**Decoder Conditioning with Tabular Data for Enhanced 3D Image Segmentation**Tomasz Szczepański1, Michal K. Grzeszczyk1, Szymon Płotka1, Arleta Adamowicz2, Piotr Fudalej2, Przemysław Korzeniowski1, Tomasz Trzciński3, Arkadiusz Sitek4

1Sano Centre for Computational Medicine, Kraków, Poland
2Jagiellonian University Medical College, Kraków, Poland
3Warsaw University of Technology, Warszawa, Poland
4Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
*t.szczepanski*@*sanoscience.org***Keywords**: Conditioning · Tabular data · Non-Imaging · Segmentation

**1. Introduction**

Recent advancements in deep learning have achieved strong performance in 3D medical image segmentation across various domains. However, these models typically rely on large, expertly annotated datasets, which are expensive to produce, and often fail to generalize across clinical centers and diverse patient populations.

**2. Description of the problem**

While some studies have explored architectural refinements or loss-based improvements, the integration of structured tabular data, such as clinical metadata or label-derived properties, remains underutilized in segmentation. Such data can encode valuable contextual information but is often missing, inconsistent, anonymized, or lacking a direct semantic correspondence to pixel-level labels. Existing approaches for incorporating tabular data are more common in regression tasks, where strong, explicit correlations between covariates and outcomes exist, such as predicting disease progression or birth weight. In segmentation, where labels are high-dimensional and relationships with metadata are indirect, the challenge of effective integration is greater.

**3. Related work**

Methods such as DAFT [1] and TabAttention [2] have shown benefits in regression tasks by exploiting the clear linkage between tabular variables and target measures. In segmentation, some progress has been made, for example, with INSIDE [3], which incorporates limited non-imaging metadata like cardiac phase or anatomical slice position into 2D networks. However, these approaches rely on predefined categorical metadata and operate in constrained settings. Importantly, few works have explored decoder-level conditioning in 3D segmentation networks, and none have proposed a framework that leverages label-derived embeddings during training without requiring such data at inference.

**4. Solution of the problem**

We introduce **DeCode**, a decoder conditioning framework for 3D medical image segmentation that learns to embed shape-derived features from ground-truth labels during training. Radiomics-based shape metrics, such as surface area, compactness, and elongation, are computed using the PyRadiomics library and embedded into a conditioning vector. This vector modulates the decoder through affine transformations applied to internal feature maps, with transformation parameters generated by a hyper-network. Unlike batch normalization, this conditioning operates at the individual sample level, enabling fine-grained control of feature activations. During inference, when labels are unavailable, DeCode predicts the conditioning embedding directly from the encoder’s latent representation, allowing the network to benefit from label-informed structure without manual annotations.
We implement DeCode in a U-Net-style architecture with four encoder and four decoder blocks, inserting conditioning layers after skip connections. In addition to the primary segmentation loss, two auxiliary objectives are used: (1) minimizing the distance between predicted and true embeddings, and (2) regressing predicted radiomics features to their actual values. We evaluate DeCode on two tasks: a synthetic dataset, 3DeCode, inspired by CLEVR-Seg but extended to 3D, and a clinical dataset of cone-beam CT (CBCT) scans of teeth. On synthetic shape-based segmentation tasks, DeCode outperforms unconditioned baselines, achieving over 94% Dice score on challenging mixed-shape-and-size scenarios compared to near-random baseline performance. On CBCT scans, DeCode achieves superior accuracy on external datasets from different centers, approaching the performance of an upper bound that uses ground-truth conditioning features at inference. Ablations confirm that both the conditioning mechanism and auxiliary regression loss significantly contribute to performance, while random embeddings at inference degrade results. The method also requires fewer parameters and less training time than heavier alternatives such as VNet.

**5. Conclusions and future work**

DeCode demonstrates that decoder conditioning with label-derived embeddings can improve 3D segmentation accuracy and generalization without requiring inference-time metadata. This approach expands the toolkit for data-efficient segmentation, making it suitable for scenarios with limited computational resources. Limitations include reliance on predefined radiomics features, which do not capture spatial relationships or object context, and potential training instability on small datasets. Future research could explore end-to-end learning of embeddings directly from masks, integration of richer spatial and relational information, and extensions to multi-organ or temporal segmentation.

**Acknowledgments.** This work is supported by the EU’s Horizon 2020 programme (grant no. 857533, Sano) and the Foundation for Polish Science’s International Re- search Agendas programme (MAB PLUS/2019/13), co-financed by the EU under the European Regional Development Fund and the Polish Ministry of Science and Higher Education (contract no. MEiN/2023/DIR/3796). This research was funded in whole or in part by National Science Centre, Poland 2023/49/N/ST6/01841. For the purpose of Open Access, the author has applied a CC-BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.

**References**

1. Pölsterl, Sebastian, Tom Nuno Wolf, and Christian Wachinger. "Combining 3D image and tabular data via the dynamic affine feature map transform." International conference on medical image computing and computer-assisted intervention. Cham: Springer International Publishing, 2021.
2. Grzeszczyk, Michal K., et al. "Tabattention: Learning attention conditionally on tabular data." International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2023.
3. Jacenków, Grzegorz, et al. "INSIDE: steering spatial attention with non-imaging information in CNNs." International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer International Publishing, 2020.

*This work was originally presented at the 2024 International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI).*