— Supplementary Material —
DeepEdit: Deep Editable Learning for Interactive Segmentation of 3D Medical Images

Andres Diaz-Pinto[1,6], Pritesh Mehta[1], Sachidanand Alle[6], Muhammad Asad[1], Richard Brown[1], Vishwesh Nath[6], Alvin Ihsani[6], Michela Antonelli[1], Daniel Palkovics[2], Csaba Pinter[3], Ron Alkalay[4], Steve Pieper[5], Holger R. Roth[6], Daguang Xu[6], Prerna Dogra[6], Tom Vercauteren[1], Andrew Feng[6], Abood Quraini[6], Sebastien Ourselin[1], and M. Jorge Cardoso[1]

[1] School of Biomedical Engineering & Imaging Sciences, King's College London. London, UK. {andres.diaz-pinto,m.jorge.cardoso}@kcl.ac.uk
[2] Department of Periodontology, Semmelweis University, Hungary
[3] EBATINCA, S.L. Canary Islands, Spain.
[4] Beth Israel Deaconess Medical Center, MA, USA.
[5] Isomics, Inc., MA, USA.
[6] NVIDIA Santa Clara, CA, USA.

Abstract. Automatic segmentation of medical images is a key step for diagnostic and interventional tasks. However, achieving this requires large amounts of annotated volumes, which can be tedious and time-consuming task for expert annotators. In this paper, we introduce DeepEdit, a deep learning-based method for volumetric medical image annotation, that allows automatic and semi-automatic segmentation, and click-based refinement. DeepEdit combines the power of two methods: a non-interactive (i.e. automatic segmentation using nnU-Net, UNET or UN-ETR) and an interactive segmentation method (i.e. DeepGrow), into a single deep learning model. It allows easy integration of uncertainty-based ranking strategies (i.e. aleatoric and epistemic uncertainty computation) and active learning. We propose and implement a method for training DeepEdit by using standard training combined with user interaction simulation. Once trained, DeepEdit allows clinicians to quickly segment their datasets by using the algorithm in auto segmentation mode or by providing clicks via a user interface (i.e. 3D Slicer, OHIF). We show the value of DeepEdit through evaluation on the PROSTA-TEx dataset for prostate/prostatic lesions and the Multi-Atlas Labeling Beyond the Cranial Vault (BTCV) dataset for abdominal CT segmentation, using state-of-the-art network architectures as baseline for comparison. DeepEdit could reduce the time and effort annotating 3D medical images compared to DeepGrow alone. Source code is available at https://github.com/Project-MONAI/MONAILabel

Keywords: Interactive Segmentation · Deep Learning · CNNs

1   Segmentation of Metastatic Spines

1.1   Clinical Motivation

In 2020, an estimated 1,806,590 new cancer cases will be diagnosed in the United States, and 606,520 people will die from the disease. Vertebral bone metastases are found in up to 70% of patients with advanced cancers [1], and for many patients, it is the first sign of malignancy [2]. Pathologic vertebral fractures (PVF) occur in 15-30% of patients with spinal bone metastasis with up to 50% of the patients suffering from severe neurological deficits, leading to shortened patient survival with further complications that may be fatal. The associated morbidity incurs an annual cost of USD 12 billion. However, in the absence of quantitative fracture risk assessment, standard-of-care imaging modalities simply do not provide precise estimates of fracture risk [3]. This limited ability to predict risk of PVF has a profound, negative impact on the clinical care of these patients [4]. Clinicians are thus forced to guess when deciding whether stabilizing surgery should be offered or withheld, with interventions largely offered as reactive to pain and new neurological deficits. This can be particularly tragic when surgery or vertebral augmentation could have been offered before development of these complications.

Invasion of the vertebral bone with metastatic tumors [5] destroys vertebral anatomy, causes remarkable alteration or destruction of bone architecture, and degrades bone material properties. PVF will occur when the culminated effect of these changes renders the affected vertebra unable to withstand applied spinal loading generated in daily activities. Computed tomography (CT) is the most spatially accurate imaging modality to assess 3D osseous vertebral anatomy and spatial bone morphology [6]. Computational analysis, derived from high resolution [7] and clinical [8] CT, demonstrated that directly modeling the lesion's effect on the localized modulus and spatial distribution of the bone within the pathologic vertebral volume predicts up to 78% of the variance in modulus and strength of the affected vertebra, irrespective of the type of metastasis, or degree of anatomical involvement. Establishing a clinical image-based, patient-specific methodology for precisely predicting the effect of metastatic bone lesions on vertebral strength as a contributor to fracture risk will provide actionable information for the managing clinician addressing a critical gap in the management of patients' spine metastases [9].

