

A finite element-based machine learning approach for modeling the mechanical behavior of the breast tissues under compression in real-time

F. Martínez-Martínez^{a,*}, M.J. Rupérez-Moreno^b, M. Martínez-Sober^a, J.A. Solves-Llorens^c,
D. Lorente^a, A.J. Serrano-López^a, S. Martínez-Sanchis^b, C. Monserrat^c, J.D. Martín-Guerrero^a

^a Intelligent Data Analysis Laboratory (IDAL), University of Valencia, Av. de la Universidad s/n, 46100 Burjassot (Valencia), Spain

^b Centro de Investigación en Ingeniería Mecánica (CIIM), Departamento de Ingeniería Mecánica y de Materiales, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain

^c Departamento de Sistemas Informáticos y Computación, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain

ARTICLE INFO

Keywords:

Breast biomechanics
Finite element methods
Machine learning
Modeling
Breast compression

ABSTRACT

This work presents a data-driven method to simulate, in real-time, the biomechanical behavior of the breast tissues in some image-guided interventions such as biopsies or radiotherapy dose delivery as well as to speed up multimodal registration algorithms. Ten real breasts were used for this work. Their deformation due to the displacement of two compression plates was simulated off-line using the finite element (FE) method. Three machine learning models were trained with the data from those simulations. Then, they were used to predict in real-time the deformation of the breast tissues during the compression. The models were a decision tree and two tree-based ensemble methods (extremely randomized trees and random forest). Two different experimental setups were designed to validate and study the performance of these models under different conditions. The mean 3D Euclidean distance between nodes predicted by the models and those extracted from the FE simulations was calculated to assess the performance of the models in the validation set. The experiments proved that extremely randomized trees performed better than the other two models. The mean error committed by the three models in the prediction of the nodal displacements was under 2 mm, a threshold usually set for clinical applications. The time needed for breast compression prediction is sufficiently short to allow its use in real-time (<0.2 s).

1. Introduction

Breast cancer is one of the major causes of mortality and morbidity in women today. Its mortality is related to the tumor size and the time of detection [1]. A wide range of imaging methods are available for detection and diagnosis, from projection X-ray mammography (XRM), magnetic resonance imaging (MRI), and ultrasound (US), to the more recent technologies such as digital breast tomosynthesis (DBT), positron emission mammography (PET) and ultrasound tomography (UST) [2]. However, each imaging modality displays the information in a different way and in consequence, undetectable tumors in one modality can be detected by other modalities, and lesions considered to be suspicious in one modality can be classified as healthy by others [3]. Radiologists have found that a combination of various imaging methods together with the histopathological information of a lesion can lead to a more effective diagnosis [4,5]. However, performing such a multimodality analysis is a complex task. The breast is a highly deformable structure that adopts

different forms depending on the patient's positioning in each one of the imaging modalities. The patient is in prone position during an MRI acquisition, standing up during an XRM or in supine position during an US scanning process. Moreover, the breast is compressed by the mammographic plates in an XRM or by the US probe during an US scanning process. It is also immobilized by plates during some types of biopsies such as MRI-guided biopsy. This complicates the visual comparison of the imaging modalities for the purposes of breast screening, cancer diagnosis (including image guided biopsy), tumor staging, as well as for treatment monitoring, surgical planning and simulation of the effects of surgery and wound healing, etc. [2]. Therefore, the development of non-rigid registration algorithms based on the simulation of the biomechanical behavior of the breast tissues is nowadays an active field of research (see Ref. [2] for a review). These algorithms allow the fusion of information from different breast imaging modalities. However, some limitations in the simulation of the breast biomechanical modeling have prevented the rapid transition of fusion algorithms to clinical practice.

* Corresponding author.

E-mail address: francisco.martinez-martinez@uv.es (F. Martínez-Martínez).

These limitations include establishing a realistic constitutive model for the breast tissues, obtaining their patient specific elastic constants, using the adequate boundary conditions, or achieving clinically acceptable levels of computational cost.

The finite element method (FEM) is one of the most common numerical methods employed in Biomechanics to model deformation of the soft tissues. FEM approximates a discrete solution of the equations associated with a mechanical behavior on a complex discrete geometry called mesh (formed by nodes and elements) [6]. FEM has been used to simulate the breast compression by many authors [7–13] and more (see Ref. [2] for a review) due to the considerable interest in capturing the transformation of the breast among the different imaging modalities and among the breast images and the physical space of an intervention such as biopsy or surgical resection. For example, in Ref. [13] very good results were obtained for the registration between MR breast images and X-ray images. The committed error for the detection of small lesions suspicious to be a tumor was about 5 mm, which is very low and it is accepted clinically. One of the main challenges in transitioning the simulation methods to clinical practice is the computation time. The computational cost of FEM is proportional to the size and complexity of the finite element (FE) mesh, as well as the complexity of the problem to solve. As an example, a computation time of approximately 120 min was needed in Refs. [11] and [13] to perform a FE simulation of the breast tissue deformation under compression. This is not practical for diagnosis, where the radiologists need a faster answer in the comparison of different image explorations. The same occurs for planning an intervention or for a computer-aided surgery, where it is crucial to calculate tissue deformations in real-time.

Admittedly, real-time FEM has already been proposed by different authors; it combines pre-computed deformations [14] or/and uses parallel processing [15–17]. Currently, fast GPU-based FE solvers have been proposed in the literature. For instance, NiftySim. NiftySim is a GPU-based nonlinear finite element package for simulation of soft tissue biomechanics [18]. NiftySim has been applied to the simulation of the soft tissue interactions with other organs and instrumentals. For instance, Eiben et al. [19] aimed to improve the results of registration of breast MRI from a prone to a supine patient position. In their experiments, the algorithm required 19 simulations to converge both from the supine and prone configurations to the zero-gravity reference configuration. The simulations took on average 80 and 83 s on a nVidia GeForce GTX 580 for meshes with 10,455 and 10,741 nodes, respectively. Another example can be found in Ref. [20] where an algorithm was presented for recovering suitable material parameters from MR images for the accurate modeling of breasts undergoing large deformations. The algorithm was used to estimate material parameters for up to four different types of tissue within a model: fat, fibro-glandular, muscle, and tumor tissue. The inputs were the following: a segmented image of the initial and final configurations, and a set of initial guesses for the material parameters that were obtained from the literature. Individual simulations took 19 s to complete with NiftySim, compared with 104 min with ABAQUS standard and 312 min with ABAQUS explicit on an Intel dual-core 3.4 GHz CPU with a GeForce GTX 285 GPU.

