Magdalena Bloier*, Florian Hinterwimmer, Sebastian Breden, Sarah Consalvo, Jan Neumann, Nikolas Wilhelm, Rüdiger von Eisenhart-Rothe, Daniel Rueckert, and Rainer Burgkart

Detection and Segmentation of Heterogeneous Bone Tumours in Limited Radiographs

https://doi.org/10.1515/cdbme-2022-1019

Abstract: Bone tumours are a rare and often highly malignant entity. Early clinical diagnosis is the most important step, but the difficulty of detecting and assessing bone malignancies is in its radiological peculiarity and limited experience of non-experts. Since X-ray imaging is the first imaging method of bone tumour diagnostics, the purpose of this study is to develop an artificial intelligence (AI) model to detect and segment the tumorous tissue in a radiograph. We investigated which methods are necessary to cope with limited and heterogeneous data. We collected 531 anonymised radiographs from our musculoskeletal tumour centre. In order to adapt to the complexity of recognizing the malignant tissue and cope with limited data, transfer learning, data augmentation as well as several architectures, some of which were initially designed for medical images, were implemented. Furthermore, dataset size was varied by adding another bone tumour entity. We applied a data split of 72%, 18%, 10% for training, validation and testing, respectively. To provide statistical significance and robustness, we applied a cross-validation and image stratification with respect to tumour pixels present. We achieved an accuracy of 99.72% and an intersection over union of 87.43% for hold-out test data by applying several methods to tackle limited data. Transfer learning and additional data brought the greatest performance increase. In conclusion, our model was able to detect and segment tumorous tissue in radiographs with good performance, although it was trained on a very limited amount of data. Transfer Learning and data augmentation proved to significantly mitigate the issue of limited data samples. However, to accomplish clinical significance, more data has to be acquired in the future. Through minor adjustments, the model could be adapted to other musculoskeletal tumour entities and become a general support tool for orthopaedic surgeons and radiologists.

Keywords: deep learning, sarcoma, bone tumour, detection, segmentation

***Corresponding author: Magdalena Bloier,** Department for Orthopaedics and Sports Orthopaedics, Klinikum rechts der Isar, Technical University of Munich, 81675 Munich, Germany, e-mail: magdalena.bloier@tum.de
Florian Hinterwimmer, Institute for AI and Informatics in Medicine, Technical University of Munich & Department for Orthopaedics and Sports Orthopaedics, Klinikum rechts der Isar, Technical University of Munich, 81675 Munich, Germany
Sebastian Breden, Sarah Consalvo, Nikolas Wilhelm, Rüdiger von Eisenhart-Rothe, Rainer Burgkart, Department for Orthopaedics and Sports Orthopaedics, Klinikum rechts der Isar, Technical University of Munich, 81675 Munich, Germany
Jan Neumann, Institute for Diagnostic and Interventional Radiology and Paediatric Radiology, Klinikum rechts der Isar, Technical University of Munich 81675 Munich, Germany
Daniel Rueckert, Institute for AI and Informatics in Medicine, Technical University of Munich, 81675 Munich, Germany

1 Introduction

Bone tumours are a rare disease overall [8], but are among the most common cancers in children and adolescents (>10% of all paediatric cancers). The complex and time-consuming diagnosis in a specialised centre includes clinical, radiological and histopathological steps as well as the subsequent interdisciplinary assessment in a specialised tumour board. A general practitioner, on the other hand, usually has only X-ray diagnostics and, because of the incidence, statistically encounters bone tumours less than three times in his/her professional life. As a result, sarcomas are often misdiagnosed [8] and prognostically essential time is lost, and patients are delayed in being referred to specialised sarcoma centres. Hence, new and more sophisticated techniques for early and reliable detection and evaluation of bone tumours are urgently needed. Deep learning (DL) is poised to reshape medicine and potentially improve the experience of physicians as well as patients [9]. DL has already had ample success in many medical disciplines [12]. In comparison, the impact and number of publications of DL in orthopaedics are very limited [4, 11]. Most certainly this can be explained by the low incidence of bone tumours (and soft tissue tumours as well) and the lack of sufficient data infrastructures. While limited data is a common obstacle to DL applications in medicine, even more so for image interpretation of bone malignancies. Thus, we developed a segmentation framework for heterogeneous and limited radiographs of

bone tumours, focusing on Transfer Learning [2], data augmentation [1, 3, 7] and leveraging various segmentation models [5, 10, 14] and configurations. In summary, we make the following contributions:

1. We demonstrate a novel approach for bone tumour assessment by detecting and segmenting malignancies with DL and limited and heterogeneous radiographs.
2. We illustrate the impact of transfer learning, data augmentation, different architectures and dataset size.
3. We provide a potential support tool to identify bone tumours, not only for expert centres, but also potentially targeting outpatients clinics and young physicians.