1.2   Vertebral Segmentation

Accurate extraction of the vertebral geometry [10] and pathological alterations to this anatomy from the CT data is the fundamental, first, task in applying computational analysis for assessing the effect of bone metastasis on the affected vertebral strength. Human vertebrae exhibit complex, physically extended, multisegmented articulating anatomy, which varies significantly along the spinal column with vertebral size affected by gender while varying across populations. These articulating regions (ribs, posterior vertebral elements) pose

a significant challenge for the segmentation as current clinical CT scanners suffer from inherently insufficient resolution for separating the anatomical detail of these articulating regions in the aging and disease spine. The resulting partial volume effects yield image-based artifacts when separating similar anatomical structures in close vicinity, for example, fusing the facet joints of two adjacent vertebrae or between the vertebrae and ribs, all of which results in a highly challenging segmentation process. See Fig. 1
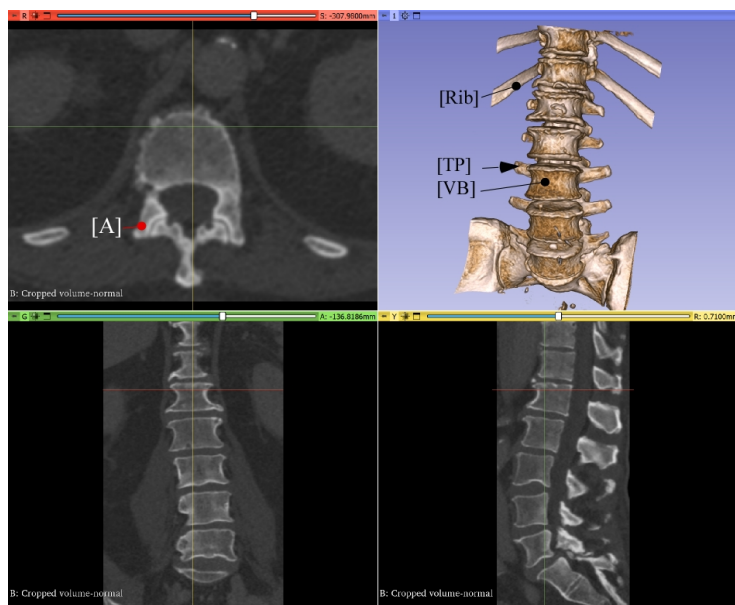
Fig. 1. Thoraco-lumbar spine image illustrating the vertebral body (VB) and transverse process (TP) are common to both lumbar and thoracic vertebrae with the ribs connected to the thoracic vertebrae. Clinical CT limited resolution regarding the anatomy of the spines posterior elements often in the "fusing" of the posterior element is indicated [A].

Manual vertebral segmentation and labeling remain the gold standard for segmenting vertebrae. This process, however, is labor-intensive, subjective [11], prone to errors, and is ill-suited for large-scale patient studies. Early methods for unsupervised image-based segmentation of vertebrae included region growing with boundary adjustment [12], watershed [13], graph-cut [14], level set [15] as well as applied mathematical morphology and watershed approaches due to their ability to handle the complex topological merging and breaking in the vertebrae. Model-based approaches using geometric models [16], Markov random fields [17], statistical shape models [18], and active contours [19] involve deforming shape priors to the spine to match the given spine. However, these approaches' deep dependence on a priori information may limit their generalizability. In recent years,

machine learning approaches allowed data-driven learning of the vertebral shape using deep neural networks [20]. Continual improvement involving localization of the spine, followed by U-Net multi-class [21], or 3D convolutional neural networks (CNN) [22] segmentation, was proposed to achieve vertebral segmentation from the lower-resolution masks. Able to utilize the increasing availability of curated, public CT data with good-quality annotations established as part of the computational spine imaging (CSI) workshops at MICCAI 2014-2020, these models have shown high-performance spine segmentation when applied to the Large Scale Vertebrae Segmentation challenge (VerSe) [23].