On the other hand, the Simulation Open Framework Architecture (SOFA) is another FE-based solver to work in real-time. An example of the computational needs of SOFA can be found in Ref. [21]. In this study, the authors present a prostate model that incorporates the anatomy of the male pelvic region. The model was used to simulate the prostate deformation during needle insertion and it was implemented in SOFA. SOFA simulations were compared with experimental results for two scenarios: indentation and needle insertion. An experimental phantom was developed using anatomically accurate magnetic resonance images and populated with elasticity properties obtained from ultrasound-based Acoustic Radiation Force Impulse imaging technique. Markers were placed on the phantom surface to identify the deformation during indentation experiments. Regarding indentation, the needle guide was pushed 5 mm towards the rectal wall with a velocity of 0.5 mm/s. The

indentation simulation was performed with a time-step of 0.01 s and the total simulation was computed in 80 s. The needle insertion experiment was performed with an insertion velocity of 1 mm/s. For the needle insertion simulation, a time-step of 1 s was used without compromising the model accuracy. The computation time for the total simulation was 60 s.

These two GPU-based FE solvers are very fast and accurate. However, machine learning techniques are a viable alternative to reduce even more these low computational times, as this paper will show.

Other researchers proposed model reduction techniques for dimensional reduction of FE models [22–24]. The main problem of this technique is that it is still not able to work in real-time for non-linear/complex models. For instance, it cannot model contact problems such as the compression of a breast due to the contact with the plates of the imaging devices.

Machine learning (ML) is a field that focuses on models and algorithms that are able to learn and to make predictions from data [25]. During a process called training, ML models learn a function that maps inputs and outputs without previous knowledge of the problem. The main advantage of ML models is that once the mapping function is estimated off-line, they are able to predict solutions in real-time. This situation opens a possibility to use FEM to generate data off-line that ML models can use to estimate a function that maps inputs (mechanical properties, geometry mesh, boundary conditions, ...) and outputs (nodal displacements and stresses). Once the mapping functions are estimated, ML models can make predictions for complex biomechanical behaviors in real-time.

FEM-based ML approaches have been recently proposed in the literature to build models that predict deformations of internal organs such as the liver [26,27], the stomach [28] or the prostate [29]. These approaches used the organ geometry from a unique patient and recursively simulated the deformations caused by loads with different directions and magnitudes in order to obtain the data that fed the ML model. As a result, a ML model was obtained that was specific for a unique geometry, but able to predict deformations caused by loads that were not used during training. The approach is useful for surgical planning or guidance but it implies to build a different model for each patient.

This paper presents a data-driven method to simulate, in real-time, the mechanical behavior of the breast under compression. The aim of this work is to model the biomechanical behavior of the breast tissues in some image-guided interventions such as biopsies, resections or radiotherapy dose delivery as well as to speed up multimodal registration algorithms. As far as the authors know, the only similar work was performed by our research group in an application dealing with liver deformations [30].

Ten real breast images were used in this work. Their deformation due to the displacement of two plates was performed off-line using FEM. Three machine learning models were trained with the data from those simulations. Then, they were used to predict in real-time the deformation of the breast tissues under compression. The mean error committed by the three models in the prediction of the nodal displacements was very low and the breast compression prediction took less than 0.2 s, thus proving that the three models could be suitable for clinical practice.

2. Methods

This work presents a data-driven method to simulate in real-time the breast compression due to two plates that could be from an X-ray mammogram or the plates of a MRI-guided biopsy device. The method answers the question of how well a machine learning approach can estimate the FE simulation of the breast compression. Fig. 1 shows an overview of the proposed method. First, the breast tissues were segmented from MR images. FE meshes were generated from those segmented images and the breast compression was simulated using FEM. The calculated displacements after the simulation, namely, the plate displacement, the fraction of skin, fat, and glandular tissue from each

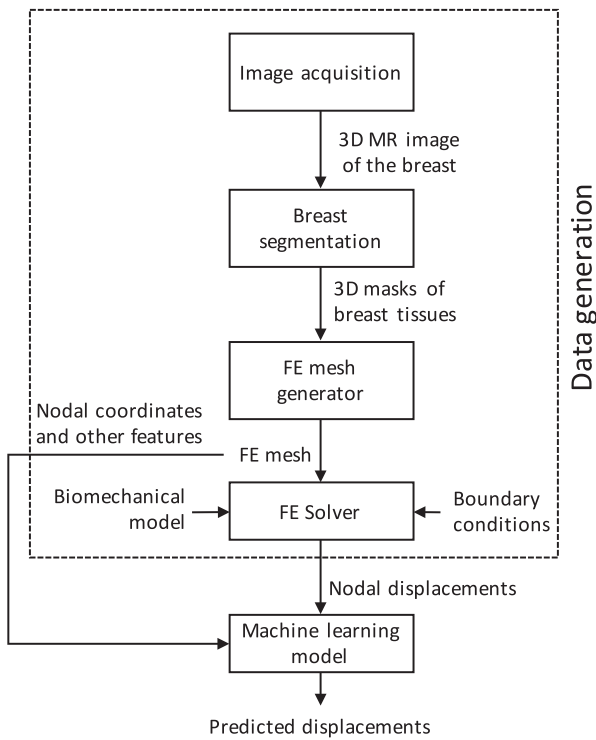


Fig. 1. Flow chart of the proposed method to simulate in real-time the breast compression in mammogram.

breast and features from the meshes (the nodal coordinates and the number of elements from skin, fat, and glandular tissue to which each node belongs) were used to train several machine learning models. During training, these models learned the relationship between the mesh features and the nodal displacements, thus being able to predict in real-time deformed states of the breast that were not used during training.

2.1. Image acquisition

MR images of ten real breast were obtained in DICOM format. These images were acquired with a *PHILIPS ACHIEVA 1.5 T* scanner. The T2 sequence was selected because of its ability to clearly display the breast structures (with less internal noise than other sequences), its number of slices, and because it does not employ any signal suppression to mask tissues. T2 was configured with TR = 5000 ms, TE = 120 ms, ip angle = 90°, slice thickness of 2 mm, image size of 512 × 512 pixels corresponding to 0.78 × 0.78 mm, and a matrix size of 448 × 512. It had 80 slices with a separation of 2 mm and 65,535 different gray levels.

2.2. Breast tissue segmentation

MR segmentation was performed following the process presented in Ref. [13]. MR segmentation begins with an image preprocessing that removes the background noise that does not belong to the breast and removes the pectoral muscle. The background noise is erased with the skin segmentation method. The pectoral muscle is detected and removed using a Hessian-based filter with eigenvalues in order to differentiate a specific geometrical structure in 3D images [31] (in this case, the pectoral muscle).

In this application, a method that uses a Curvature Flow filter, which was presented in our previous work [32], was implemented to segment the skin as accurately as possible. After the Curvature Flow filter, a cluster analysis with K-Means was performed using four clusters dividing the image into four groups: background (the darkest pixels), fatty tissue (brighter pixels), and two clusters (the brightest pixels) that mix skin and glandular tissue. Pixels that belong to these last two clusters were

Table 1

Elastic constants used for each tissue ([33,34]) for the simulation of the breast compression.

tissue	C_{01} (MPa)	C_{10} (MPa)	d (MPa ⁻¹)
skin	0.0040	0.0020	3.36
fat	0.0020	0.0013	6.04
glandular	0.0035	0.0023	3.45

reclassified with a new K-Means analysis, revealing pixels that surround the breast (the skin layer). Since some pixels of glandular tissue were still considered to be skin near the nipple region, a dynamic search was used at the breast boundary. The final pixels that were classified as skin were used as a mask to separate the breast of interest from background and to segment breast tissue with no skin [13]. At the end of the MR image segmentation process, a segmented DICOM image was obtained with the breast of interest segmented into three types of tissues: glandular tissue, fat, and real skin.

