Fast Segmentation of Convex Cyst-Like Structures in Gelatin Soft Tissue Phantoms under Ultrasound Imaging with Artifacts and Limited Training Samples

Dalia Drozd[1], Adam Ciszkiewicz[1*]

[1] Faculty of Mechanical Engineering, Cracow University of Technology, ul. Warszawska 24, 31-155 Cracow, Poland
* Corresponding author's e-mail: adam.ciszkiewicz@pk.edu.pl

ABSTRACT

Ultrasound imaging is commonly used in surgical training and development of medical robotics systems. Recent advancements in surgical training often utilize soft-tissue phantoms based on gelatin, with additional objects inserted to represent different, typically fluid-based pathologies. Segmenting these objects from the images is an important step in the development of training and robotic systems. The current study proposed a simple and fast algorithm for segmenting convex cyst-like structures from phantoms under very low training sample scenarios. The algorithm is based on a custom two-step thresholding procedure with additional post-processing with two trainable parameters. Two large phantoms with convex cysts are created and used to train the algorithm as well as evaluate its performance. The train/test procedures were repeated 60 times with different dataset splits and proved the viability of the solution with only 4 training images. The DICE coefficients were on average at 0.92, while in the best cases exceeded 0.95, all with fast performance in single-thread operation. The algorithm might be useful for development of surgical training systems and medical robotic systems in general.

Keywords: thresholding, medical imaging, computer vision.

INTRODUCTION

Ultrasound (US) imaging is a common method for visualization in clinical practice. The image is created by probing the tissue with a series of acoustic signals in the US range. Then, these signals are captured as they reflect, with some energy loss, from different tissues. The final image is composed using a numerical algorithm. This approach is relatively inexpensive, which makes the equipment widely available. While the image is generally of lower quality than what can be obtained with other methods, such as magnetic resonance imaging [1, 2], it allows for diagnosis as well as increases the safety of the patient and the quality of the medical procedure.

Increasing interest can be observed in the field of producing soft tissue phantoms for use in medical training [3–6]. The use of phantoms is also very common in development of medical robotic

systems, for instance in breast biopsy [7, 8]. Depending on their purpose, the phantoms differ in structure. In terms of low-cost, easy to prepare phantoms, a popular option is to utilize actual animal soft tissues [9]. Another common solution is to use gelatin [10, 11], often with additional ingredients, such as corn flour, agar and more [12]. Gelatin based phantoms can also mimic inflamed tissue [13] This is both inexpensive and sustainable, in addition to offering good results in terms of visualization under US. To mimic the different layers in human tissue, different elements can be added [14, 15] usually based on common ingredients, such as tofu for skin modeling and nitrile gloves for cyst-like structures.

While US imaging is very common and affordable, the quality of the obtained images can vary, as the method is susceptible to various artifacts [16, 17]. US devices offer a large number of parameters available to the user, which can be

tuned to increase the quality of the image. Nevertheless, setting these parameters requires experience and can be difficult. At the same time, the parameters can greatly affect the signal-to-noise ratio in the image and the ability to discern certain objects in it. The artifacts are difficult to remove and soft tissue phantoms are especially susceptible to them, due to lower mechanical endurance of the top layer, when compared to actual human tissue. This makes the images difficult to interpret, while the interpretation is mostly subjective. This problem is more exacerbated by the fact that the phantoms are often used for medical training and development of medical robotics in biopsy with US. Therefore, automatic segmentation of pathological tissue in tissue phantoms has been an active area of study [18, 19]. Nevertheless, due to the large variety of phantoms, US artifacts and objects representing pathologies, the problem still remains open. This is only compounded by the fact that novel and advanced techniques in machine learning typically require large datasets [20, 21], while the results highly depend on the chosen architecture and the training process [22]. While the performance of deep learning methods is generally higher than that of classical methods [23], the need for large input sets can be a limiting factor in some applications.