1.1 Related work

With the rise of DL, especially the task of segmentation [6] of medical images became more and more popular over the past decade [9]. Nonetheless, segmentation of bone tumours has only been presented a few times. Zhang et al. [16] proposed a multiple supervised residual network to segment osteosarcomas in CT images with good results (Dice score 0.89). In contrast, Schacky et al. [15] demonstrated a multi-task DL approach to classify, detect and segment bone tumours in radiographs with fair segmentation performance (Dice score 0.6). CT imaging is usually the imaging modality of choice for bone pathologies, because it provides more detailed information about the potential destruction of cortical bone. However, similar to Schacky et al., this study focused on detecting and segmenting bone tumours in standardised radiographs and identifying the most impactful methods for an imperfect dataset.

2 Materials and Methods

2.1 Dataset

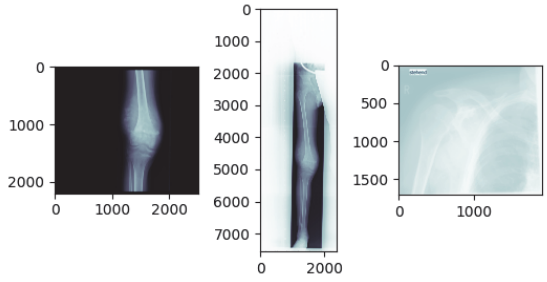

Fig. 1: Sample images before preprocessing.

We collected 531 preoperative radiographs from our musculoskeletal tumour centre from paediatric sarcoma patients. The dataset includes two subsets. The first subset with 44% of the entire dataset contains osteosarcoma and the second subset with 56% of the entire dataset contains chondrosarcoma. Typically, sarcoma occur in joints and long tubular bones and no restrictions regarding anatomical regions were applied. Also, the images available are heterogeneous in character, as can be seen in figure 1. They vary in dimension, resolution and data quality containing black or white background and marks. External images were also included in the dataset. No meta-information for statistical analysis was available and no further restrictions regarding age, musculoskeletal features or sex were made. Additionally, masks of the X-ray images including the location of the tumour were created. The masks are a binarised representation of tumour tissue and healthy tissue which were manually segmented by orthopaedic surgeons.

2.2 Model training

Model training and inference was conducted on a DGX Station A100 with four 80GB graphical processing units (Nvidia Corporation, Santa Clara, CA), 64 2.25 GHz cores and 512 GB DDR4 system memory running on a Linux/Ubuntu 20.04 distribution (Canonical, London, England). Preprocessing and model implementation were performed in Python 3.8.5 (https://www.python.org/) using PyTorch 1.10.2 and cuda toolkit 11.3 (https://pytorch.org/). The pretrained ConvNet model of this study will be provided upon publication.

2.3 Algorithm and experimental setup

Fig. 2: Illustration of workflow.

We developed a DL framework to train a segmentation network and evaluate it through the metrics accuracy and intersection over union (IoU). The dataset was preprocessed by removing the background areas, padding to create a square image, scaling to 256x256 and normalizing. The data split for pretraining was 72%, 18%, 10% for training, validation and hold-out testing, respectively. To provide statistical significance and robustness, images from all subsets were equally distributed with respect to tumour pixels present. An additional 5-fold cross-validation supported the task.
The initial dataset contained only the images with osteosar-