2.3. Mesh generation

The first step for generating the virtual compression of the breast is to obtain a finite element mesh from the segmented MR image. The external software, Simpleware[®] v5.1, was used to build that mesh. Before building the mesh, the segmented MR images were smoothed with an Island Removal Filter and a Gaussian Filter; the parameters for those operations were obtained experimentally (5 for Island Removal and 0.75 for the sigma value). This ensured that the FEM model could be built and guaranteed the convergence of the FE model. Simpleware provided a mesh of tetrahedrons for three types of tissues: glandular tissue, fat, and skin. The constructed meshes had a mean of 313,000 elements and 62,000 nodes with an average tetrahedron edge size of 2 mm. The compressed meshes were exported to ANSYS[®] v13.0. ANSYS[®] is a commercial finite element software package that was used to obtain the deformations produced in the process of mammogram compression.

2.4. Biomechanical model of the breast tissues

The physical behavior of the breast was described by the material model and the boundary conditions. Obtaining the true material model for each dataset from a patient is not a feasible task. In the application presented in this paper, the material model was the same for all patients and their elastic constants were constant for each type of tissue. To simulate the physical behavior of the breast tissues under compression, the hyperelastic model proposed in Ref. [33] for the glandular and fatty tissues of the breast was used in this paper. For the skin tissue, the model proposed in Ref. [34] for the human skin (which was also assumed as hyperelastic) was used. For the three tissues, the model was a Mooney-Rivlin model, whose strain energy potential, W , is defined by Equation (1) [35]:

$$W = C_{10} \left(\bar{I}_1 - 3 \right) + C_{01} \left(\bar{I}_2 - 3 \right) + \frac{1}{d} (J - 1)^2 \quad (1)$$

where I_1 , I_2 and I_3 stand for the invariants of the deformation tensor in the three dimensions of the space, C_{01} and C_{10} are material constants, J stands for the determinant of the deformation gradient, and d is the incompressibility parameter. The used elastic constants for each tissue are shown in Table 1.

2.5. FE simulations of the breast compression

ANSYS was used to simulate the compression of the breast (Fig. 2).

¹ <https://www.simpleware.com>.

² <http://www.ansys.com>.

In a cranio-caudal projection during an X-ray mammogram, the breast is compressed due to the vertical movement of a superior plate that pushes the breast against a second plate that is fixed. The breast mesh was located between two rigid bodies that emulated the two plates of an X-ray mammogram. The movement of the breast nodes that belonged to the chest wall was restricted in the anterior-posterior direction. The movement of two other sets of nodes was restricted to avoid rigid motion of the breast: the displacements of the nodes initially in contact with the superior plate were restricted in the lateral direction and the nodes closest or in touch with the inferior plate were constrained in all directions. Finally, the upper plate was displaced in the direction of the lower plate along the longitudinal axis while the lower plate was totally fixed. The problem was solved as a non-friction contact problem between a rigid body (both plates) and a deformable body (the breast).

2.6. Dataset generation

Ten healthy patients signed an informed consent form to use their clinical data. The average age of these patients was 56.8 years, ranging between 34 and 90 years old. MR images from 10 breasts were used for the experiments. The volume of skin, fat, and glandular tissue for each breast as well as the total volume is reported in Table 2. Ten corresponding meshes were obtained by the methodology explained in Section 2.3. A displacement that corresponded to the 20% of the breast height in the longitudinal axis was applied to the upper plate. ANSYS solved the problem in several load steps obtaining a simulation of the deformation at each one. These load steps were automatically chosen by ANSYS for each breast for the sake of convergence. Table 3 shows the number of nodes for the ten breast meshes and the number of steps used to simulate breast deformations. After this procedure, a total of 162 deformations were obtained from the 10 breast meshes. Each instance of the dataset corresponded to one node from the 162 deformations; finally, the dataset contained a total of 9,816,283 instances.

In this work, the data used to train the models were the results from the FE simulations of the breast compression. In particular, the data set consisted of 11 features (input variables):

- The three nodal Euclidean coordinates from each breast mesh.
- The fraction (unit basis) of skin, fat, and glandular tissue of each breast. These values were estimated from Table 2 as Equation (2) shows:

$$\hat{V}_{tissue} = \frac{V_{tissue}}{V_{skin} + V_{fat} + V_{gland}} \quad (2)$$

where V_{tissue} stands for the volume of skin, fat or glandular tissue.

Table 2

Volume of skin, fat, and glandular tissue and total volume of each breast.

Volume (cm ³)				
breast	skin	fat	glandular tissue	total
#1	26.39	152.61	38.18	217.18
#2	25.59	149.17	48.85	223.61
#3	29.11	229.16	50.57	308.84
#4	83.15	464.00	60.74	607.89
#5	29.55	507.12	69.48	606.15
#6	30.03	235.53	47.21	312.77
#7	29.07	239.45	39.50	308.02
#8	29.33	321.86	172.99	524.19
#9	78.62	839.73	44.82	963.17
#10	82.21	979.03	43.19	1104.43

Table 3

Number of nodes for the ten breast meshes and the number of steps chosen by ANSYS to simulate breast deformations.

breast mesh	number of nodes	number of simulations
#1	31,745	15
#2	46,248	19
#3	29,930	14
#4	121,210	11
#5	58,507	15
#6	55,999	24
#7	37,772	15
#8	77,507	21
#9	84,649	10
#10	80,376	18

- The normalized volume of each breast, which is estimated by means of Equation (3):

$$\hat{V}_i = \frac{V_i - \bar{V}}{\sigma} \quad (3)$$

where V_i , \bar{V} and σ stand for the total volume of breast i , the mean volume, and the standard deviation estimated over the 10 breasts respectively.

- The normalized number of skin, fat, and glandular elements from the mesh to which each node belongs. The values were estimated as in the case of Equation (2), ensuring that the sum of skin, fat, and glandular elements was equal to the unity.

- The value of the plate displacement in each step.