Therefore, the aim of this study was to propose a simple and fast algorithm for segmenting cyst-like structures in gelatin-based soft tissue phantoms under US imaging with limited number of training samples. For this purpose, two large soft tissue phantoms with convex cyst-like structures were prepared. These phantoms were then registered using US under varied parameters. The obtained images were manually segmented and served as the ground-truth for the analysis. In the next step a computer-vision algorithm was developed and trained on 20% of the images and then tested on the remaining ones. This procedure was repeated 60 times in order to minimize the effects of random number generation. The DICE coefficient was chosen to measure the quality of results. The following sections describe the approach in detail.

MATERIALS AND METHODS

As it was mentioned in the previous section, the aim of this study was to develop a simple and fast algorithm for segmenting cyst-like structures in gelatin-based soft tissue phantoms under US imaging. However, this involved preparation of the phantoms and the database for segmentation. All of these steps were described in detail in the following sections.

Preparing soft tissue phantoms

The first step of this research was to prepare the soft tissue phantoms. The base recipe was assumed after [10] and consisted of 50 g of gelatin with 160 g of corn flour and 1000 ml of warm water. In the first step, the corn flour was mixed with the water and heated on an induction stove for 2 minutes. After that, the gelatin was added and stirred into the mix. This step was performed with care to prevent the mixture from clamping up. Then, the solution was further heated and stirred for another 10 minutes to thicken it up. The resulting mixture was poured into a bread-form container with a rectangular shape and additional elements were implanted into it. These elements represented the cysts and the different layers of the skin. Specifically, cut nitrile gloves were filled with water and tied to substitute fluid-based pathologies, while the skin layer was represented with a rectangular tofu slice, which was 1 to 2 cm wide. After cooling down, the model was stored in the fridge for approximately 12 hours, before being evaluated with US. In total, two models were prepared with this recipe – see Figure 1 for one of the samples. The models were relatively large and contained multiple elements, which allowed for registration of a varied set of images.

Registering soft tissue phantoms under ultrasound

MEDISON SonoAce PICO US was equipped with a linear probe L5-9EC, frequency of 9MHz, and used to register the phantoms, see Figure 2 and Figure 3. In order to introduce as many typical conditions into dataset as possible, images were registered under slightly different parameters of the US. Specifically, for each phantom, US was manually set so that the cyst-like objects were visible. Then, the objects were registered with varying TCG curve, gain, focus and probe alignment to increase the robustness of the obtained results. The images were then saved into lossless bmp files with a resolution of 640 × 480, a sample was presented in Figure 4a. In total, 19 varied images were obtained with this method.

Figure 1. One of the obtained soft tissue phantoms

Obtaining the database for a computer vision algorithm

After the registration, each image was manually segmented in GIMP. The cyst-like object was marked in green, as seen in Figure 4b. This resulted in a database of 19 input images and corresponding 19 output images. This set was then shuffled and split into two parts. The first part – training – was used to prepare the algorithm for segmenting the cysts, while the second part was needed for the validation of the procedure on unseen cases – more details were given in the section 2.5.

An algorithm for automatic segmentation of cysts in the soft tissue phantoms

The image processing algorithm consisted of several steps. In the first step, the parts of the image corresponding to the graphical user interface of the US device were cropped, leaving only the region of interest (see Fig. 4a). Then, custom thresholding method was applied onto the image and followed with labelling, simple classification and convex hull. The thresholding method featured two stages, similarly to [24]. In the first step, the threshold was set to the level thr_1, then the second threshold was computed using the thresholded image and scaled with the second parameter thr_2. These two parameters could be tuned to the problem and this was the approach undertaken in this study – more details were given in the section 2.5. The final equation used for obtaining the thresholded image was as follows:

$$I_{thr} = be(I > (mfn[I \cdot (I > thr_1)] \cdot thr_2)) \quad (1)$$

where: I_{thr} – the image obtained after thresholding, $be()$ – a function, which performs an erosion operation on a binary image (here: the version from Scikit-image [25] was used), I – the input image, $mfn()$ – function, which computes the mean value of a matrix using only the nonzero elements thr_1 – the first trainable parameter, denoting the initial threshold value, thr_2 – the second trainable parameter, denoting the strength of the second threshold.