coma. As a baseline, we used U-Net-architecture as common performance baseline with ResNet34 to train our neural network and successively adapted several extensions to boost the performance of the model as shown in figure 2. In the first setup, instead of training from scratch, a pretrained model with Imagenet weights was implemented. Afterwards, the impact of data augmentation techniques was investigated in setup 2. Therefore, the current best model of setup 1 was extended by 12 different data augmentations including geometric transformations, cropping, filtering and intensity operations. Then, different data augmentation methods were combined by successively extending the amount of data augmentations using the operations with the highest IoU of the 12 data augmentation methods first. Among other things, so-called unrealistic data augmentations were used, which describe in particular operations that alter the images to such an extent that they no longer correspond to a medically realistic X-ray image as shown in figure 3 [3]. After determining the data augmentations leading to the highest IoU, we varied the model architecture. Model architectures selected were UNet++, DeepLabv3, DeepLabv3+ and MA-Net. In the last setup the data were duplicated by using the chondrosarcoma dataset to determine the influence of the amount of data.

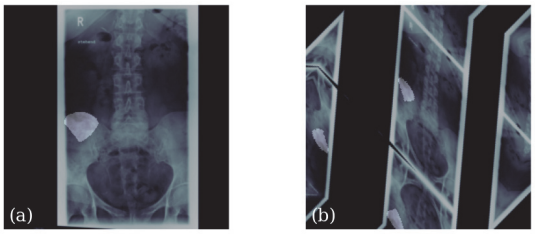

Fig. 3: Preprocessed sample image (a) and sample image with excessive data augmentation (b).

3 Results

Tab. 1: Overall results of the test dataset.

Setup number	Description	Accuracy	IoU
Baseline	U-Net with ResNet34	99.51 %	73.78 %
Setup 1	Transfer learning	99.59 %	79.58 %
Setup 2	Data augmentation	99.62 %	82.70 %
Setup 3	Model architecture	99.64 %	83.39 %
Setup 4	**Amount of data**	**99.72 %**	**87.43 %**

Transfer learning improved the baseline results to an IoU of 5.80% in setup 1 as shown in table 1. The best combination of additional data augmentations from configurations selected was the combined application of several affine transformations as shown in figure 1. In setup 3, we choose UNet++ architecture as model selected with the highest IoU compared to the model architectures U-Net, MA-Net, DeepLabv3 and DeepLabv3+. While MA-Net also had a relatively high IoU and accuracy of 83.19% and 99.62%, DeepLabv3 only reached an IoU 10.78% lower than UNet++. Through adding the chondrosarcoma images, we reached a final accuracy of 99.72% and an IoU of 87.43%. A prediction of a sample image of the final framework is shown in figure 4.

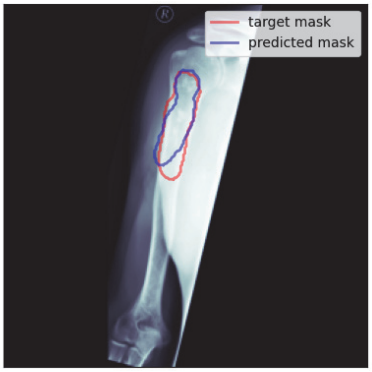

Fig. 4: Sample image with predicted and target mask.

4 Discussion

The main finding of this study was that significant results in detecting and segmenting heterogeneous bone tumour appearances in limited radiographs can be reached by implementing several methods to fine-tune the algorithm and tackle the issue of small datasets.

An improvement in accuracy and IoU could be achieved in each setup through an extending framework. For this task, using a pretrained neural network and adding more data leads to the biggest improvement of the segmentation task. It should be emphasized that the added data is that of another entity. Nevertheless, the datasets contain similar image features and therefore lead to improved performance. Additionally, data augmentation methods have shown to support image segmentation tasks. Further, while in literature realism is a goal for

many authors [1, 3], in our approach unrealistic data augmentations through for example affine transformations lead to better segmentation results than the realistic ones. However, the use of more data augmentations is not guaranteed to be beneficial as they might lead to poorer results. Choosing an appropriate model is crucial to get good segmentation results. In our task, UNet++ [14] and MA-Net [5, 13] lead to the best performance.

The major limitation of this study is the low amount of data available, hence, robustness of the model has to be further validated. In addition, the interpretation of the results must take into account that there is no gold standard for segmentation labels. Therefore, the segmentation labels need to be evaluated by an interdisciplinary team.