The prediction of the breast compression was tackled by a multi-output approach, where the nodal displacements in the three Euclidean coordinates d_1, d_2, d_3 were considered as the target variables. Fig. 3 shows the histogram of the norm of the nodal displacements described in

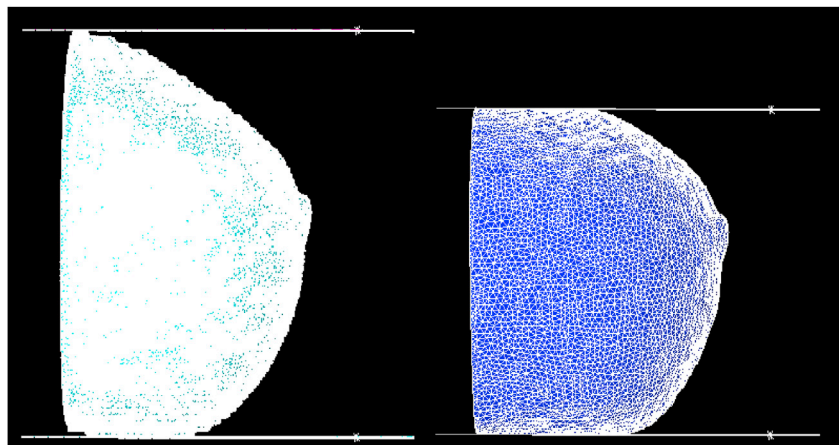


Fig. 2. Sagittal view of the breast before (left) and after (right) the FE simulation of the compression.

Equation (4) for the whole dataset and the deviation estimated over the 162 deformations.

$$\|d\| = \sqrt{\sum_{i=1}^3 (d_i)^2} \quad (4)$$

2.7. Real-time predictions of the breast compression

Several machine learning models from *Scikit Learn 0.17* [36], a package for *Python 3.5*, were used to predict the breast compression in real-time. A preliminary study with a set of regression models was performed. A dummy model (used as a baseline) that always returns the mean value, a linear regression model, artificial neural networks, a decision tree model, and two tree-based ensemble methods (extremely randomized trees and random forest) were used. The conclusion of the study was that tree-based ensemble methods provided the best results and this is why the rest of the experimentation was focused on them. For this reason, the following three models were used to predict the deformations: decision tree (DT) [37], random forest (RF) [38] and extremely randomized trees (ERTs) [39].

DT is a supervised learning algorithm that predicts values of a target variable by learning simple decision rules from the data. RF and ERTs are tree-based ensemble methods, i.e., they combine several trees to improve results over a single estimator; they are considered as state-of-the-art methods in machine learning [40]. In particular, these two methods create a set of estimators and add randomness to them, giving as final predictions the averaged predictions of each estimator. The main difference between both is the way they select the best split during the construction of the tree. RF selects the best split among a random subset of the features using the most discriminative threshold. ERTs also selects the best split from a random subset of features but using the best threshold from a randomly generated set of thresholds for each feature in the subset.

The hyperparameters used to train DT, RF and ERTs are the following:

- All the features were used for each decision split.
- The nodes of the tree/s were expanded until all leaves contained less than two samples.
- At least two samples were required to split an internal node.
- Mean Squared Error was used as the function to measure the quality of the splits.
- A total of 20 trees were used to build RF and ERTs.

3. Experiments

The dataset was split into training and validation sets. The training set was used by the models to learn the relationship between the input and output data spaces. The validation set was used to assess the models' performance on new or unseen data. This strategy guarantees the correct generalization of the model avoiding overfitting (a typical problem that appears when a model provides a perfect score in samples from the training set but fails predicting unseen data).

In this work, two experiments based on different partition approaches were carried out to evaluate the performance of the three models: leave-one-deformation-out and hold-out validation. The first experiment is a geometry-based partition, while the second one is not. A geometry-based partition means that an entire geometry is taken for training or validation but never split. In contrast, hold-out validation is an instance-based partition, i.e., the geometry is not considered during partitions so it may happen that some nodes of a given geometry may belong to the training set with others belonging to the validation set.

3.1. Hold-out validation

In this experiment the dataset was randomly split into two partitions:

70% of the data was used to train the models and the remaining 30% to validate them. As previously mentioned, the geometry of the breast was not considered for data splitting. As one of the main limitations of working with medical data is the amount of available data, this experiment gives insight about how the model would scale and behave with an extensive breast dataset when we leave a whole breast geometry for validation. This approximation is a standard approach in machine learning framework and it was designed aimed to assess the performance of the models in a real scenario with many breasts available for modeling.

3.2. Leave-one-deformation-out validation

A unique deformation from the 162 simulations performed in ANSYS was used for validation, and the remaining deformations were used for training. This experiment was performed for the 162 deformations. This experiment is closer to the clinical practice, where the models are trained with a set of breasts and deformations and then used to predict the behavior of a new state of deformation for a whole breast geometry. However, this is a challenging experiment due to the number of deformed states (162 in our particular case). In this experiment, the models were validated with one deformation from a breast mesh that was used during training.

3.3. Error estimation

A metric must be considered to evaluate the performance of the models. In this work, the error at each node was calculated as the 3D Euclidean distance between the predicted displacement by the models (d^*) and the observed displacement after the FE simulation of the breast deformation (d), as shown in Equation (5):

$$\varepsilon(d, d^*) = \sqrt{\sum_{i=1}^3 (d_i - d_i^*)^2} \quad (5)$$

The mean of the error over all the nodes from the validation set, Equation (6), as well as the percentage of nodes with error ε greater than 1 mm, 2 mm and 3 mm were calculated to present the results. These thresholds were considered to assess the feasibility of applying the models in real clinical practice. In particular, 3 mm is an acceptable threshold for the committed error, 2 mm a good threshold and 1 mm corresponds with an extremely accurate prediction.

$$\bar{\varepsilon}(d, d^*) = \frac{1}{N} \sum_{n=1}^N \varepsilon_n(d, d^*) \quad (6)$$

where ε_n stands for the error ε of the n -th node, and N is the total number of nodes.

4. Results

It must be emphasized that all results presented in this paper are referred to the validation data set, this giving an indication of the performance of the model on unseen data different from those used for training.

4.1. Results of hold-out experiment

Table 4 shows the results obtained by the three models in the hold-out validation experiment. The lowest errors correspond with ERTs: there were no nodes with errors greater than 2 mm and 3 mm for this model and only 0.03% of the nodes had an error higher than 1 mm. The mean nodal error was equal to 0.046 mm; it corresponds to the 0.066% of the maximum nodal displacement from the whole dataset (69.69 mm), to the 0.4% of the mean displacement (11.58 mm) and to the 0.43% of the standard deviation (10.62 mm).

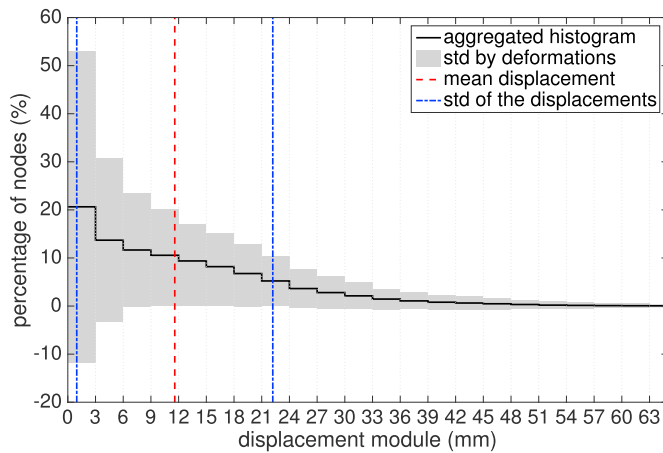


Fig. 3. Histogram of the norm of the displacement $\|d\|$ and the deviation over the deformations. The histogram is normalized, i.e., the accumulated sum is equal to 100%. The mean displacement \pm standard deviation is superimposed in dashed lines (11.58 ± 10.62 mm).

