Benchmarking CNNs and Transformers for Knee Cartilage MRI Segmentation Without Pre-Training

James Battye, Ping Lu, Philip Conaghan, Samuel Relton

UKRI Centre for Doctoral Training in Artificial Intelligence for Medical Diagnosis and Care, University of Leeds

scjb@leeds.ac.uk

UK Research and Innovation

UNIVERSITY OF LEEDS

Introduction

- Segmentation of knee joint structures from MRI supports assessing osteoarthritis biomarkers
- CNN-based segmentation models have had significant success, but recently transformers have demonstrated SOTA performance. However, they often require significant pre-training which is resource intensive

This study compared CNN and transformer models trained under purely supervised conditions when segmenting knee cartilage from MRI

Methods

Data

OAI subset of 176 DESS MRI scans from 88 patients (baseline and after one year) [1,2]
Scanner: Siemens 3T Trio
Size: 384 x 384 x 160 volumes
Resolution: 0.36 x 0.36mm x 0.70mm

Models

CNN: nnU-Net [3]
Transformer: Segformer3D, SwinUNETR-V2 [4,5]

Evaluation Metrics

Volume: Dice similarity coefficient (DSC), volumetric overlap error (VOE)
Surface-distance: Hausdorff distance (HD95), average symmetric surface distance (ASSD)
Clinical: Average cartilage thickness error

Model Training

Models trained on 3 NVIDIA L40S 48GB GPUs. nnU-Net training followed the original protocol with an ensemble of 2D and 3D models, and transformer-based model training protocol is shown in Figure 1