Figure 2. The US device MEDISON SonoAce PICO used in this study

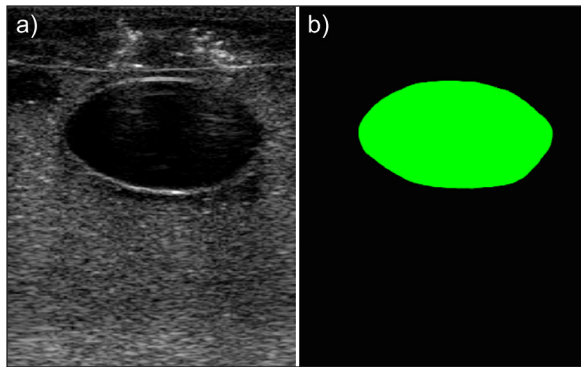**Figure 3.** The linear probe used for registration of images

Figure 4. (a) An image of the cyst-like structure in the phantom, (b) a manually segmented mask of the cyst-like structure

This approach was used to focus the thresholding on the cyst, rather than the full image, which may contain other similar objects. In the third step, a labelling procedure was performed on the thresholded image and cyst-like objects were classified using a simple and fixed decision tree, in which questions included relative positioning and relative area of the object. Finally, after the cyst-like object was selected from the image, a convex-hull operation was performed on it, to remove common artifacts present in US imaging from the final image.

Training and testing the algorithm

In order to analyze the predictive capabilities of the algorithm, the dataset was split into two subsets, one for training and one for testing. The training percentage was set to 20%, while the remaining 80% of the images were used to test the procedure. Two parameters of the algorithm – thr_1 and thr_2 – were set as trainable and their values were determined with a numerical search. These parameters represented the threshold values for the operations mentioned in the previous sections and were allowed to change in the following ranges: from 0.784 (or 200/255) to 0.933 (or 238/255) with a step of 0.008 (or 2/255) for thr_1 and from 0.9 to 1.0 with a step of 0.01 for thr_2. This resulted in 220 variants of the algorithm (based on the values of thr_1 and thr_2), trained and tested. The test/train procedure was repeated 60 times under different seed of the random number generator with dataset shuffling, in order to minimize the effect of luck on the results. The quality of each variant of the procedure was assessed using the DICE coefficient. For each rerun of the procedure, the

segmentation of the test set was performed using the variant with the highest average DICE from the training runs.

Assessing the quality of the results obtained from the algorithm

To assess the quality of the obtained results, the DICE coefficient [26] was used. The coefficient measures the relative difference between two objects and is widely employed in segmentation problems. The exact equation for it can be written as follows:

$$DICE = \frac{2|A \cap B|}{|A| + |B|} \qquad (2)$$

where: DICE – the DICE coefficient used to measure the quality of the segmentation (between 0.0 and 1.0, higher values signified better results), A – the segmented image, B – the ground truth image.

Furthermore, as the algorithm was meant for real-time augmentation of US images, the adopted measure was also the average time needed to segment a single frame from US on a consumer-grade hardware.

RESULTS AND DISCUSSION

As it was mentioned before, the acquired images were split into train and test subsets with 20:80 ratio with the training dataset containing only 4 samples. The different variants of the algorithm were tested over 60 reruns of the procedure with a different RNG seed for dataset splitting. The following subsections provide detailed results of the algorithm performance, as well as the summary of the best results.