Fig. 4 shows the histograms of the error ε committed by the three models in the hold-out validation experiment. In order to compute the histograms, ε was uniformly quantified in the interval $[0, 1]$ mm with a bin width equal to 0.1 mm (a total of 10 bins). Each histogram was normalized by dividing each height bin by the number of nodes in the validation set, i.e., the cumulative sum of each histogram was equal to 1. Finally, the histograms were multiplied by 100 to provide percentages. The histograms show that there are practically no nodes with errors greater than 0.5 mm for any of the three models. They also show that ERTs yield a better performance than DT and RF. Almost 90% of the nodes from the validation set show an error smaller than 0.1 mm for ERTs.

Fig. 5 compares the norm of the observed displacements from the validation set $\|d\|$ described in Equation (4) with the norm of predicted displacements d^* . The nodes were sorted by the displacement in ascending order for the sake of clarity.

Fig. 6 (left) shows the error ε defined in Equation (5) committed by the three models using the same sorting displacement strategy as in Fig. 5. Since the error may be smaller for small deformations and larger for large deformations, the relative errors $\hat{\varepsilon}$ were also estimated, as shown in Equation (7):

$$\hat{\varepsilon}(d, d^*) = \frac{\varepsilon(d, d^*)}{\|d\|} \quad (7)$$

The relative errors $\hat{\varepsilon}$ take values between 0 and 1 while $\varepsilon \leq \|d\|$. Fig. 6 (right) shows the relative errors and it can be appreciated that their values were larger for small displacements, while the models performed better for large displacements.

4.2. Results of leave-one-deformation-out experiment

This experiment was carried out 162 times, one for each deformation. Table 5 shows the errors committed by the three models. For the sake of simplicity, the table shows the errors averaged over the 162

Table 4

Percentage of nodes with errors larger than 1 mm, 2 mm and 3 mm committed by each model in the hold-out validation. The error at each node is calculated by means of Equation (5). The mean error ($\bar{\varepsilon}$) is estimated over all the nodes from the validation set.

model	% of nodes with $\varepsilon > 1$ mm	% of nodes with $\varepsilon > 2$ mm	% of nodes with $\varepsilon > 3$ mm	$\bar{\varepsilon}$ (mm)
DT	0.41	0.03	0.01	0.123 ± 0.167
ERTs	0.03	0	0	0.046 ± 0.067
RF	0.08	0.01	0	0.064 ± 0.094

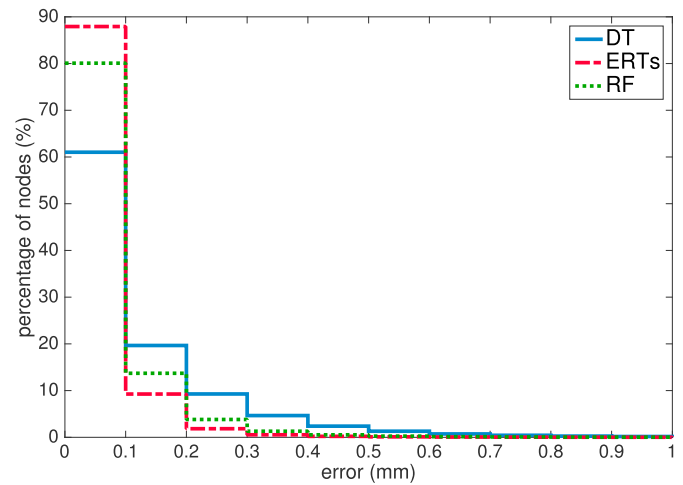


Fig. 4. Distribution of the errors committed by Decision Tree (DT), Extremely Randomized Trees (ERTs) and Random Forest (RF) in the hold-out validation.

deformations. On average, there was only a 2.72% of nodes with an error greater than 3 mm for the best model, ERTs, that obtained a mean nodal error averaged over the ten deformations equal to 0.624 mm. This value corresponded to the 0.89% of the maximum plate displacement, to the 5.39% of the mean displacement and the 5.88% of the standard deviation.

Fig. 7 shows the histograms of the error ε committed by the three models in the leave-one-deformation-out validation experiment. The histograms were computed as explained in Section 4.1 for the hold-out validation experiment. The only difference is that here ε was uniformly quantified in the interval $[0, 10]$ instead of $[0, 1]$. The histograms show that there were practically no nodes with errors greater than 5 mm for any of the three models. Around 80% of the nodes from the validation set committed an error smaller than 1 mm for ERTs.

DT and RF predicted worse than ERTs. The fact that DT performed worse than ERTs was expected because ERTs is an ensemble method, i.e., it is more complex since combines a set of simple DTs. The difference between the performance of the two ensemble methods (ERTs and RF) was not that expected although there are several reasons that could explain why ERTs performed better than RF in this particular case. The first reason is that during training, ERTs built a more complex ensemble. The mean depth leaf, estimated as the average of the depth leaves over all the trees in the ensemble and the 162 trained models (one for each validation deformation), was $\sim 12\%$ larger for ERTs. The mean number of nodes was also larger for ERTs, specifically $\sim 37\%$ larger. This also occurred during training in the hold-out experiment, but the error differences between models were less pronounced, since as explained in Section 3.2 leave-one-deformation-out experiment was more difficult to be modeled due to the limited amount of deformed states (162 in our case). The second reason is the extra-randomness added by ERTs selecting the best split thresholds.

4.3. Summary of results

It is concluded that the best models' performance was achieved in hold-out validation, as expected. Although this is the typical ML approach, is farther away from the clinical practice than leave-one-deformation validation experiment given the nature of our problem, where each instance within the dataset (nodal features) is part of something bigger (an organ geometry). Focusing on leave-one-deformation-out experiment, which is a geometry-based partition method and closer to the clinical practice, the results show that the performance is worse, but still acceptable. This experiment is highly recommended since it is very important validating with unseen plate

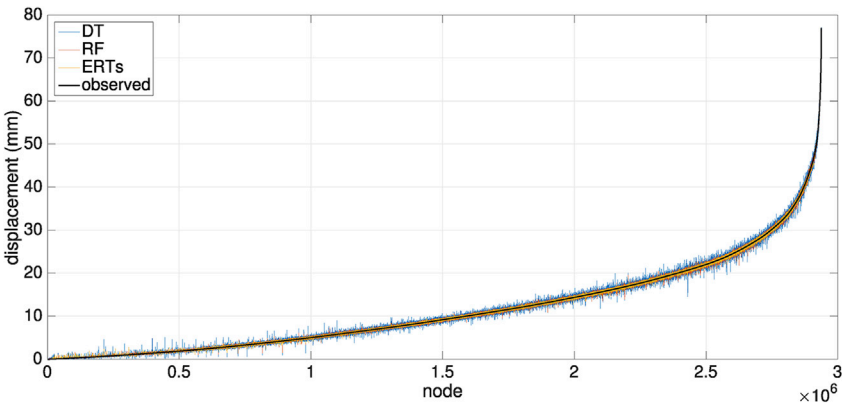


Fig. 5. Sorted norm of the displacements observed after the FE simulations of the breasts from the hold-out validation and the norm of predicted displacements for the three models.