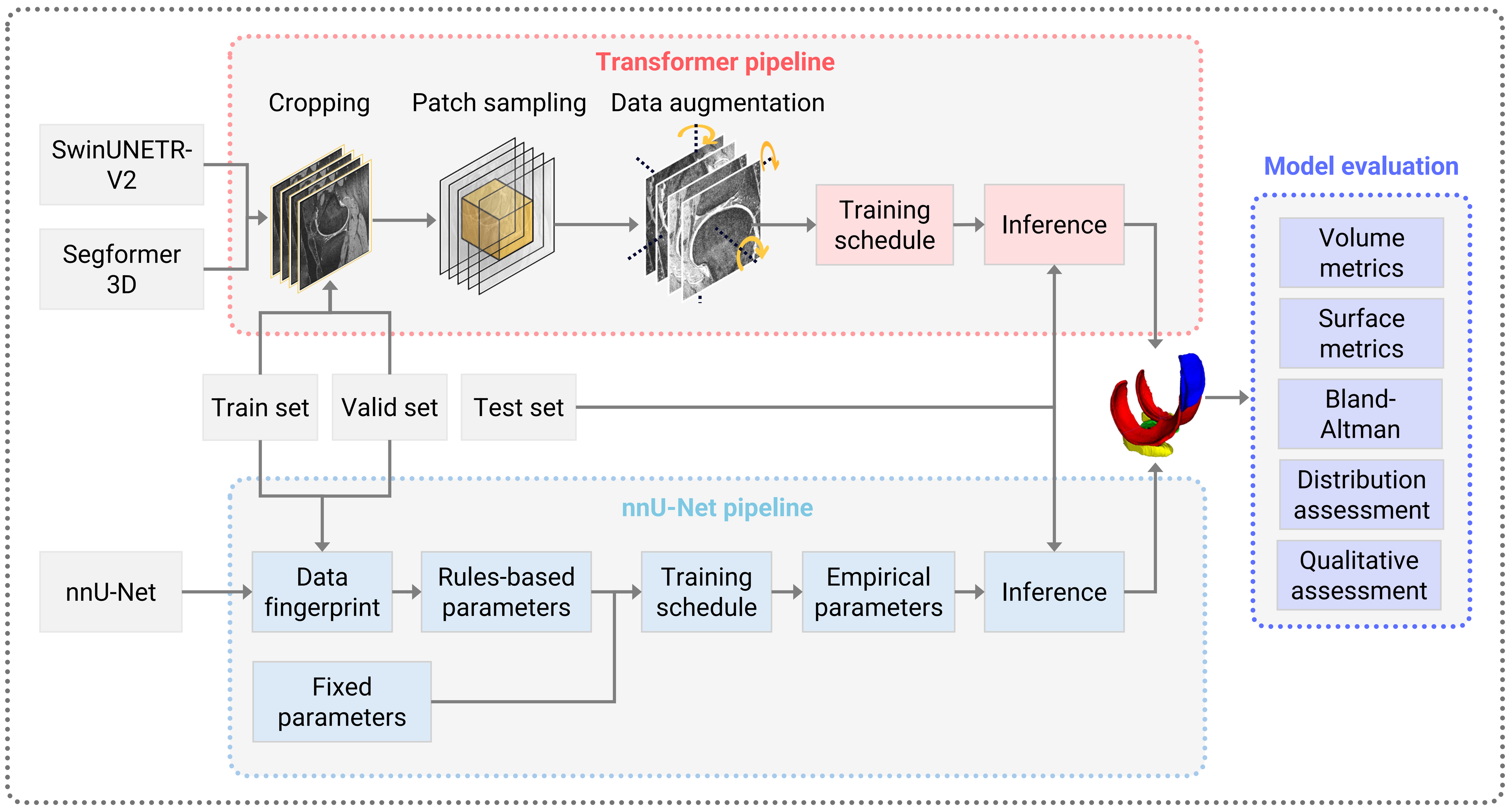

Figure 1: Model training, inference, and evaluation protocols

Results

Table 1: Mean segmentation performance for each model (best performance in bold). TE for the meniscus and HD for all tissues were not recorded in the IWOAI.

Tissue	Metric	IWOAI Best	nnU-Net	Segformer 3D	Swin-UNETR-V2
Femoral cart.	DSC	0.90±0.02	**0.91±0.02**	0.81±0.04	0.90±0.02
	VOE	0.17±0.03	**0.16±0.03**	0.31±0.05	0.19±0.03
	ASSD	0.20±0.03	**0.18±0.03**	0.47±0.11	0.24±0.04
	TE	**0.04±0.03**	0.05±0.03	0.15±0.09	0.05±0.04
	HD		**0.78±0.21**	2.02±0.96	0.94±0.27
Tibial cart.	DSC	0.89±0.03	**0.90±0.03**	0.81±0.06	0.88±0.03
	VOE	0.20±0.04	**0.18±0.04**	0.31±0.08	0.21±0.05
	ASSD	0.26±0.15	**0.25±0.16**	0.44±0.22	0.29±0.18
	TE	**0.04±0.03**	0.05±0.04	0.13±0.10	0.05±0.04
	HD		**1.94±2.36**	2.60±2.42	2.02±2.41
Patellar cart.	DSC	0.86±0.07	**0.87±0.08**	0.73±0.09	0.85±0.07
	VOE	0.24±0.09	**0.22±0.10**	0.42±0.10	0.25±0.10
	ASSD	0.26±0.08	**0.23±0.13**	0.72±0.30	0.30±0.24
	TE	**0.06±0.05**	0.08±0.05	0.26±0.26	0.10±0.07
	HD		**1.11±0.79**	4.27±4.98	1.56±2.01
Meniscus	DSC	0.88±0.03	**0.89±0.03**	0.80±0.03	0.87±0.03
	VOE	0.22±0.04	**0.20±0.04**	0.34±0.05	0.23±0.04
	ASSD	0.33±0.08	**0.30±0.09**	0.62±0.15	0.36±0.08
	TE		**0.09±0.07**	0.12±0.11	0.10±0.08
	HD		**1.53±0.60**	2.23±0.80	1.68±0.61