The performance of the algorithm

The predictive capabilities of the algorithm were tested on the subset of data unseen in training, which consisted of 14 images. As it was mentioned before, the training and testing procedures were repeated 60 times with different seeds for the RNG, which resulted in different shuffling of the subsets. The DICE coefficients obtained on the 60 test sets were presented in Figure 5. As it can be seen in Table 1, the performance of the algorithm in the studied cases was relatively good. The DICE coefficient never dropped below
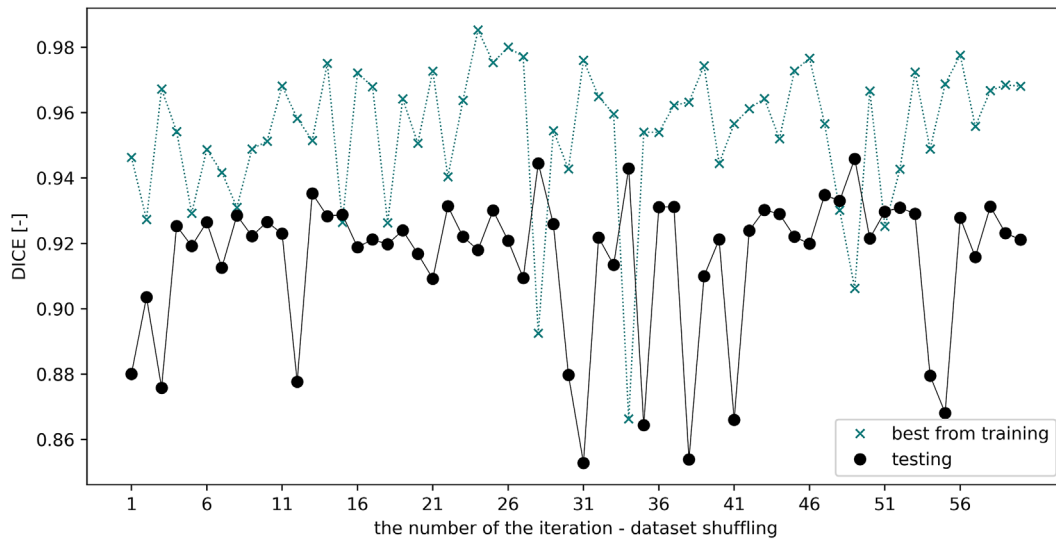
Figure 5. The DICE coefficients obtained on the 60 test sets for both training and testing of the procedure

Table 1. The summary of the obtained results in terms of DICE

Specification	Best DICE [-]	Worst DICE [-]	Average DICE [-]
Train	0.985	0.866	0.954 ± 0.022
Test	0.946	0.853	0.916 ± 0.022

0.85, while in the best test case it almost reached 0.95. On average, the algorithm returned a DICE of nearly 0.92 on the test set over 60 reruns of the procedure. The performance in training was slightly higher than in testing, which might suggest some overfitting, given the low number of training cases. On the other hand, the good results signify that the procedure did not require many samples to be optimized, which was its advantage. Note that these results should be taken with caution, as the studied cases included mainly convex-like objects, albeit with a lot of artifacts, high signal-to-noise ratio and low resolution. Nevertheless, in general the obtained results were promising.

Additionally, the procedure was able to achieve an averaged framerate of 35.7 FPS on a single CPU thread using modern consumer-grade hardware with unoptimized Python implementation, making it inexpensive to run and couple with US systems.

Selected results of the segmentation

Selected results obtained on the test data were presented in Figure 6. The algorithm was capable of segmenting images with high levels of artifacts caused by US. This was visible in most cases, but especially in the top-right case in Figure 6.

This performance was achieved by the finishing convex hull operation coupled with the trainable thresholds. Interestingly, the training procedure did not try to optimize the threshold alone, which with this level of noise would not be sufficient. Instead, its focus was shifted to thresholding the best image for the subsequent convex hull operation.

The relationship between the trainable thresholds

As it can be seen in Figure 7, the relationship between the two trainable parameters: thr_1 – threshold #1 and thr_2 – threshold #2 was nonlinear, albeit only very slightly. The best results in terms of DICE, higher than 0.92, were obtained for a variety of different combinations of parameters in the middle band of the graph.