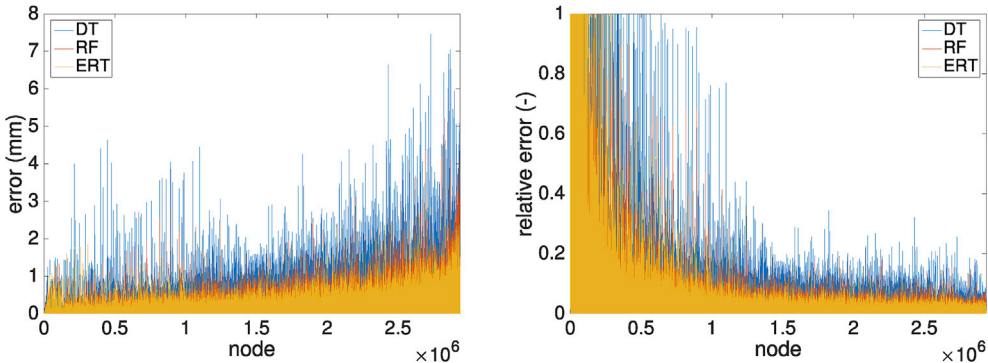


Fig. 6. Error ϵ (left) and relative error $\hat{\epsilon}$ error (right) calculated through Equations (5) and (7) committed by the three models during the hold-out validation.

displacement (something that will also occur in clinical practice). It gives a lot of information about the real performance of the models since forces the model to interpolate and produce outcomes for a truly unseen scenario. In this experiment ERT outperformed DT and RF drastically, something that was not visible in the hold-out experiment.

The time that took the prediction of the breast compression was measured using an equipment with 2.60 GHz Intel (R) Xeon (R) CPU and it was computed as the sum of CPU time, I/O time and the communication channel delay. On average, the prediction of a whole deformed breast took 0.02 s, 0.10 s and 0.15 s with DT, ERTs and RF, respectively.

Pogorelov et al. [41] defined the term ‘near to real-time’ if a system is able to process 10 frames per second, which corresponds to a computation time of 0.10 s. According to this, our best model (ERTs) provides the same value, while DT is below and RF above. In our case, we consider real-time if our system allows the decision making in a real clinical environment ($t < 1$ s).

On the other hand, the time that took to train a model was around 127 s on average. This computation time is still considerable smaller than FEM-based approaches, which can take about 2 h for breast compression [11,13].

Table 5
Percentage of nodes with errors larger than 1 mm, 2 mm and 3 mm committed by each model in the leave-one-deformation-out validation. The error at each node is calculated by means of Equation (5). The results are shown averaging over the 162 deformations.

model	% of nodes with $\epsilon > 1$ mm	% of nodes with $\epsilon > 2$ mm	% of nodes with $\epsilon > 3$ mm	$\bar{\epsilon}(\text{mm})$
DT	45.85	20.96	11.37	1.59 ± 2.755
ERTs	20.64	6.43	2.72	0.624 ± 0.881
RF	45.58	20.66	10.90	1.525 ± 2.647

5. Discussion

It has been demonstrated that it is crucial to test the model for loads (plate displacement for breast compression simulation) not used during training, since the applied loads are highly correlated with the obtained deformations.

The presented study has shown the feasibility to use machine learning models to predict in real-time the biomechanical behavior of the breast, concretely the deformation that the breast experiences under compression, due to two reasons: the mean error committed by the method was

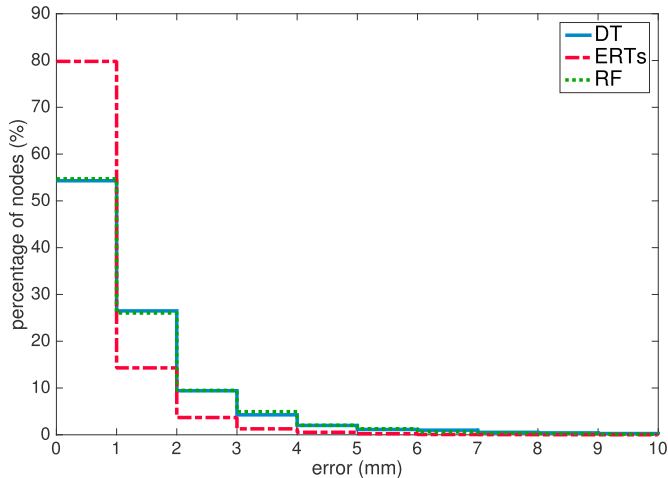


Fig. 7. Distribution of the errors committed by Decision Tree (DT), Extremely Randomized Trees (ERTs) and Random Forest (RF) in the leave-one-deformation-out validation.

under 0.2 mm and 2 mm for the hold-out and leave-one-deformation-out experiments, respectively, and the breast compression prediction took 0.1 s on average with ERTs (the most complex of the three models). Besides, there were practically no nodes in the whole validation set with errors greater than 2 mm and 5 mm for hold-out and leave-one-deformation-out, respectively. The best results were achieved in the hold-out validation experiment with ERTs, where the average error was equal to 0.046 mm, and there were no nodes with errors greater than 2 mm.

The compression of the real breasts was limited to 20% in order to avoid convergence problems of the Finite Element (FE) solution. In fact, we found convergence problems for some breast meshes when higher compression ratios were applied. The problems were related to distorted elements and these might easily appear when the skin layer is considered. This is mainly due to the small size of the skin; as a result of this, the elements are very small and few elements define the layer. Therefore, under large deformations some of these elements are distorted and then, the FE solution does not converge. In order to obtain a deformed state for each one of the real breasts, we decided to limit this deformation to 20%. This does not mean that non-linearity is not taken into account, since the effects of non-linearity appear at around 5% of the deformation for the soft tissues [42].

It should be emphasized that this is not a limitation of the ML models that can reproduce this solution in real time with a very low error, but rather a problem of the FE model used to virtually simulate the breast compression. It is true that simpler models that allow compression rates above 20% can be chosen (without the skin) to make the registration between MRI and X-ray [7,11–13]. However, they do not provide clinically acceptable errors (between 11 mm and 15 mm in the localization of lesions suspicious to be a tumor).