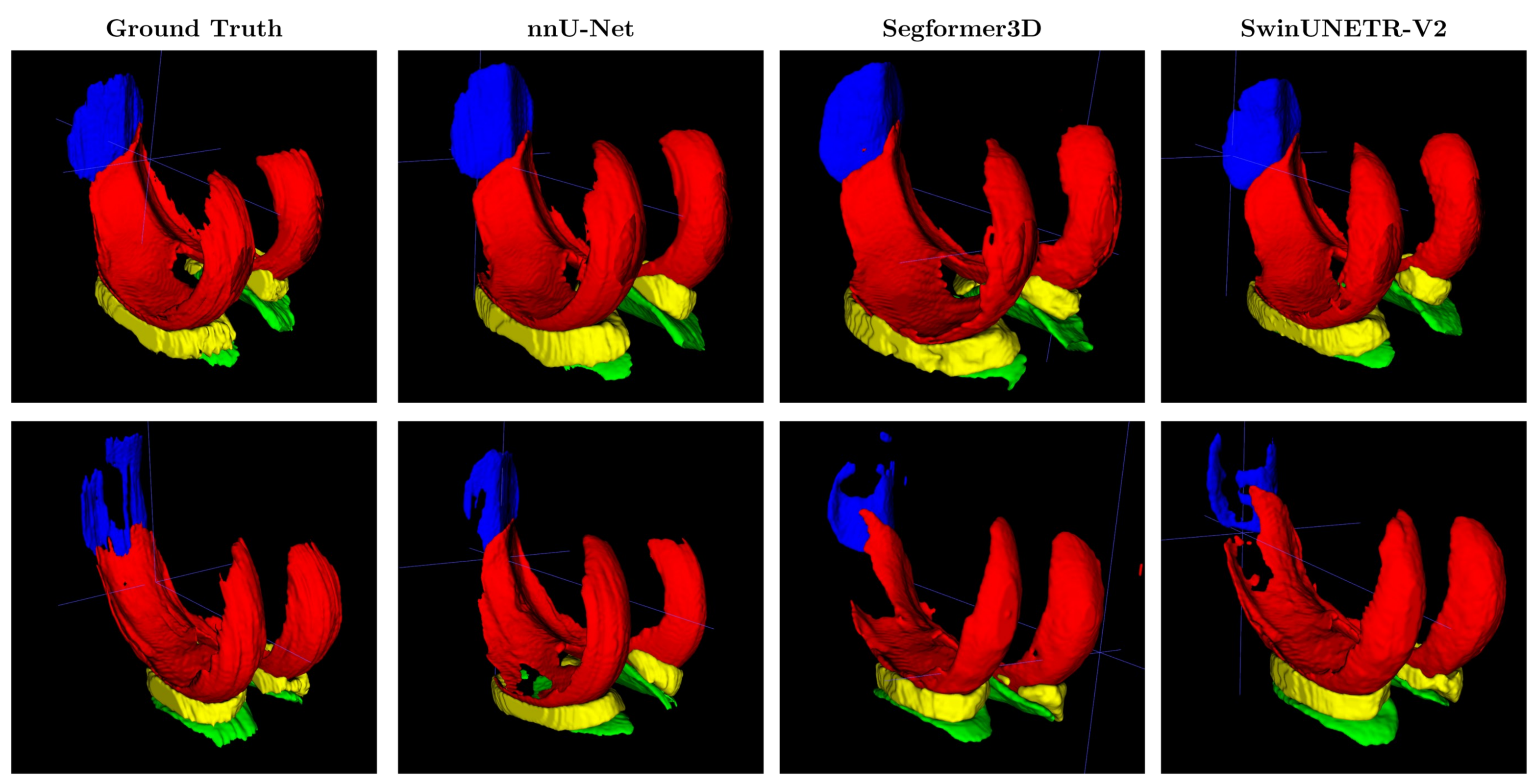

Figure 2: Surface meshes of the ground truth and predicted masks of the best (top row) and worst-performing (bottom row) test samples as measured by Dice score. The 3D masks show the femoral (red), tibial (green), and patellar (blue) cartilage, and the meniscus (yellow).

