Florian Hinterwimmer*, Sarah Consalvo, Jan Neumann, Carina Micheler, Nikolas Wilhelm, Jan Lang, Rüdiger von Eisenhart-Rothe, Rainer Burgkart, and Daniel Rueckert

From Self-supervised Learning to Transfer Learning with Musculoskeletal Radiographs

https://doi.org/10.1515/cdbme-2022-1003

Abstract: Ewing sarcomas are malignant neoplasm entities typically found in children and adolescents. Early detection is crucial for therapy and prognosis. Due to the low incidence the general experience as well as according data is limited. Novel support tools for diagnosis, such as deep learning models for image interpretation, are required. While acquiring sufficient data is a common obstacle in medicine, several techniques to tackle small data sets have emerged. The general necessity of large data sets in addition to a rare disease lead to the question whether transfer learning can solve the issue of limited data and subsequently support tasks such as distinguishing Ewing sarcoma from its main differential diagnosis (acute osteomyelitis) in paediatric radiographs. 42,608 unstructured radiographs from our musculoskeletal tumour centre were retrieved from the PACS. The images were clustered with a DeepCluster, a self-supervised algorithm. 1000 clusters were used for the upstream task (pretraining). Following, the pretrained classification network was applied for the downstream task of differentiating Ewing sarcoma and acute osteomyelitis. An untrained network achieved an accuracy of 81.5%/54.2%, while an ImageNet-pretrained network resulted in 89.6%/70.8% for validation and testing, respectively. Our transfer learning approach surpassed the best result by 4.4%/17.3% percentage points. Transfer learning demonstrated to be a powerful technique to support image interpretation tasks. Even for small data sets, the impact can be significant. However, transfer learning is not a final solution to small data sets. To achieve clinically relevant results, a structured and systematic data acquisition is of paramount importance.

Keywords: transfer learning, self-supervised learning, radiographs, sarcoma

***Corresponding author: Florian Hinterwimmer,** Technical University of Munich, Institute for AI and Informatics in Medicine & Department for Orthopaedics and Sports Orthopaedics, 81675 Munich, Germany, e-mail: florian.hinterwimmer@tum.de
Sarah Consalvo, Nikolas Wilhelm, Rüdiger von Eisenhart-Rothe, Rainer Burgkart, Technical University of Munich, Klinikum rechts der Isar, Department for Orthopaedics and Sports Orthopaedics, 81675 Munich, Germany
Jan Neumann, Technical University of Munich, Klinikum rechts der Isar, Institute for Diagnostic and Interventional Radiology and Paediatric Radiology, 81675 Munich, Germany
Daniel Rueckert, Technical University of Munich, Institute for AI and Informatics in Medicine, 81675 Munich, Germany
Carina Micheler, Technical University of Munich, Klinikum rechts der Isar, Department for Orthopaedics and Sports Orthopaedics, 81675 Munich, Germany & Institute for Machine Tools and Industrial Management, School of Engineering and Design, Technical University of Munich, Garching near Munich, Germany
Jan Lang, Technical University of Munich, Klinikum rechts der Isar, Department for Orthopaedics and Sports Orthopaedics, 81675 Munich, Germany & Chair of Non-destructive Testing, School of Engineering and Design, Technical University of Munich, Munich, Germany

1 Introduction

Ewing sarcomas are highly malignant tumour entities that occur predominantly in children and adolescents. Early detection and differentiation from other entities, especially acute osteomyelitis, are critical for therapy and prognosis and thus patient survival [17]. Because of the low incidence, experience especially in outpatient clinics is usually limited, so the chance of early detection is low [18]. Hence, new sophisticated diagnostic support tools are required. Deep learning (DL) has achieved great success in image interpretation in many other disciplines [11]. However, a common obstacle to the application of DL in medicine is the availability of a sufficient amount of data. Several techniques to cope with small data sets, such as data augmentation [15, 16], data synthesis, or transfer learning, have emerged. The general need of DL models for sufficient (training) data poses a challenge in the context of rare diseases. The question arises whether transfer learning can solve the problem of limited data and support specific tasks such as distinguishing Ewing sarcoma from acute osteomyelitis in pediatric radiographs. The presented study investigated if and how 42,608 unstructured radiographs can be integrated in a transfer learning approach to support minimal data sets in a classification task. In summary, we make the following contributions:

1. We demonstrate a novel transfer learning approach specifically developed with musculoskeletal radiographs by subsequent training of an already ImageNet-pretrained model.
2. We leverage a state-of-the-art self-supervised model to obtain weak auxiliary labels from 42,608 unstructured radiographs.