Our aim was to obtain a specific model for the mammogram compression but generic for the geometry. It was not designed to generalize for a different deformation but for new patients. Our approach has been based on using several geometries during training in order to generalize for new patients. Once the model is constructed, there is no need to re-train the model for a new patient.

The most realistic experimental setup should consist of testing the models in completely new patients, not previously seen by the model. The only reason that prevented us from doing it is the limited number of breasts (only 10) that form our data set; a leave-one-patient-out validation would involve to test the model obtained in the previous nine patients with the remaining one, that may have very different characteristics, and thus, the model may not have enough information to provide a correct prediction. Obviously, if the number of breasts was higher (as we expect in the near future) we would evaluate the model that way. Our point is that the experimentation shown in this paper is a fair estimation of the accuracy of the models leaving one patient out in a scenario with many breasts.

Biomechanical modeling of the breast tissues has been used for a great number of registration methods, namely: registration of prone MRI to PET, PET/CT, registration of prone MRI to supine MRI, registration of prone MRI to supine surgical surface/depth data, MRI to X-ray registration, and registration of ultrasound CT and X-ray mammography [2]. However, despite these developments, the accuracy of biomechanical modeling as well as the high computational costs of the algorithms have limited its applicability in clinical practice. Obtaining accurate material models is an ongoing research topic in this field; in particular, more accurate patient-specific constants for the model as well as more adequate boundary conditions are needed. The high computational cost involved in obtaining numerical solutions for those models can be approached by the ML methods presented in this work due to their great modeling capabilities.

The realism of the FE simulations depends on many factors such as, the process of segmentation, the constitutive model assumed to characterize the mechanical behavior of the tissues, the boundary conditions chosen to solve the problem, etc [2]. It is true that these simulations were

used as ground truth in this work. However, we used the same database used in Ref. [13] for the registration between the X-ray mammogram and MRI, which provided an error under 5 mm in this process. The simulations of the mammogram compression lasted about 2 h in this application. Therefore, the high reduction of the computational time is a remarkable contribution of this work. The main goal of the approach presented in this paper is to provide a methodology able to accelerate FE simulations of the mammogram compression for applications like this, thus improving its usability in clinical practice.

The orientation or the pose of the patient has not been considered as an input parameter of our model since we are accelerating the FE simulation of the breast compression during a mammogram or a MRI-guide biopsy. It is true that new machine learning models could be constructed simulating the deformations that the breast undergoes for different load states. This is an important challenge for our future work.

6. Conclusions

This paper has described a method to estimate in real-time the deformation undergone by the breast under compression. The finite element method was used to perform off-line simulations of the breast deformation caused by two compression plates. Three machine learning models were used to learn the relationship between inputs (geometry, breast tissues, and loads) and outputs (nodal displacements) from the simulations. A study was performed by means of two experiments to assess the performance of the three models, concluding that extremely randomized trees provided the best performance. Nevertheless, the other ML approaches used in this work also provided a mean error under 2 mm, which is an acceptable performance for clinical practice.

Conflicts of interest statement

The authors of the manuscript entitled “A finite element-based machine learning approach for modeling the mechanical behavior of the breast tissues under compression in real time” certify that they have NO financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Acknowledgments

This work has been funded by the Spanish Ministry of Economy and Competitiveness (MINECO) through research projects TIN2014-52033-R and DPI2013-40859-R with the support of European FEDER funds.

References

- [1] A. C. Society, Breast cancer facts & figures, Am. Cancer Soc. INC. 1 (34) (2011–2012).
- [2] J.H. Hipwell, V. Vavourakis, L. Han, T. Mertzaniidou, B. Eiben, D.J. Hawkes, A review of biomechanically informed breast image registration, *Phys. Med. Biol.* 61 (2) (2016) R1.
- [3] W.A. Kaiser, H. Fischer, J. Vagner, M. Selig, Robotic system for biopsy and therapy of breast lesions in a high-field whole-body magnetic resonance tomography unit, *Investig. Radiol.* 35 (8) (2000) 513–519.
- [4] K. Drukker, K. Horsch, M.L. Giger, Multimodality computerized diagnosis of breast lesions using mammography and sonography 1, *Acad. Radiol.* 12 (8) (2005) 970–979.
- [5] F.B. De Abreu, W.A. Wells, G.J. Tsongalis, The emerging role of the molecular diagnostics laboratory in breast cancer personalized medicine, *Am. J. Pathol.* 183 (4) (2013) 1075–1083.
- [6] O.C. Zienkiewicz, R.L. Taylor, *The Finite Element Method. Volume 1., Basic Formulation and Linear Problems*, Mc Graw-hill, London, New York, Paris, 1989.
- [7] N. Ruiter, R. Stotzka, T.O. Muller, H. Gemmeke, J.R. Reichenbach, W.A. Kaiser, Model-based registration of X-ray mammograms and MR images of the female breast, in: *IEEE Transactions on Nuclear Science*, vol. 53, 2006, pp. 204–211.
- [8] J.H. Chung, V. Rajagopal, P.M.F. Nielsen, M.P. Nash, Modelling mammographic compression of the breast, in: *medical image computing and computer-assisted intervention – MICCAI*, in: 11th International Conference, New York, NY, USA, Proceedings, Part II, 2008, 2008, pp. 758–765.
- [9] P. Pathmanathan, D.J. Gavaghan, J.P. Whiteley, S.J. Chapman, J.M. Brady, Predicting tumor location by modeling the deformation of the breast, *IEEE Trans. Biomed. Eng.* 55 (10) (2008) 2471–2480.

