Toward Precision Knee MRI Segmentation for Bone and Cartilaginous Tumor Interventions

Faisal Al-Qawasmi[1], Layth Alkhani[2], Hossam Zaki[3], Yusef Qazi[2], Wali Badar[4], Osman Ahmed[5]

University of Illinois College of Medicine-Peoria,[1] Stanford University[2], Warren Alpert Medical School of Brown University[3], Department of Radiology, University of Illinois Chicago[4], Joint & Vascular Institute[5]

Purpose

- Segmentation of musculoskeletal structures is essential for tumor characterization, intervention planning, and treatment monitoring in interventional oncology.
- Manual segmentation is time-consuming, variable across operators, and limits reproducibility.
- Deep learning–based segmentation offers potential to automate workflows and enhance precision.
- This study evaluates three advanced architectures:
  - SwinUNETR (transformer-based)
  - SegResNet (convolutional)
  - DiNTS (neural architecture search)
- Focused on knee MRI structures: patella, articular cartilage, and medial/lateral menisci.
- Goal: Identify models best suited for clinical integration in bone and cartilaginous tumor interventions.

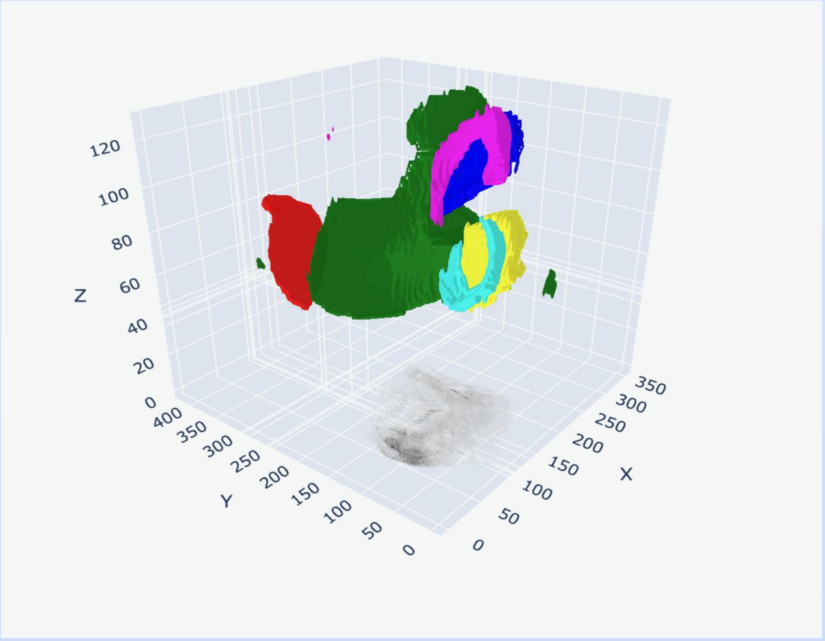

Figure 1. Example 3D segmentation of an MRI scan, illustrating the spatial delineation of anatomical structures produced by the model