1.3   Segmentation of pathologic vertebrae

The effect of metastatic bone lesions on vertebral anatomy, bone structure, and architecture [24] poses unique challenges for vertebral segmentation (See Fig. 2). Depending on the type and origin of the metastasis, vertebral anatomy may present, singly or combined, extensive degradation and partial or extensive destruction of vertebral anatomy [24]. Vertebral fractures are common, yielding markedly altered vertebral shape and loss of vertebral dimensions [9]. Notably, the bone radiological appearance within the vertebral volume may undergo extensive alteration due to the metastases, resulting in a highly non-uniform distribution of HU values. For example, areas of markedly higher HU values due to increased bone density in osteosclerotic (bone-building) lesions or markedly diminished HU values In osteolytic (bone destroying) lesions. The latter significantly reduces or completely obliterates the bone's image and textural contrast to other soft tissues [13]. The resulting low local contrast between the bone, the marrow, and the surrounding soft tissues [25], yielding overlapping intensity ranges between the bone and the soft tissues, i.e., fat, marrow, and the intervertebral disc, further complicating the separation of functional and structural skeletal tissues.

1.4   Clinical Data Set

Clinically, CT scanning is employed to assess osseous destruction or reactive sclerosis within the spinal column offering an isotropic spatial resolution of about 0.5mm and an in-plane acquisition matrix of 512×512, yielding at best a pixel resolution of 0.3125mm in-plane. Image-based differentiation of musculoskeletal tissues is achieved by linearly transforming the distribution of the linear attenuation coefficients to integer-valued Hounsfield units (HU). Typical musculoskeletal values range from 20-80 HU for muscle and soft tissue, 150-300 HU for the trabecular bone (foam-based bone structure comprising the majority of the vertebral body), and 200-500 HU for the vertebral cortex (a thickened shell that surrounds the trabecular bone) [26]. With an average of 1000-1300 axial slices acquired for a complete thoraco-lumbar scan, the resulting volumetric data set ranges from 0.8-1.3 GB. Our expected data sets comprise of up to 430 cancer patients, with each patient scanned at radiation planning and four following
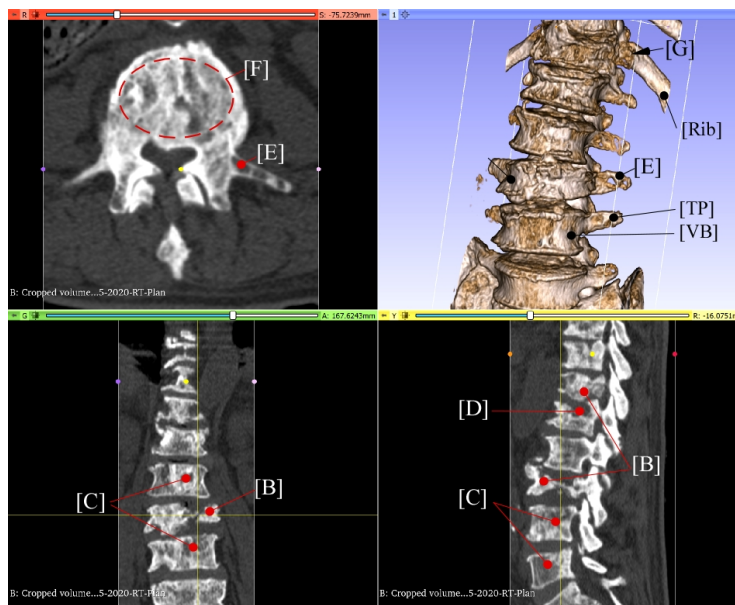
Fig. 2. Human thoraco-lumbar spine afflicted with prostate metastatic spine disease: VB stands for the vertebral body, TP for transverse process and Rib for the thoracic ribs. The CT image illustrates a critical vertebral fracture [B] with associated loss of neural canal space, which may result in compression of the neural cord, causing significant morbidity and may be fatal. At the vertebral body, such fractures often are radiologically displayed by spatial structural discontinuitywithin the VB cross-section [F]. The non-uniform radiological presentation [C] and the degradation of vertebral anatomy [D] manifest the metastasis effect on vertebral bone composition and architecture. Common is the destruction of vertebral structures a the TP [E] and Osteolytic mediated destruction of the rib-vertebral joint [G] affects the mechanical stability of the spine.

time points, i.e., 3,6,9 and 12 months post-therapy. In total, we are expecting to analyze high-resolution 2150 CT volumes.