3. We underline the importance of sufficient data by showing that transfer learning is a powerful technique, but not a sole solution to limited data sets.

1.1 Related work

Transfer learning was first proposed in 1976 by Bozinovski and Fulgosi [4]. Since then, it has found various applications and shown great impact [2–4, 6–8]. The most popular transfer learning models are pretrained on ImageNet [8]. While these models are trained on every day images such as landscape-, cat- and dog-images, the pretraining still shows significant improvement also in medical image interpretation. Recently, several transfer learning approaches in the context of Covid19 detection and classification tasks have been published [2, 7]. These studies, due to the nature of the disease, focus on thorax images. However, to our knowledge, no model generally pretrained for musculoskeletal features in radiographs has been demonstrated.

2 Materials and methods

2.1 Data sets

The data set consisted of 42,608 unstructured, pseudonymised radiographs from a musculoskeletal tumour centre. All images belonged to patients with sarcoma associated ICD codes. Sarcomas typically occur in extremities and joints. Additionally, the data set contained images, which were initiated to check for metastases or monitor progress after surgery or therapy. Therefore, it is to be expected that any possible musculoskeletal region is included. The DICOM images were retrieved from the local PACS (Picture Archiving and Communication System) at Klinikum rechts der Isar (Munich). The imaging data was gathered over the past 25 years and contained corrupted and false data as well as heterogeneous data quality, resolution and external images. The DICOM header information was fully blinded, so that no meta-information for statistical analysis remained. For assessment of the transfer learning approach, a second data set consisting of 63 images (22 acute osteomyelitis, 41 Ewing sarcoma) from patients under 18 years of age was used. No further restrictions regarding age, musculoskeletal features or sex were made.

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, UK). Preprocessing and model implementation were performed in Python 3.9.6 (https://www.python.org/) using PyTorch 1.9.0 and cuda toolkit 11.1 (https://pytorch.org/). The pretrained model of this study will be provided upon publication.

2.3 Algorithm

We developed a two step deep learning framework to pretrain a classification network on an upstream task and subsequently evaluate it on a downstream task with different data from the same domain (musculoskeletal radiographs). In step one, the unstructered data set was clustered by a self-supervised model [5] into several clusters: DeepCluster presents a self-supervising approach to learning image representations. It iteratively groups features using k-means and uses the subsequent assignments as labels to update the weights of a network. The optimal number of clusters was determined through test runs measured by highest pretraining classification scores. In step two, the cluster assignments were used as "auxiliary" class labels for a classification task, whereby a ResNet50 [13] was pretrained. The data split for pretraining was 80%, 10%, 10% for training, validation and hold-out testing. Next, the pretrained model was applied to a two-entity classification task with limited samples for both entities with a data split of 80%, 10%, 10% for assessment of the transfer learning approach. To provide statistical robust results and avoid cross-contamination, a cross-validation was implemented. Accuracy values were calculated to evaluate the results. Figure 1 displays the workflow including the five steps from unstructured data to the final pretrained model.

2.4 Hyperparameters and runtime

For the upstream task a batch size of 512, a learning rate of 0.05 and 500 epochs were chosen. The runtime was $\sim$7.5 hours. For the downstream task a batch size of 4, a learning rate of 0.0001 and 100 epochs were chosen. The runtime for the all cross-validation folds was $\sim$2 hours. The inference step for all folds took $\sim$7 minutes.