- [10] C. Tanner, M. White, S. Guarino, M.A. Hall-Craggs, M. Douek, D.J. Hawkes, Large breast compressions: observations and evaluation of simulations, *Med. Phys.* 38 (2) (2011) 682–690.
- [11] T. Hopp, M. Dietzel, P. Blatzer, P. Kreisel, W.A. Kaiser, H. Gemmeke, N.V. Ruiters, Automatic multimodal breast image registration using biomechanical FEM models and intensity-based optimization, *Med. Image Anal.* 17 (2) (2013) 209–218.
- [12] T. Mertzaniidou, J. Hipwell, S. Johnsen, L. Han, B. Eiben, Z. Taylor, S. Ourselin, H. Huisman, R. Mann, U. Bick, et al., MRI to X-ray mammography intensity-based registration with simultaneous optimisation of pose and biomechanical transformation parameters, *Med. Image Anal.* 18 (4) (2014) 674–683.
- [13] J.A. Solves-Llorens, M.J. Rupérez, C. Monserrat, E. Feliu, M. García, M. Lloret, A complete software application for automatic registration of x-ray mammography and magnetic resonance images, *Med. Phys.* 41 (2014) 081903.
- [14] S. Cotin, H. Delingette, N. Ayache, A hybrid elastic model for real-time cutting, deformations, and force feedback for surgery training and simulation, *Vis. Comput.* 16 (2000) 437–452.
- [15] D.D. Pokrajac, A.A. Maidment, P. Bakic, Optimized generation of high resolution breast anthropomorphic software phantoms, *Med. Phys.* 39 (4) (2012) 2290–2302.
- [16] S. Capell, S. Green, B. Curless, T. Duchamp, Z. Popovic, Interactive skeleton-driven dynamic deformations, in: *ACM SIGGRAPH 2002 Papers, SIGGRAPH '02*, ACM, 2002.
- [17] Y. Inoue, Y. Masutani, K. Ishii, N. Kumai, F. Kimura, I. Sakuma, Development of surgical simulator with high-quality visualization based on finite-element method and deformable volume rendering, *Syst. Comput. Jpn.* 37 (2006) 67–76.
- [18] S.F. Johnsen, Z.A. Taylor, M.J. Clarkson, J. Hipwell, M. Modat, B. Eiben, L. Han, Y. Hu, T. Mertzaniidou, D.J. Hawkes, S. Ourselin, Niftysim: a GPU-based nonlinear finite element package for simulation of soft tissue biomechanics, *Int. J. Comput. Assist. Radiol. Surg.* 10 (7) (2015) 1077–1095.
- [19] B. Eiben, L. Han, J.H. Hipwell, T. Mertzaniidou, S. Kabus, T. Bülow, C. Lorenz, G. Newstead, H. Abe, M. Keshtgar, S. Ourselin, D.J. Hawkes, Biomechanically guided prone-to-supine image registration of breast MRI using an estimated reference state, in: *10th IEEE International Symposium on Biomedical Imaging: from Nano to Macro, ISBI 2013*, 7–11 April, 2013, San Francisco, CA, USA, Proceedings, 2013, pp. 214–217.
- [20] L. Han, J.H. Hipwell, C. Tanner, Z. Taylor, T. Mertzaniidou, J. Cardoso, S. Ourselin, D.J. Hawkes, Development of patient-specific biomechanical models for predicting large breast deformation, *Phys. Med. Biol.* 57 (2) (2012) 455.
- [21] P. Moreira, I. Peterlik, H. Mark, C. Duriez, S. Cotin, S. Misra, Modelling prostate deformation: SOFA versus experiments, *Mech. Eng. Res.* 3 (2) (2013) p64.
- [22] S. Niroomandi, I. Alfaro, E. Cueto, F. Chinesta, Real-time deformable models of non-linear tissues by model reduction techniques, *Comput. Meth. Programs Biomed.* 91 (3) (2008) 223–231.
- [23] S. Niroomandi, I. Alfaro, E. Cueto, F. Chinesta, Accounting for large deformations in real-time simulations of soft tissues based on reduced-order models, *Comput. Meth. Programs Biomed.* 105 (1) (2012) 1–12.
- [24] S. Niroomandi, D. González, I. Alfaro, F. Bordeu, A. Leygue, E. Cueto, F. Chinesta, Real-time simulation of biological soft tissues: a PGD approach, *Int. J. Numer. Meth. Biomed. Eng.* 29 (5) (2013) 586–600.
- [25] C.M. Bishop, *Pattern Recognition and Machine Learning* (Information Science and Statistics), Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
- [26] K. Morooka, X. Chen, R. Kurazume, S. Uchida, K. Hara, Y. Iwashita, M. Hashizume, Real-time nonlinear FEM with neural network for simulating soft organ model deformation, in: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008: 11th International Conference*, New York, NY, USA, September 6–10, 2008, Proceedings, Part II, 2008, pp. 742–749.
- [27] W. Abdelrahman, S. Farag, S. Nahavandi, D.C. Creighton, A comparative study of supervised learning techniques for data-driven haptic simulation, in: *SMC, IEEE*, 2011, pp. 2842–2846.
- [28] D. Deo, S. De, Phyness: a physics-driven neural networks-based surgery simulation system with force feedback, in: *EuroHaptics Conference, 2009 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2009. Third Joint*, 2009, pp. 30–34.
- [29] S.R. Mousavi, I. Khalaji, A.S. Naini, K. Raahemifar, A. Samani, Statistical finite element method for real-time tissue mechanics analysis, *Comput. Meth. Biomech. Biomed. Eng.* 15 (6) (2012) 595–608.
- [30] D. Lorente, F. Martínez-Martínez, M. Rupérez, M.L.M. Martínez-Sober, P. Escandell-Montero, J. Martínez-Martínez, S. Martínez-Sanchis, A. Serrano-López, C. Monserrat, J. Martín-Guerrero, A framework for modelling the biomechanical behaviour of the human liver during breathing in real time using machine learning, *Expert Syst. Appl.* 71 (2017) 342–357.
- [31] L. Wang, K. Filippatos, O. Friman, H.K. Hahn, Fully automated segmentation of the pectoralis muscle boundary in breast MR images 7963 (2011) 796309–796309–8.
- [32] J.A. Solves-Llorens, M.J. Rupérez, C. Monserrat, E. Feliu, M. García, M. Lloret, Segmentation of the breast skin and its influence in the simulation of the breast compression during an X-ray mammography, *Sci. World J.* 2012 (2012) 876489.
- [33] H.M. Yin, L.Z. Sun, G. Wang, T. Yamada, J. Wang, M.W. Vannier, Imageparser: a tool for finite element generation from three-dimensional medical images, *Biomed. Eng. OnLine* 3 (1) (2004) 31.
- [34] C.C.M.M. Niels, I. Horváth, Using finite element model of the human body for shape optimization of seats: optimization material properties, in: *International Design Conference - Design, Design*, 2002.
- [35] W.K. Belytschko, B. Moran, *Nonlinear Finite Elements for Continua and Structures*, Wiley, New York, 2000.
- [36] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: machine learning in Python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830.
- [37] L. Breiman, J. Friedman, R. Olshen, C. Stone, *Classification and Regression Trees*, Wadsworth and Brooks, Monterey, CA, 1984.
- [38] L. Breiman, Random forests, *Mach. Learn.* 45 (1) (2001) 5–32.
- [39] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, *Mach. Learn.* 63 (1) (2006) 3–42.
- [40] M. Fernández-Delgado, E. Cernadas, S. Barro, D. Amorim, Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res.* 15 (2014) 3133–3181.
- [41] K. Pogorelov, M. Riegler, P. Halvorsen, P.T. Schmidt, C. Griwodz, D. Johansen, S.L. Eskeland, T.D. Lange, GPU-accelerated real-time gastrointestinal diseases detection, in: *2016 IEEE 29th International Symposium on Computer-based Medical Systems (CBMS)*, 2016, pp. 185–190.
- [42] F. Martínez-Martínez, M. Lago, M. Rupérez, C. Monserrat, Analysis of several biomechanical models for the simulation of lamb liver behaviour using similarity coefficients from medical image, *Comput. Meth. Biomech. Biomed. Eng.* 16 (7) (2013) 747–757, <https://doi.org/10.1080/10255842.2011.637492> PMID: 22463393.