1.5   DeepEdit Results

Given these challenges, applying DeepEdit allows to perform automatic inference as using standard deep learning models while providing clinicians with the ability to refine the predicted labels using clicks. In the absence of publicly available curated data sets for metastatic spines, the model integration of active learning strategies provides a highly adaptive mechanism to account for the bone metastases mediated degradation of the vertebral anatomy. We aim to develop an accurate method for automated segmentation of human spines invariant to the effect of bone metastases on spinal anatomy. Such a robust segmentation approach is critical to developing precise, patient-specific prediction

of fracture risk in pathologic vertebral, addressing a critical gap in managing patients' spine metastases. We further see this approach as instrumental in creating labeled spine CT datasets for metastatic spines, necessary for developing and optimizing machine learning approaches to segment pathological spines.

As a first step to address this clinical problem, we trained DeepEdit on the VerSe'19 dataset [23]. For this we used 80% (80 volumes) for training and 20% (20 volumes) for validation. Best, average and low quality segmentations are shown in Fig 3:


(a)   (b)   (c)
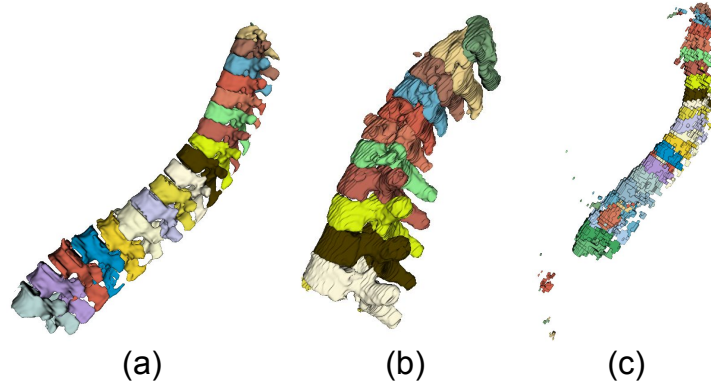
Fig. 3. Results obtained on the VerSe'19 dataset: DeepEdit applied to segment vertebra on the VerSe'19 dataset. Different quality of the obtained results from the DeepEdit in automatic segmentation mode: (a) best result, (b) average result, and (c) low quality result.

2 Treatment Planning in Reconstructive Periodontal Surgery

2.1 Clinical Motivation

Periodontitis is an infectious disease, as a result complex periodontal- and/or alveolar bone defects may develop that jeopardize general health in the oral cavity. During the complex multi-staged surgical – prosthetic rehabilitation of advanced periodontal cases, the primary determining local factor in the selection of the best possible surgical approach and regenerative strategy is the morphology of the bone defect. Standard radiographic diagnosis processes (intraoral radiographs, orthopantomograms) provide provide 2D images, in which overlapping anatomical structures make it difficult to determine the true defect morphology. Consequently, the applied surgical technique may not be sufficient for the successful treatment of the selected defect morphology. For this reason various authors have suggested the application of cone-beam coputed tomography

(CBCT) scans for diagnostic purposes in periodontology, however mostly planar images or 3D renders reconstructed with global thresholding methods were utilized which do not provide an accurate visualization of defect morphologies (Fig. 4.). To overcome these limitations Palkovics et al. [27] utilized a semi-automatic segmentation method for the 3D reconstruction of CBCT scans for diagnosis and treatment planning of periodontal- and alveolar bone defects. In their described method alveolar bone and teeth were segmented separately to acquire a more realistic virtual model of the clinical sitation. A notable drawback of this method, however is the lengthy duration and repetitive nature of the segmentation process. Other segmentation methods in utilized in dento-maxillofacial radiology (DMFR) include: (i) local thresholding methods, (ii) region growing, (iii) watershed and more recently (iv) AI-based segmentation methods [28].
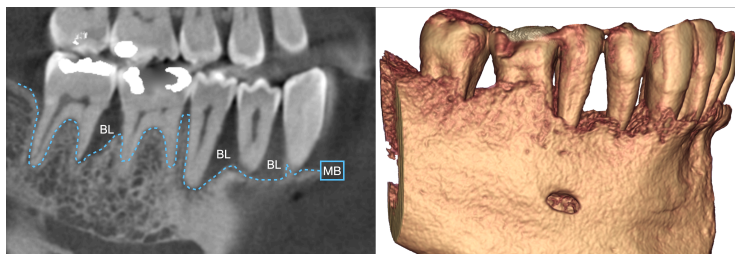
Fig. 4. CBCT image of a periodontally compromised patient, demonstrating horizontal and vertical periodontal bone defects (BL: bone loss; MB: marginal bone).