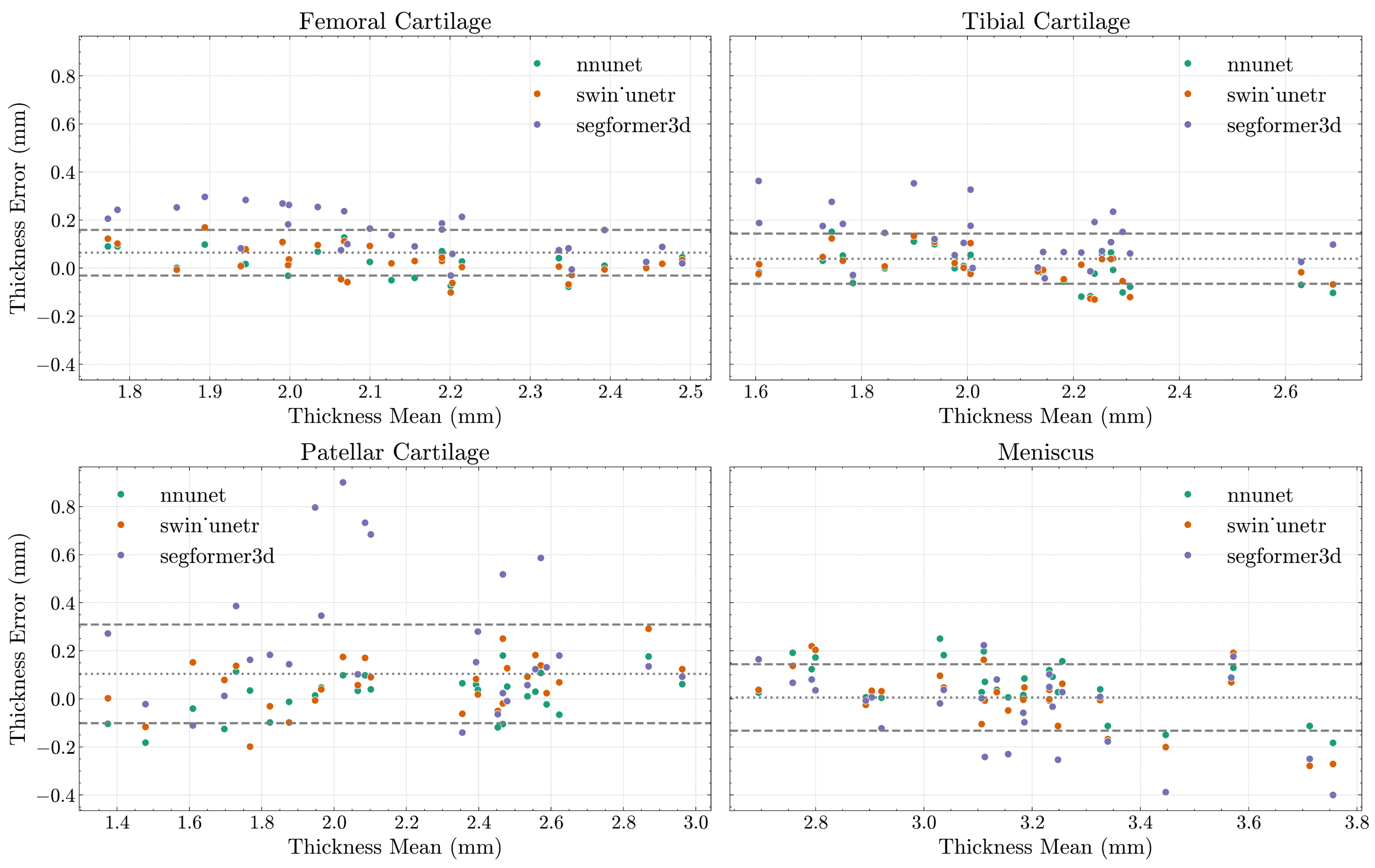

Figure 3: Bland Altman plots for nnU-Net (green), SwinUNETR-V2 (orange) and Segformer3D (purple) showing TE against the ground truth average thickness.

Discussion, Conclusion and Further Work

Discussion

- nnU-Net outperformed transformer-based models and all IWOAI entrants, achieving SOTA accuracy for the dataset, but the performance difference was small
- All models performed worse when tissue was significantly degenerate

Conclusion and Further Work

- In resource-limited settings without pre-training, a well-designed CNNs remains the most effective choice for knee cartilage segmentation from MRI
- Test models on full OAI cohort (4,796 participants), linking clinical information to segmentation mask morphology

References

[1] Branch, NIAMS Science Communications and Outreach. 'Osteoarthritis Initiative'. *National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIAMS*
[2] Desai, Arjun D., et al. 'The International Workshop on Osteoarthritis Imaging Knee MRI Segmentation Challenge: A Multi-Institute Evaluation and Analysis Framework on a Standardized Dataset', *Radiology: Artificial Intelligence*
[3] Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 2021 Feb;18(2):203–11.
[4] Perera S, Navard P, Yilmaz A. SegFormer3D: an Efficient Transformer for 3D Medical Image Segmentation [Internet]. arXiv; 2024 [cited 2025 Apr 17]. Available from: http://arxiv.org/abs/2404.10156
[5] He Y, Nath V, Yang D, Tang Y, Myronenko A, Xu D. SwinUNETR-V2: Stronger Swin Transformers with Stagewise Convolutions for 3D Medical Image Segmentation. In: Greenspan H, Madabhushi A, Mousavi P, Salcudean S, Duncan J, Syeda-Mahmood T, et al., editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. Cham: Springer Nature Switzerland; 2023. p. 416–26.

Acronyms

MRI = Magnetic Resonance Imaging
SOTA = State of the Art
OAI = Osteoarthritis Initiative
DESS = *Double Echo Steady State*

Acknowledgements

This work was funded by the UKRI Centre for Doctoral Training in Artificial Intelligence for Medical Diagnosis and Care at the University of Leeds [grant number EP/S024336/1/] and undertaken on the Aire HPC system at the University of Leeds, UK. The OAI dataset is publicly available. Ethical approval and informed consent were obtained by OAI for all participants.