In general, multiple variants of the algorithm resulted in good DICE coefficients. However, with some combinations DICE could be as low as 0.7, which meant that the trainable parameters had a significant impact on the results.

Limitations

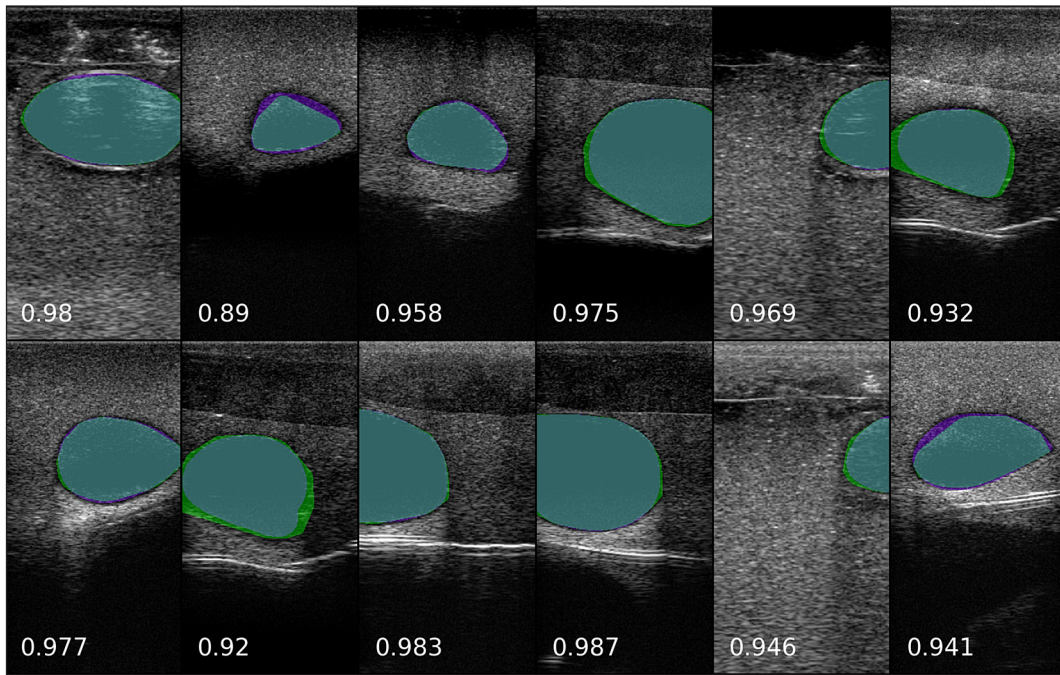In general, the results obtained from the procedure were very good. Nevertheless, in some

Figure. 6. Selected results obtained from segmentation on the test dataset. The numbers on represent the DICE coefficient. Colors correspond to the following: "green" – segmentation only, "violet" – ground truth only, "cyan" – ground truth and segmentation
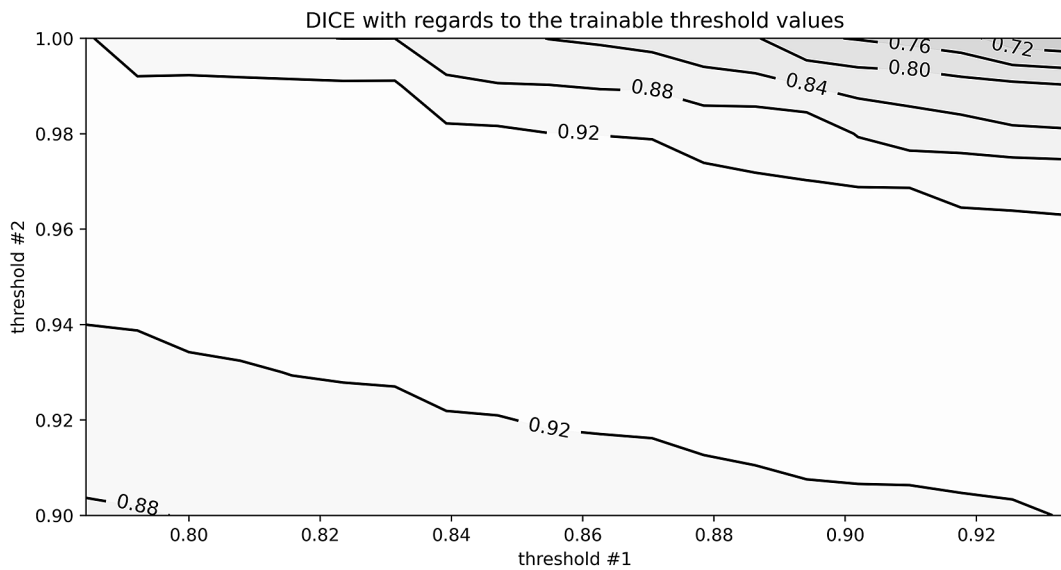

