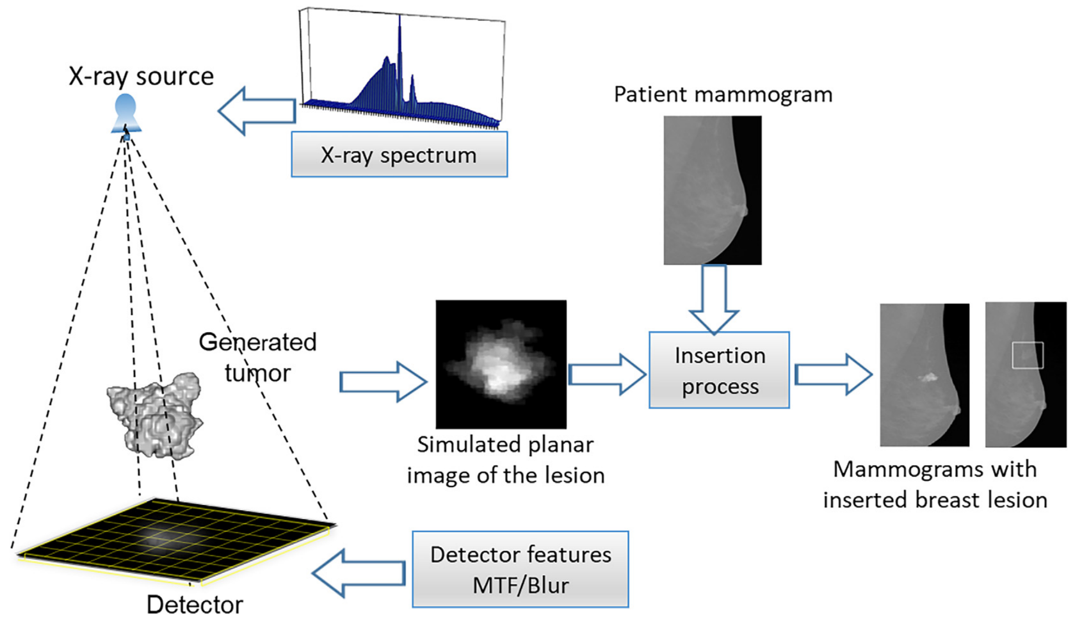

Fig. 14. Main steps in creating images with breast lesions. For visualization purposes, the projection image of the lesion is enhanced in terms of contrast to make the lesion more visible (left image in mammograms with inserted breast lesion) compared to the real lesion insertion (right image in mammograms with inserted breast lesion).

and in educational activities. In the demonstrated case, the elemental composition of the lesion was assigned to water, and the obtained simulated planar image is shown in Fig. 14.

4. Conclusions

Both computational and physical models play a significant role in breast imaging research. In particular, they are very important when new technology is under design, development, testing and optimization. This paper addressed the development of a database application dedicated to three-dimensional models of breast cancers segmented from *i*) x-ray three-dimensional patient breast images as well as whole body and breast cadavers CT images, and *ii*) mathematical algorithms developed to generate breast lesions, with irregular shape. The developed database will serve as an imaging data source for researchers, working on breast imaging and early breast cancer detection. In addition to boosting the research activities, the database is expected to represent also an attractive educational tool in the general field of Medical Physics and Biomedical Engineering.

Acknowledgments

This work is supported by both the Bulgarian National Science Fund under grant agreement DN17/2 and the MaXIMA project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 692097.

References

[1] Lewin J. Comparison of contrast-enhanced mammography and contrast-enhanced breast MR imaging. Magn Reson Imaging Clin N Am 2018;26(2):259–63.
[2] Patel BK, Lobbes MBI, Lewin J. Contrast enhanced spectral mammography: a review. Semin Ultrasound CT MR 2018;39(1):70–9.
[3] Sarno A, et al. A Monte Carlo model for mean glandular dose evaluation in spot compression mammography. Med Phys 2017;44(7):3848–60.
[4] Contillo A, et al. A novel approach to background subtraction in contrast-enhanced dual-energy digital mammography with commercially available mammography devices: Polychromaticity correction. Med Phys 2015;42(11):6641–50.
[5] Contillo A, et al. A novel approach to background subtraction in contrast-enhanced dual-energy digital mammography with commercially available mammography devices: Noise minimization. Med Phys 2016;43(6):3080–9.
[6] Taibi A, et al. Dual-energy imaging in full-field digital mammography: a phantom study. Phys Med Biol 2003;48(13):1945–56.
[7] Sechopoulos I. A review of breast tomosynthesis. Part I. The image acquisition process. Med Phys 2013;40(1):014301.
[8] Sechopoulos I. A review of breast tomosynthesis. Part II. Image reconstruction, processing and analysis, and advanced applications. Med Phys 2013;40(1).
[9] Sarno A, et al. Dose volume distribution in digital breast tomosynthesis: a phantom study. IEEE Trans Radiat Plasma Med Sci 2017;1(4):322–8.
[10] Sarno A, et al. Monte Carlo evaluation of glandular dose in cone-beam X-ray computed tomography dedicated to the breast: Homogeneous and heterogeneous breast models. Phys Med 2018;51:99–107.
[11] Sarno, A., et al., A Framework for Iterative Reconstruction in Phase-contrast Computed Tomography Dedicated to the Breast. 2016 Ieee Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (Nss/Mic/Rtsd), 2016.
[12] Longo R, et al. Towards breast tomography with synchrotron radiation at Elettra: first images. Phys Med Biol 2016;61(4):1634–49.
[13] Bakic PR, Albert M, Brzakovic D, Maidment AD. Mammogram synthesis using a 3D simulation. I. Breast tissue model and image acquisition simulation. Med Phys 2002;29(9):2131–9.
[14] Bliznakova K, et al. A three-dimensional breast software phantom for mammography simulation. Phys Med Biol 2003;48(22):3699–719.
[15] Kiarashi N, et al. Development and application of a suite of 4-D virtual breast phantoms for optimization and evaluation of breast imaging systems. IEEE Trans Med Imaging 2014;33(7):1401–9.
[16] Elangovan P, et al. Design and validation of realistic breast models for use in multiple alternative forced choice virtual clinical trials. Phys Med Biol 2017;62(7):2778–94.
[17] Bliznakova K, Suryanarayanan S, Karellas A, Pallikarakis N. Evaluation of an improved algorithm for producing realistic 3D breast software phantoms: application for mammography. Med Phys 2010;37(11):5604–17.
[18] Yang XI, Sechopoulos B. Fei Automatic tissue classification for high-resolution breast CT images based on bilateral filtering. Lake Buena Vista, Florida, USA: SPIE; 2011.
[19] Youn H, et al. Numerical generation of digital mammograms considering imaging characteristics of an imager. Nucl Instrum Methods Phys Res Sect 2011;652(1):810–4.
[20] Huang SY, et al. The characterization of breast anatomical metrics using dedicated breast CT. Med Phys 2011;38(4):2180–91.
[21] Li CM, et al. Methodology for generating a 3D computerized breast phantom from empirical data. Med Phys 2009;36(7):3122–31.
[22] Saunders R, Samei E, Baker J, Delong D. Simulation of mammographic lesions. Acad Radiol 2006;13(7):860–70.
[23] De Sisternes L, et al. A computational model to generate simulated three-dimensional breast masses. Med Phys 2015;42(2):1098–118.
[24] Elangovan P, et al. Simulation of spiculated breast lesions. Medical Imaging 2016: Physics of Medical Imaging. SPIE; 2016.
[25] Hintsala H, Bliznakova K, Pallikarakis N, Jamsa T, Modelling of Irregular Breast Lesions. In: World Congress on Medical Physics and Biomedical Engineering, Vol 25, Pt 4: Image Processing, Biosignal Processing, Modelling and Simulation, Biomechanics, 2010. 25: p. 2024–2027.
[26] Rashidnasab A, et al. Simulation and assessment of realistic breast lesions using