To reduce the timeframe of model acquisition process, an AI-based segmentation method could be developed for this purpose. Current trends in scientific research are geared towards the development of convolutional neural networks (CNN) for CBCT segmentation [28]. Few articles describe the segmentation of CBCT dataset using CNNs for the detection of periapical pathologies [29], or the localization of anatomical landmarks for the digital planning of orthodontic/orthognathic treatments [30]. This is most likely due to the fact that these applications are the least effected by the limitations of CNNs, which is the heterogeneity of sample databases [28]. Due to the morphological diversity of periodontal- and alveolar ridge defects, there is no available literature data on a reliable CNN or other AI-based method available for this application.

Hence, our aim is to develop an accurate and reliable method for fully automatic segmentation of CBCT datasets to acquire realistic 3D models of dental and alveolar structures for diagnostic and treatment planning purposes.

2.2 Clinical Data Set

In the current investigation 20 anonymized CBCT datasets of real patients, previously diagnosed with periodontitis have been prepared for training. Slice

interval was set to 0.2 mm in all directions (x,y,z) for all datasets. Based on literature data this is the lowest image resolution that is still feasible for surgical planning. For the preliminary testing it was selected to reduce processing time. Image sizes varied between $8.18cm \times 8.18cm \times 5.12cm$ ($409 \times 409 \times 256$ slices) and $16cm \times 16cm \times 16cm$ ($800 \times 800 \times 800$ slices).

Semi-automatic segmentation of the mandible and teeth have been carried out in 3D Slicer utilizing region growing. Separate binary labelmaps were generated for the alveolar bone and the teeth during the process (Fig. 5.). Based on the number of remaining teeth, label numbers varied between 4 and 16. After segmentation a custom terminology was created to ensure standardized labeling between the datasets.
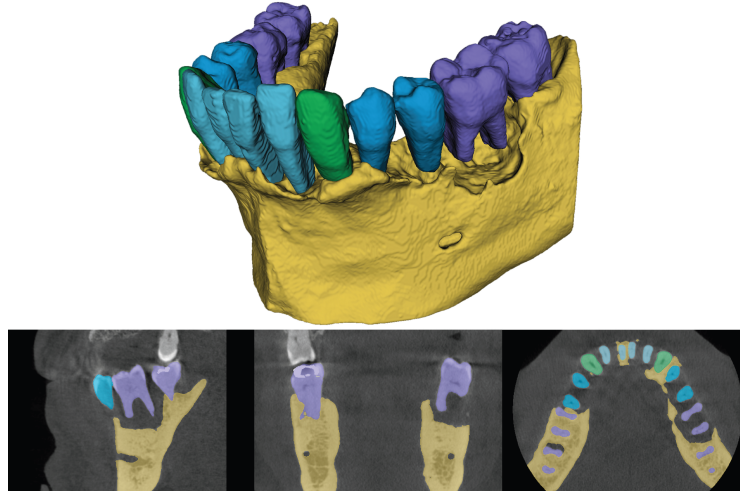
Fig. 5. Labels used to train DeepEdit on teeth segmentation: Multilabel application for teeth and alveolar bone segmentation.

2.3   DeepEdit Results

DeepEdit is a feasible tool for both automatic segmentation and label refinement of CBCT datasets in the field of periodontology and reconstructive dentoalveolar surgery. As DeepEdit allows easy integration of active learning strategies, the obtained model could be robust enough for such a difficult task. Other deep learning applications can only provide automatic inference or only semiautomatic inference (at least one click is needed), which makes the segmentation task less flexible for the clinician to annotate teeth and alveolar bone. Preliminary results of using DeepEdit on the CBCT dataset (16 volumes for training and 4 for validation) are presented in Fig 6.
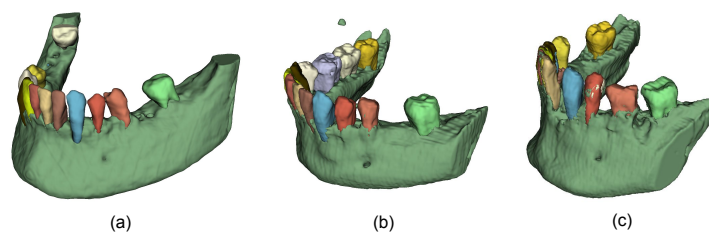
Fig. 6. Obtained results from DeepEdit on automatic inference mode: Different quality of the obtained results from the DeepEdit in automatic segmentation mode: (a) best result, (b) average result, and (c) low quality result.