Figure. 7. The contour plot detailing the DICE obtained on the training set with regards to the two trainable parameters of the procedure: thr_1 – threshold #1 and thr_2 – threshold #2

cases the DICE coefficient could drop below 0.9, which was exemplified in Table 1. However, even these results would still be sufficient for some of the further operations in computer-aided-surgery and surgical training, such as computing the target point for aspiration.

While the performance of the algorithm was sufficient in the analyzed cases, it could potentially degrade with highly concave objects – this is a limitation of this approach and should be taken into consideration with the type and geometry of objects inserted into the phantom.

Although the segmented image already enhances the visual representation of the cyst-like object in a meaningful way, the algorithm should be coupled with other solutions, which can

transform the segmented image into a geometrically-optimal target point for aspiration or even compute the minimally-invasive path for the needle, such as in [27]. Furthermore, the generally good results under very limited training data signify that the method could potentially be used to automatically generate larger datasets for training more robust algorithms, with deep learning being one of the primary examples.

CONCLUSIONS

To summarize, the paper featured a trainable procedure for segmenting cyst-like objects in soft-tissue phantoms based on gelatin using limited number of training samples. The dataset was registered using an US device, while the procedure was tested under different seeds of the random number generator. The obtained results suggested the viability of the procedure. The DICE coefficients were on average at 0.92 over 60 reruns of the test/train procedure, while in the best cases exceeded 0.95. Nonetheless, the dataset was focused on cyst-like structures and might not represent other medical cases studied using soft tissue phantoms. The procedure was based on the assumption that the cyst-objects were close to convex, which should also be taken into account. The proposed method could be employed in systems for testing medical robots as well as frameworks for surgeon training, especially in regard to augmented visualization of tissues.

REFERENCES

1. Kuhl C.K., Schrading S., Leutner C.C., Morakkabati-Spitz N., Wardelmann E., Fimmers R., Kuhn W., and Schild H. Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. J. Clin. Oncol. Wolters Kluwer 2005, 23(33): 8469–76, https://doi.org/10.1200/JCO.2004.00.4960

2. Schoots I.G., Roobol M.J., Nieboer D., Bangma C. H., Steyerberg E.V., Hunink M.G.M. Magnetic resonance imaging–targeted biopsy may enhance the diagnostic accuracy of significant prostate cancer detection compared to standard transrectal ultrasound-guided biopsy: a systematic review and meta-analysis. Eur. Urol. 2015, 68(3): 438–50, https://doi.org/https://doi.org/10.1016/j.eururo.2014.11.037

3. Li J.W., Karmakar M.K., Li X., Fanzca W.H.K., Kee W.D.N. Gelatin-agar lumbosacral spine phantom. J. Ultrasound Med. John Wiley & Sons, Ltd 2011, 30(2): 263–72, https://doi.org/https://doi.org/10.7863/jum.2011.30.2.263