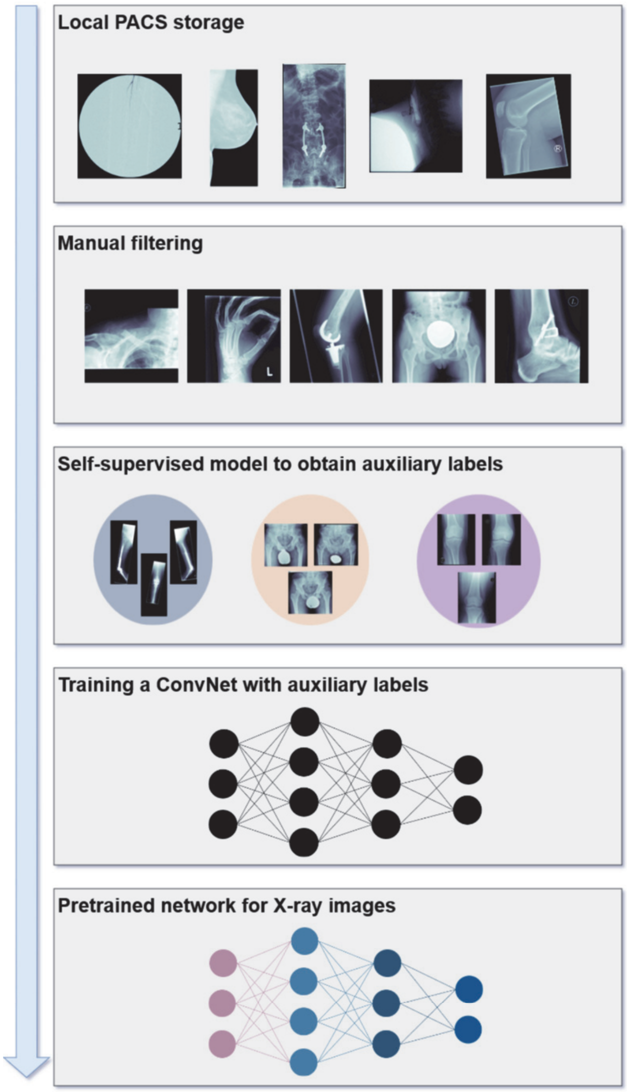

Fig. 1: Illustration of workflow with respective data samples.

Tab. 1: Classification results of Ewing sarcoma vs. acute osteomyelitis

Model	Val Acc	Test Acc
ResNet50	81.5 %	54.2 %
ResNet50 pretrained (ImageNet)	89.6 %	70.8 %
ResNet50 pretrained (our approach)	**94.0 %**	**88.1 %**

3 Results

3.1 Upstream task: from clustering to pretraining

The highest pretraining classification scores were achieved with 1000 clusters. The clustering of 42,608 images into 1000 clusters resulted in a normalised mutual information of 0.930. The smallest cluster was comprised of 2, the largest of 135 image samples with a first quartil of 20, a median of 46 and a third quartil of 55. The pretraining of the ResNet50 achieved an accuracy of 86.7%/80.0% for validation and testing respectively.

3.2 Downstream task: final classification

An untrained network achieved an accuracy of 81.5%/54.2%, while an ImageNet-pretrained network resulted in 89.6%/70.8% for validation and testing in the downstream task, respectively. Our transfer learning approach surpassed the best result by 4.4 and 17.3 percentage points (table 1).

4 Discussion

The most important finding of this study was that 42,608 unstructured radiographs can be utilised for transfer learning by leveraging a modern self-supervised model, thus significantly improving downstream classification tasks.

The obstacle of insufficient data for state-of-the-art deep learning applications is very common in medicine and especially in a field, such as orthopaedic oncology, where incidence is low and consequently data is limited. While collecting more quality data is probably the most effective way to improve deep learning applications in medicine, new techniques also need to be (further) developed. For example data augmentation [15, 16] or image synthesis [14] have shown to support various image interpretation tasks. We developed a transfer learning approach specified for radiographs with bone and soft tissue tumours. Most certainly though, our pretrained network will also improve other tasks working with radiographs of human

patients.
Another noticeable finding is that the not-pretrained network seemed to be overfitting and both pretrained networks seemed to mitigate this effect, thus, underlining the positive impact of pretraining with bigger data sets.
The major limitation of this study is that in contrast to the common data sets applied for pretraining (for example ImageNet [8], currently more than 14 million images), our data set is still comparably small. Therefore, the overall validity is still to be proven. However, for the particular task of distinguishing radiographs of Ewing and acute osteomyelitis patients, we achieved noticeable improvement. Although we were able to increase accuracy scores significantly (1), we did not reach clinically relevant results, yet. In the future, systematic and structured data collection will be of utmost importance for the improvement of DL applications.