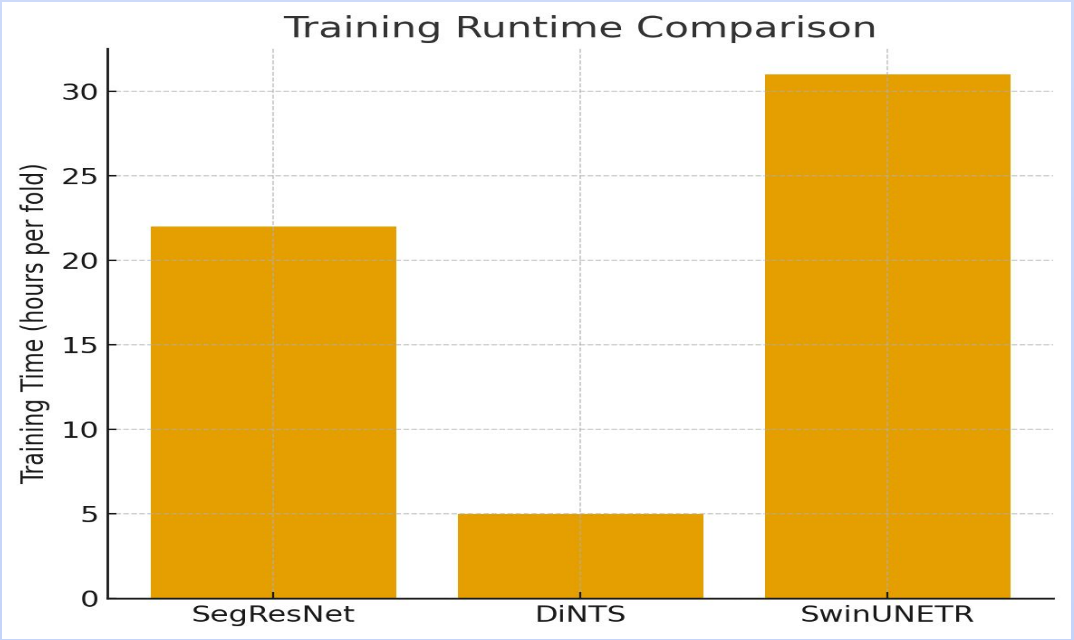

Figure 2. Training runtime (in hours per fold) comparison across different models (SegResNet, DINTS, and SwinUNETR)

Methods

- Dataset: 155 T2-weighted dESS MRI volumes with manual segmentations from the SKM-TEA dataset.
- Preprocessing: Images were standardized and normalized prior to training.
- Splitting: Data randomly divided into 90% training and 10% testing.
- Models Tested:
  1. SwinUNETR: Vision Transformer–based U-Net hybrid with strong context modeling.
  2. SegResNet: Deep convolutional encoder-decoder with residual connections.
  3. DiNTS: Architecture discovered automatically using differentiable search.
- Training:
  1. Framework: MONAI Auto3DSeg
  2. Cross-validation: 5-fold
  3. Metrics: Dice coefficients (weighted by structure size and unweighted across all masks).
- Evaluation Criteria:
  1. Segmentation accuracy (Dice coefficient for each structure).
  2. Runtime efficiency (training time per fold).
  3. Qualitative performance in capturing fine anatomical details (cartilage borders, meniscal horns).

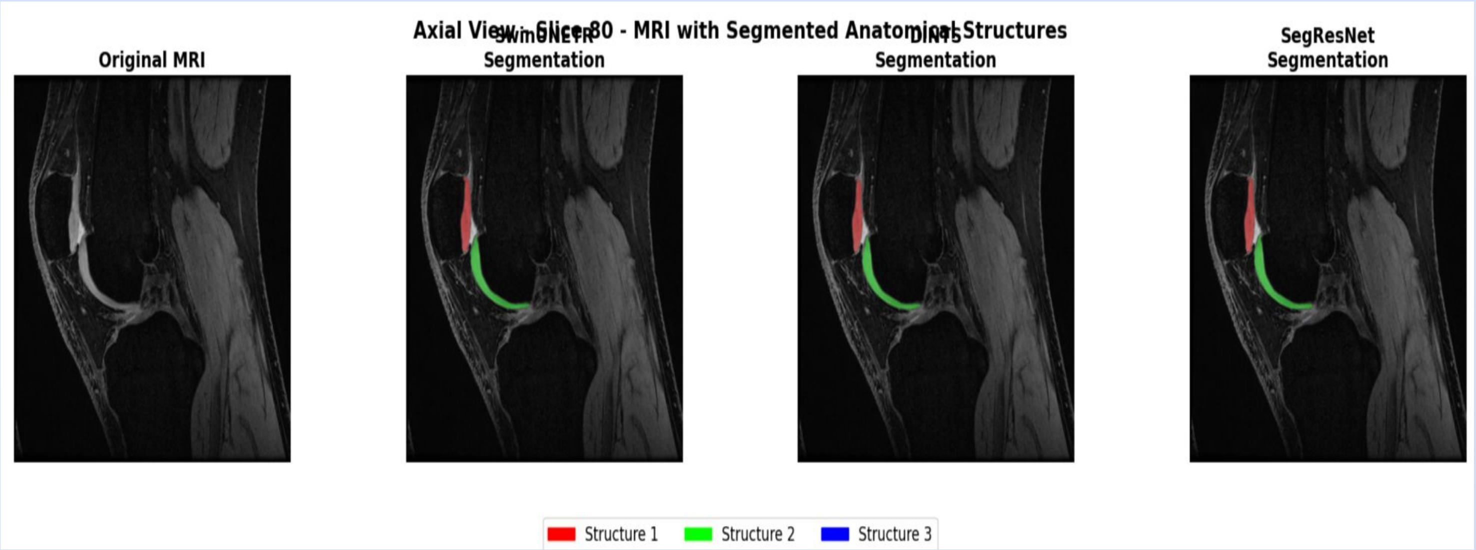

Figure 3. Example of model-based MRI segmentation. The same input MRI is shown with overlaid segmentations from different models, demonstrating how each model delineates anatomical structures