4. Richardson C., Bernard S., and Dinh V.A. A cost-effective, gelatin-based phantom model for learning ultrasound-guided fine-needle aspiration procedures of the head and neck. J. Ultrasound Med. John Wiley & Sons, Ltd 2015, 34(8): 1479–84, https://doi.org/https://doi.org/10.7863/ultra.34.8.1479

5. Selame L.A., Risler Z., Zakaria S.J., Hughes L.P., Lewiss R.E., Kehm K., Goodsell K.,

6. Kalwani R., Mirsch D., Kluger S.B., Au A. A comparison of homemade vascular access ultrasound phantom models for peripheral intravenous catheter insertion. J. Vasc. Access SAGE Publications Ltd STM 2020, 22(6): 891–97, https://doi.org/10.1177/1129729820961941

7. Zhao X., Ersoy E., Ng D.L. Comparison of low-cost phantoms for ultrasound-guided fine-needle aspiration biopsy training. J. Am. Soc. Cytopathol. 2023, 12(4): 275–83, https://doi.org/https://doi.org/10.1016/j.jasc.2023.03.005

8. Liang K., Light E.D., Rogers A.J., von Allmen D., Smith S.W. 3-D Ultrasound guidance of autonomous surgical robotics: Feasibility studies. Proc. - IEEE Ultrason. Symp. 2009, 582–85, https://doi.org/10.1109/ULTSYM.2009.5441429

9. Liang K., Rogers A.J., Light E.D., von Allmen D., Smith S.W. Three-dimensional ultrasound guidance of autonomous robotic breast biopsy: feasibility study. Ultrasound Med. Biol. 2010, 36(1): 173–77, https://doi.org/10.1016/j.ultrasmedbio.2009.08.014

10. Liang K., Rogers A.J., Light E.D., von Allmen D., Smith S.W. Simulation of autonomous robotic multiple-core biopsy by 3D ultrasound guidance. Ultrason. Imaging NIH Public Access 2010, 32(2): 118–27.

11. Giannotti E., Jethwa K., Closs S., Sun R., Bhatti H., Jameset J., Clarke C. Promoting simulation-based training in radiology: a homemade phantom for the practice of ultrasound-guided procedures. Br. J. Radiol. 2022, 95(1137): 20220354, https://doi.org/10.1259/bjr.20220354

12. Feng Y., Qiu S., Chen Y., Ma S. Viscoelastic Characterization of soft tissue-mimicking gelatin phantoms using indentation and magnetic resonance elastography. JoVE MyJoVE Corp 2022, (183): e63770, https://doi.org/doi:10.3791/63770

13. Antoniou A. and Damianou C. MR relaxation properties of tissue-mimicking phantoms. Ultrasonics 2022, 119: 106600, https://doi.org/https://doi.org/10.1016/j.ultras.2021.106600.

14. Jonasson H., Anderson C.D., and Saager R.B. Water and hemoglobin modulated gelatin-based phantoms to spectrally mimic inflamed tissue in the validation of biomedical techniques and the modeling of

microdialysis data. J. Biomed. Opt. 2022, 27(7): 74712, https://doi.org/10.1117/1.JBO.27.7.074712

15. Anugrah M.A., Suryani S., Ilyas S., Mutmainna I., Fahri A.N., Jusmawang, Tahir D. Composite gelatin/Rhizophora SPP particleboards/PVA for soft tissue phantom applications. Radiat. Phys. Chem. 2020, 173: 108878, https://doi.org/https://doi.org/10.1016/j.radphyschem.2020.108878

16. Houten E.E.W. Van, Viviers D., McGarry M.D.J., vR., Perriñez P.R., Perreard I.I., Weaver J. B., Paulsen K.D. Subzone based magnetic resonance elastography using a Rayleigh damped material model. Med. Phys. John Wiley & Sons, Ltd 2011, 38(4): 1993–2004, https://doi.org/https://doi.org/10.1118/1.3557469