4.1 Conclusion

Transfer learning and an increased quantity of data even from another entity lead to the largest improvement of segmentation results, while varying the model architecture leads to the biggest differences in IoU and accuracy. In addition, unrealistic data augmentation through affine transformations supported the task. To achieve clinically relevant results, a systematic and structured collection of data to increase dataset size is of paramount importance.

Author Statement

Research funding: The author state no funding involved. Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board or equivalent committee.

Equal Contribution

Magdalena Bloier and Florian Hinterwimmer equally contributed to this article (shared first).

References

[1] Abdollahi, B., N. Tomita and S. Hassanpour (2020). Data augmentation in training deep learning models for medical image analysis. Deep learners and deep learner descriptors for medical applications, Springer: 167-180.

[2] Al-Rakhami, M. S., M. M. Islam, M. Z. Islam, A. Asraf, A. H. Sodhro and W. Ding (2021). "Diagnosis of COVID-19 from X-rays using combined CNN-RNN architecture with transfer learning." MedRxiv: 2020.2008. 2024.20181339.

[3] Chlap, P., H. Min, N. Vandenberg, J. Dowling, L. Holloway and A. Haworth (2021). A review of medical image data augmentation techniques for deep learning applications. Journal of Medical Imaging and Radiation Oncology 65(5): 545-563.

[4] Do, N.-T., S.-T. Jung, H.-J. Yang and S.-H. Kim (2021). Multilevel seg-unet model with global and patch-based X-ray images for knee bone tumor detection. Diagnostics 11(4): 691.

[5] Fan, T., Wang, G., Li, Y. and Wang, H. (2020). MA-Net: A Multi-Scale Attention Network for Liver and Tumor Segmentation. IEEE Access, 8, 179656–179665.

[6] Hesamian, M. H., Jia, W., He, X. and Kennedy, P. (2019). Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. Journal of Digital Imaging, 32(4), 582–596.

[7] Hussain, Z., F. Gimenez, D. Yi and D. Rubin (2017). Differential data augmentation techniques for medical imaging classification tasks. AMIA annual symposium proceedings, American Medical Informatics Association.

[8] Picci, P., M. Manfrini, D. M. Donati, M. Gambarotti, A. Righi, D. Vanel et al. (2020). Diagnosis of Musculoskeletal Tumors and Tumor-like Conditions: Clinical, Radiological and Histological Correlations-the Rizzoli Case Archive, Springer.

[9] Rajpurkar, P., E. Chen, O. Banerjee and E. J. Topol (2022). AI in health and medicine. Nat Med 28(1): 31-38.

[10] Ronneberger, O., P. Fischer and T. Brox (2015). U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention, Springer.

[11] Hinterwimmer, F., Lazic, I., Suren, C., Hirschmann, M., Pohlig, F., Rueckert, D., et al. (2022). "Machine learning in knee arthroplasty: specific data are key—a systematic review." Knee Surgery, Sports Traumatology, Arthroscopy: 1-13.

[12] Singh, P., N. Singh, K. K. Singh and A. Singh (2021). Diagnosing of disease using machine learning. Machine Learning and the Internet of Medical Things in Healthcare, Elsevier: 89-111.

[13] Shang, R., J. Zhang, L. Jiao, Y. Li, N. Marturi and R. Stolkin (2020). Multi-scale adaptive feature fusion network for semantic segmentation in remote sensing images. Remote Sensing 12(5): 872.

[14] Zhou, Z., M. M. Rahman Siddiquee, N. Tajbakhsh and J. Liang (2018). Unet++: A nested u-net architecture for medical image segmentation. Deep learning in medical image analysis and multimodal learning for clinical decision support, Springer: 3-11.

[15] von Schacky, C. E., N. J. Wilhelm, V. S. Schäfer, Y. Leonhardt, F. G. Gassert, S. C. Foreman, et al. (2021). Multitask deep learning for segmentation and classification of primary bone tumors on radiographs. Radiology 301(2): 398-406.

[16] Zhang, R., L. Huang, W. Xia, B. Zhang, B. Qiu and X. Gao (2018). Multiple supervised residual network for osteosarcoma segmentation in CT images. Computerized Medical Imaging and Graphics 63: 1-8.