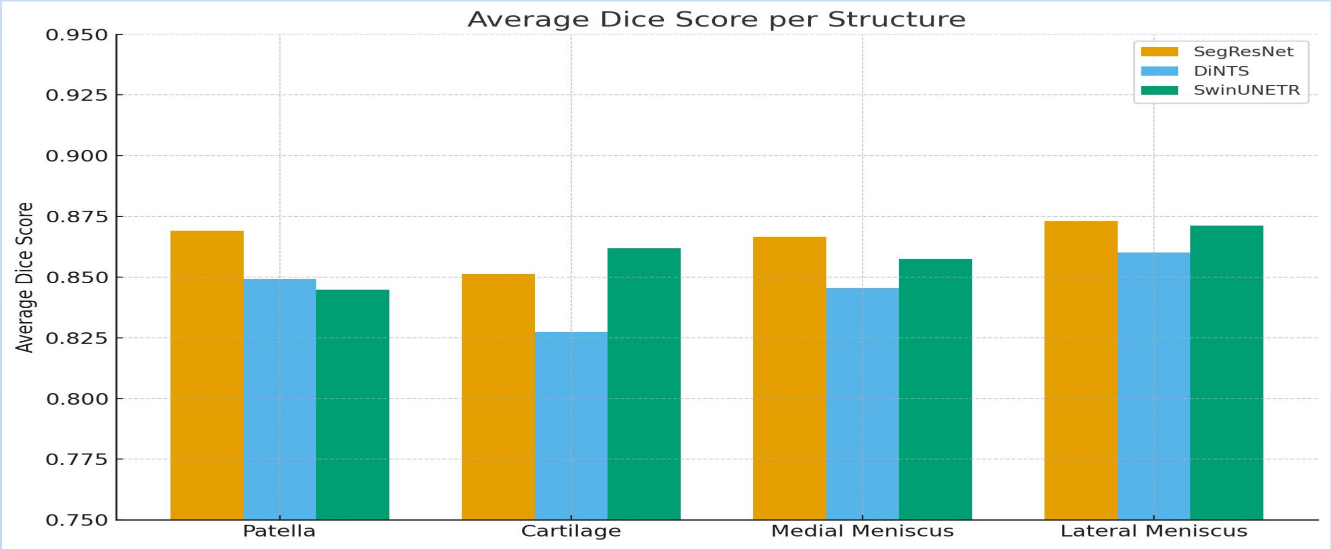

Figure 4. Model performance by anatomical structure. Average Dice similarity scores are shown for each structure (patella, cartilage, medial meniscus, lateral meniscus) across models (SegResNet, DINTS, and SwinUNETR)

Results

Overall Accuracy

- SegResNet: Best average performance (UW Dice 0.8653, W Dice 0.8731).
- SwinUNETR: Nearly equivalent overall (UW Dice 0.8588, W Dice 0.8683).
- DiNTS: Lower accuracy but consistent across structures (UW Dice 0.8456, W Dice 0.8515).

Structure-Specific Performance

- SwinUNETR: Superior in patella (0.9047) and articular cartilage (0.8889), highlighting transformer models' ability to integrate global context for thin, curved structures.
- SegResNet: Outperformed others in lateral meniscus (0.885 / 0.852) and medial meniscus (0.874 / 0.858), where CNNs better captured localized features.
- DiNTS: Did not achieve the highest score for any structure, but produced balanced, moderate segmentationsacross all masks.

Runtime

- SwinUNETR: 31 hours per fold.
- SegResNet: 22 hours per fold.
- DiNTS: 5 hours per fold (fastest, though with trade-off in accuracy).

Visual Evaluation

- SegResNet produced sharper meniscal boundaries.
- SwinUNETR captured continuous cartilage surfaces with fewer gaps.
- DiNTS occasionally underestimated cartilage thickness but provided stable patella segmentation.

Conclusions

- SwinUNETR was most effective for patella and cartilage, highlighting transformer strengths in capturing global context.
- SegResNet achieved the highest Dice for medial and lateral menisci, reflecting CNN advantages in localized features.
- DiNTS delivered balanced but lower accuracy, with the benefit of much faster training times.
- Overall, performance varied by structure, suggesting model choice should depend on clinical need.
- Future directions include fusion approaches and optimizing runtimes for broader clinical adoption.