References

1. R. Siegel, C. DeSantis, K. Virgo, K. Stein, A. Mariotto, T. Smith, D. Cooper, T. Gansler, C. Lerro, S. Fedewa, *et al.*, "Cancer treatment and survivorship statistics, 2012," *CA: a cancer journal for clinicians*, vol. 62, no. 4, pp. 220–241, 2012.
2. D. Schiff, B. O'Neill, and V. J. Suman, "Spinal epidural metastasis as the initial manifestation of malignancy: clinical features and diagnostic approach," *Neurology*, vol. 49, no. 2, pp. 452–456, 1997.
3. S.-H. Sung and U.-K. Chang, "Evaluation of risk factors for vertebral compression fracture after stereotactic radiosurgery in spinal tumor patients," *Korean Journal of Spine*, vol. 11, no. 3, p. 103, 2014.
4. M. H. Weber, S. Burch, J. Buckley, M. H. Schmidt, M. G. Fehlings, F. D. Vrionis, and C. G. Fisher, "Instability and impending instability of the thoracolumbar spine in patients with spinal metastases: a systematic review," *International journal of oncology*, vol. 38, no. 1, pp. 5–12, 2011.
5. J. Fornetti, A. L. Welm, and S. A. Stewart, "Understanding the bone in cancer metastasis," *Journal of Bone and Mineral Research*, vol. 33, no. 12, pp. 2099–2113, 2018.
6. J. Yao, J. E. Burns, D. Forsberg, A. Seitel, A. Rasoulian, P. Abolmaesumi, K. Hammernik, M. Urschler, B. Ibragimov, R. Korez, *et al.*, "A multi-center milestone study of clinical vertebral ct segmentation," *Computerized Medical Imaging and Graphics*, vol. 49, pp. 16–28, 2016.
7. M. Staehler, N. Haseke, P. Nuhn, C. Tüllmann, A. Karl, M. Siebels, C. G. Stief, B. Wowra, and A. Muacevic, "Simultaneous anti-angiogenic therapy and single-fraction radiosurgery in clinically relevant metastases from renal cell carcinoma," *BJU International-British Journal of Urology*, vol. 108, no. 5, p. 673, 2011.
8. R. Alkalay, R. Adamson, A. Miropolsky, and D. Hackney, "Female human spines with simulated osteolytic defects: Ct-based structural analysis of vertebral body strength," *Radiology*, vol. 288, no. 2, pp. 436–444, 2018.
9. M. H. Bilsky, E. Lis, J. Raizer, H. Lee, and P. Boland, "The diagnosis and treatment of metastatic spinal tumor," *The oncologist*, vol. 4, no. 6, pp. 459–469, 1999.
10. R. McBroom, W. Hayes, W. Edwards, R. Goldberg, and A. White 3rd, "Prediction of vertebral body compressive fracture using quantitative computed tomography.," *JBJS*, vol. 67, no. 8, pp. 1206–1214, 1985.
11. E. Chow, L. Holden, J. Rubenstein, M. Christakis, K. Sixel, M. Vidmar, J. Finkelstein, C. Hayter, A. Loblaw, R. Wong, *et al.*, "Computed tomography (ct) evalu-

ation of breast cancer patients with osteolytic bone metastases undergoing palliative radiotherapy—a feasibility study," *Radiotherapy and Oncology*, vol. 70, no. 3, pp. 291–294, 2004.

12. Y. Kang, K. Engelke, and W. A. Kalender, "A new accurate and precise 3-d segmentation method for skeletal structures in volumetric ct data," *IEEE transactions on medical imaging*, vol. 22, no. 5, pp. 586–598, 2003.

13. H. Li and Z. Wang, "A seepage flow model for vertebra ct image segmentation," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pp. 6364–6367, IEEE, 2006.