fractal growth models. Phys Med Biol 2013;58(16):5613–27.
[27] Dukov N, et al. Models of breast lesions based on three-dimensional X-ray breast images. Phys Med Eur J Med Phys 2019;57:80–7.
[28] Ganesan K, et al. Computer-aided breast cancer detection using mammograms: a review. IEEE Rev Biomed Eng 2013;6:77–98.
[29] Qian W, Sun W, Zheng B. Improving the efficacy of mammography screening: the potential and challenge of developing new computer-aided detection approaches. Expert Rev Med Dev 2015;12(5):497–9.
[30] Samala RK, et al. Mass detection in digital breast tomosynthesis: deep convolutional neural network with transfer learning from mammography. Med Phys 2016;43(12):6654.
[31] Yousefi M, Krzyzak A, Suen CY. Mass detection in digital breast tomosynthesis data using convolutional neural networks and multiple instance learning. Comput Biol Med 2018;96:283–93.
[32] Jiao ZC, Gao XB, Wang Y, Li J. A deep feature based framework for breast masses classification. Neurocomputing 2016;197:221–31.
[33] Caballo M, et al. Development of 3D patient-based super-resolution digital breast phantoms using machine learning. Phys Med Biol 2018;63(22):225017.
[34] Bliznakova K, et al. Computer aided preoperative evaluation of the residual liver volume using computed tomography images. J Digit Imaging 2015;28(2):231–9.
[35] Tomal A, Poletti ME, Caldas LV. Evaluation of subject contrast and normalized average glandular dose by semi-analytical models. Appl Radiat Isot 2010;68(4–5):755–9.
[36] Carroll FE, et al. Attenuation of monochromatic X-rays by normal and abnormal breast tissues. Invest Radiol 1994;29(3):266–72.
[37] Chen RC, et al. Measurement of the linear attenuation coefficients of breast tissues by synchrotron radiation computed tomography. Phys Med Biol 2010;55(17):4993–5005.
[38] Johns PC, Yaffe MJ. X-ray characterisation of normal and neoplastic breast tissues. Phys Med Biol 1987;32(6):675–95.
[39] Gospodinova G, Bliznakova K, An approach of modelling of breast lesions, in Proceedings of the 1st International Conference “Applied Computer Technologies” 2018, University of Information Science and Technology “St. Paul the Apostle” – Ohrid, Macedonia: Ohrid, 21-23 June. p. 149-153.
[40] Elangovan, P., et al., Simulation of spiculated breast lesions. Medical Imaging 2016: Physics of Medical Imaging, 2016. 9783.
[41] Bliznakova K, Sechopoulos I, Buliev I, Pallikarakis N. BreastSimulator: a software platform for breast x-ray imaging research. J Biomed Graph Comput 2012;2(1):1–14.
[42] Zyganitidis C, Bliznakova K, Pallikarakis N. A novel simulation algorithm for soft tissue compression. Med Biol Eng Comput 2007;45(7):661–9.
[43] Elangovan P, et al. Lesion detectability in 2D-mammography and digital breast tomosynthesis using different targets and observers. Phys Med Biol 2018;63(9).
[44] Shaheen E, et al. The simulation of 3D microcalcification clusters in 2D digital mammography and breast tomosynthesis. Med Phys 2011;38(12):6659.
[45] Bliznakova K, et al. Experimental validation of a radiographic simulation code using breast phantom for X-ray imaging. Comput Biol Med 2010;40(2):208–14.