4.1 Conclusion

Transfer learning has proven to be a powerful technique for supporting image interpretation tasks. Even for very limited data sets, the impact can be significant. However, transfer learning is not an overall solution for small data sets. To achieve clinically relevant results, structured and systematic data collection 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.

References

[1] Aiello, M., C. Cavaliere, A. D'Albore and M. Salvatore (2019). "The challenges of diagnostic imaging in the era of big data." Journal of clinical medicine 8(3): 316.

[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] Banerjee, I., A. Crawley, M. Bhethanabotla, H. E. Daldrup-Link and D. L. Rubin (2018). "Transfer learning on fused multiparametric MR images for classifying histopathological subtypes of rhabdomyosarcoma." Comput Med Imaging Graph 65: 167-175.

[4] Bozinovski, S. and A. Fulgosi (1976). The influence of pattern similarity and transfer of learning upon training of a base perceptron B2.(original in Croatian: Utjecaj slicnosti likova i transfera ucenja na obucavanje baznog perceptrona B2). Proc. Symp. Informatica.

[5] Caron, M., P. Bojanowski, A. Joulin and M. Douze (2018). Deep clustering for unsupervised learning of visual features. Proceedings of the European conference on computer vision (ECCV).

[6] Chhikara, P., P. Singh, P. Gupta and T. Bhatia (2020). Deep convolutional neural network with transfer learning for detecting pneumonia on chest X-rays. Advances in Bioinformatics, Multimedia, and Electronics Circuits and Signals, Springer: 155-168.

[7] Das, N. N., N. Kumar, M. Kaur, V. Kumar and D. Singh (2020). "Automated deep transfer learning-based approach for detection of COVID-19 infection in chest X-rays." Irbm.

[8] Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei (2009). Imagenet: A large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition, Ieee.

[9] Ghesu, F. C., B. Georgescu, A. Mansoor, Y. Yoo, D. Neumann, P. Patel, et al. (2022). "Self-supervised Learning from 100 Million Medical Images." arXiv preprint arXiv:2201.01283.

[10] Nguyen, X.-B., G. S. Lee, S. H. Kim and H. J. Yang (2020). "Self-supervised learning based on spatial awareness for medical image analysis." IEEE Access 8: 162973-162981.

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

[12] Willemink, M. J., W. A. Koszek, C. Hardell, J. Wu, D. Fleischmann, H. Harvey, et al. (2020). "Preparing medical imaging data for machine learning." Radiology 295(1): 4-15.

[13] He, K., X. Zhang, S. Ren and J. Sun (2016). "Deep residual learning for image recognition". Proceedings of the IEEE conference on computer vision and pattern recognition.

[14] Cao, B., H. Zhang, N. Wang, X. Gao and D. Shen (2020). Auto-GAN: self-supervised collaborative learning for medical image synthesis. Proceedings of the AAAI Conference on Artificial Intelligence.

[15] 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.

[16] Gao, Y., V. Ghodrati, A. Kalbasi, J. Fu, D. Ruan, M. Cao, et al. (2021). "Prediction of soft tissue sarcoma response to radiotherapy using longitudinal diffusion MRI and a deep neural network with generative adversarial network-based data augmentation." Med Phys 48(6): 3262-3372.

[17] Picci, P., M. Manfrini, D. 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: Clinical, Radiological and Histological Correlations - The Rizzoli Case Archive.

[18] Clark, M. A. and J. M. Thomas (2005). "Delay in referral to a specialist soft-tissue sarcoma unit." Eur J Surg Oncol 31(4): 443-448.