14. M. S. Aslan, A. Ali, H. Rara, and A. A. Farag, "An automated vertebra identification and segmentation in ct images," in *2010 IEEE international conference on image processing*, pp. 233–236, IEEE, 2010.

15. Y. Li, W. Liang, J. Tan, and Y. Zhang, "A novel automatically initialized level set approach based on region correlation for lumbar vertebrae ct image segmentation," in *2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings*, pp. 291–296, IEEE, 2015.

16. D. Štern, B. Likar, F. Pernuš, and T. Vrtovec, "Parametric modelling and segmentation of vertebral bodies in 3d ct and mr spine images," *Physics in Medicine & Biology*, vol. 56, no. 23, p. 7505, 2011.

17. S. Kadoury, H. Labelle, and N. Paragios, "Spine segmentation in medical images using manifold embeddings and higher-order mrfs," *IEEE transactions on medical imaging*, vol. 32, no. 7, pp. 1227–1238, 2013.

18. A. Rasoulian, R. Rohling, and P. Abolmaesumi, "Lumbar spine segmentation using a statistical multi-vertebrae anatomical shape+ pose model," *IEEE transactions on medical imaging*, vol. 32, no. 10, pp. 1890–1900, 2013.

19. M. E. Leventon, W. E. L. Grimson, and O. Faugeras, "Statistical shape influence in geodesic active contours," in *5th IEEE EMBS International Summer School on Biomedical Imaging, 2002.*, pp. 8–pp, IEEE, 2002.

20. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Springer, 2015.

21. A. Sekuboyina, A. Valentinitsch, J. S. Kirschke, and B. H. Menze, "A localisation-segmentation approach for multi-label annotation of lumbar vertebrae using deep nets," *arXiv preprint arXiv:1703.04347*, 2017.

22. N. Lessmann, B. Van Ginneken, P. A. De Jong, and I. Išgum, "Iterative fully convolutional neural networks for automatic vertebra segmentation and identification," *Medical image analysis*, vol. 53, pp. 142–155, 2019.

23. A. Sekuboyina, M. E. Husseini, A. Bayat, M. Löffler, H. Liebl, H. Li, G. Tetteh, J. Kukačka, C. Payer, D. Štern, *et al.*, "Verse: a vertebrae labelling and segmentation benchmark for multi-detector ct images," *Medical image analysis*, vol. 73, p. 102166, 2021.

24. S. Bailey, D. Hackney, D. Vashishth, and R. N. Alkalay, "The effects of metastatic lesion on the structural determinants of bone: Current clinical and experimental approaches," *Bone*, vol. 138, p. 115159, 2020.

25. J. S. Bauer, T. M. Link, A. Burghardt, T. D. Henning, D. Mueller, S. Majumdar, and S. Prevrhal, "Analysis of trabecular bone structure with multidetector spiral computed tomography in a simulated soft-tissue environment," *Calcified tissue international*, vol. 80, no. 6, pp. 366–373, 2007.

26. J. J. Schreiber, P. A. Anderson, and W. K. Hsu, "Use of computed tomography for assessing bone mineral density," *Neurosurgical focus*, vol. 37, no. 1, p. E4, 2014.

27. D. Palkovics, F. G. Mangano, K. Nagy, and P. Windisch, "Digital three-dimensional visualization of intrabony periodontal defects for regenerative surgical treatment planning," *BMC oral health*, vol. 20, no. 1, pp. 1–11, 2020.
28. A. F. Leite, K. d. F. Vasconcelos, H. Willems, and R. Jacobs, "Radiomics and machine learning in oral healthcare," *PROTEOMICS–Clinical Applications*, vol. 14, no. 3, p. 1900040, 2020.
29. K. Orhan, I. Bayrakdar, M. Ezhov, A. Kravtsov, and T. Özyürek, "Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans," *International endodontic journal*, vol. 53, no. 5, pp. 680–689, 2020.
30. P.-J. Verhelst, A. Smolders, T. Beznik, J. Meewis, A. Vandemeulebroucke, E. Shaheen, A. Van Gerven, H. Willems, C. Politis, and R. Jacobs, "Layered deep learning for automatic mandibular segmentation in cone-beam computed tomography," *Journal of Dentistry*, vol. 114, p. 103786, 2021.