17. Huang Q., Luo Y., Zhang Q. Breast ultrasound image segmentation: a survey. Int. J. Comput. Assist. Radiol. Surg. 2017, 12(3): 493–507, https://doi.org/10.1007/s11548-016-1513-1

18. Huang J., Triedman J.K., Vasilyev N.V, Suematsu Y., Cleveland R.O., Dupont P.E. Imaging artifacts of medical instruments in ultrasound-guided interventions. J. Ultrasound Med. John Wiley & Sons, Ltd 2007, 26(10): 1303–22, https://doi.org/https://doi.org/10.7863/jum.2007.26.10.1303

19. Ramme A.J., Guss M.S., Vira S., Vigdorchik J.M., Newe A., Raithel E., Chang G. Evaluation of automated volumetric cartilage quantification for hip preservation surgery. J. Arthroplasty 2015, https://doi.org/10.1016/j.arth.2015.08.009

20. Davignon F., Deprez J.-F., Basset O. A parametric imaging approach for the segmentation of ultrasound data. Ultrasonics 2005, 43(10): 789–801, https://doi.org/https://doi.org/10.1016/j.ultras.2005.06.001

21. Noh Y.-G., Ryu W.-S., Schellingerhout D., Park J., Chung J., Jeong S.-W., Gwak D.-S., Kim B.J., Kim J.-T., Hong K.-S., Lee K.B,, Park T.H., Park S.-S., Park J.-M., Kang K., Cho Y.-J., Park H.-K., Lee B.-C., Yu K.-H, Oh M.S., Lee S.J., Kim J.G., Cha J.-K., Kim D.-H., Lee J., Park M.S., Kim D., Bang O.Y., Kim E.Y., Sohn C.-H, Kim H., Bae H.-J., Kimet D.-E. Deep learning algorithms for automatic segmentation of acute cerebral infarcts on diffusion-weighted images: Effects of training data sample size, transfer learning, and data features. medRxiv 2023, 2023.07.02.23292150, https://doi.org/10.1101/2023.07.02.23292150

22. Maxwell A., Li R., Yang B., Weng H., Ou A., Hong H., Zhou Z., Gong P., Zhang

23. C. Deep learning architectures for multi-label classification of intelligent health risk prediction. BMC Bioinformatics 2017, 18(14): 523, https://doi.org/10.1186/s12859-017-1898-z

24. Capelin M., Martinez G.A.S., Xing Y., Siquerila A. Analysis of wire rolling processes using convolutional neural networks. Adv. Sci. Technol. Res. J. 2024, 18(2): 103–14, https://doi.org/10.12913/22998624/183699

25. Plaksyvyi A., Skublewska-Paszkowska M., and Powroznik P. A comparative analysis of image segmentation using classical and deep learning approach. Adv. Sci. Technol. Res. J. 2023, 17(6): 127–39, https://doi.org/10.12913/22998624/172771

26. Ciszkiewicz A., Milewski G., Lorkowski J. Baker's Cyst Classification Using Random Forests. In: Proceedings of the Federated Conference on Computer Science and Information Systems 2018, https://doi.org/10.15439/2018F89.

27. Walt S. van der, Schönberger J.L., Nunez-Iglesias J., Boulogne F., Warner J.D., Yager N., Gouillart E., Yu T. scikit-image: image processing in Python. Peer J 2014, 2: e453, https://doi.org/10.7717/peerj.453

28. Dice L.R. Measures of the amount of ecologic association between species. Ecology John Wiley & Sons, Ltd 1945, 26(3): 297–302, https://doi.org/https://doi.org/10.2307/1932409

29. Ciszkiewicz A., Lorkowski J., Milewski G. Differential evolution and cost-maps for needle path planning in Baker's cyst aspiration. Acta Bioeng. Biomech. 2022, 24(4): 39–47, https://doi.org/10.37190/ABB-02157-2022-02