Kidney Tumor Conditional Generation via Latent Diffusion

Salme Ussanov
Supervisors:
Dmytro Fishman (PhD),
Ekaterina Sedykh (MSc)

KEY FINDING: Synthetic kidney tumor CT data can support segmentation training in low-data setting, but does not fully replace real data. The best trade-off was achieved with **modified MAISI + 75% real data.**

1 MOTIVATION

Computed tomography (CT) imaging is essential for diagnosing kidney cancer, but tumor annotation is costly, time-consuming, and requires specialized expertise [1, 2].

This creates a major challenge for deep learning segmentation models, which depend on large annotated datasets that are often limited in medical imaging [3].

Therefore, this work investigates whether synthetic kidney tumor CT images can reduce the need for annotated real data while maintaining competitive segmentation performance.

2 DATA

Dataset	CT images
Training data for generation methods	**850**
Healthy CT images for tumor generation	**160**
Synthetic tumor dataset for segmentation	**160**
Real tumor dataset for segmentation	**160**

3 PREPROCESSING

- **Orientation:** RAS
- **Spacing:** 1 x 1 x 1 mm
- **Volume size:**
  - 512 x 512 x 256 (generation)
  - 512 x 512 x 512 (segmentation)
- **HU values:**
  - [-1000; 1000] (generation)
  - [-160; 240] (segmentation)

4 GENERATION METHODS

MAISI (modified) [4] – adapted for inpainting and kidney tumor generation in compressed image representation (latent space).

2D* – generates kidney tumors slice by slice in image space.

2.5D* – generates kidney tumors in image space slice by slice using information from adjacent slices.

DiffTumor [5] – inpaints tumors in latent space. It was modified to store the synthetic images.

5 SEGMENTATION

nnU-Net [3] was trained on synthetic-only, real-only, and mixed real/synthetic datasets to compare segmentation performance.

Evaluation metrics:

Tumor Dice score (↑)
False positive rate (↓)
Sensitivity (↑)
Specificity (↑)

6 GENERATION PIPELINE (MODIFIED MAISI)

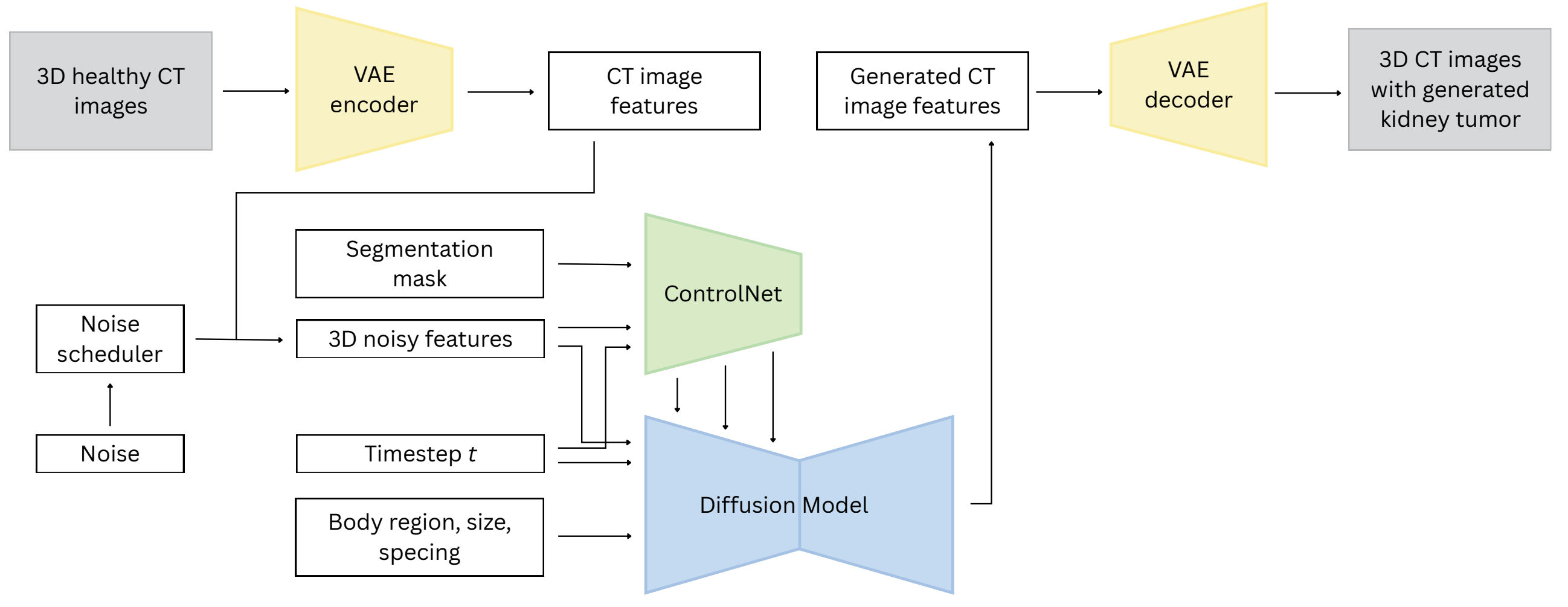

Figure 1. Kidney tumor generation pipeline using our modified MAISI [4] framework.

7 RESULTS - VISUAL ASSESSMENT

A. Fine-tuning on a single scan

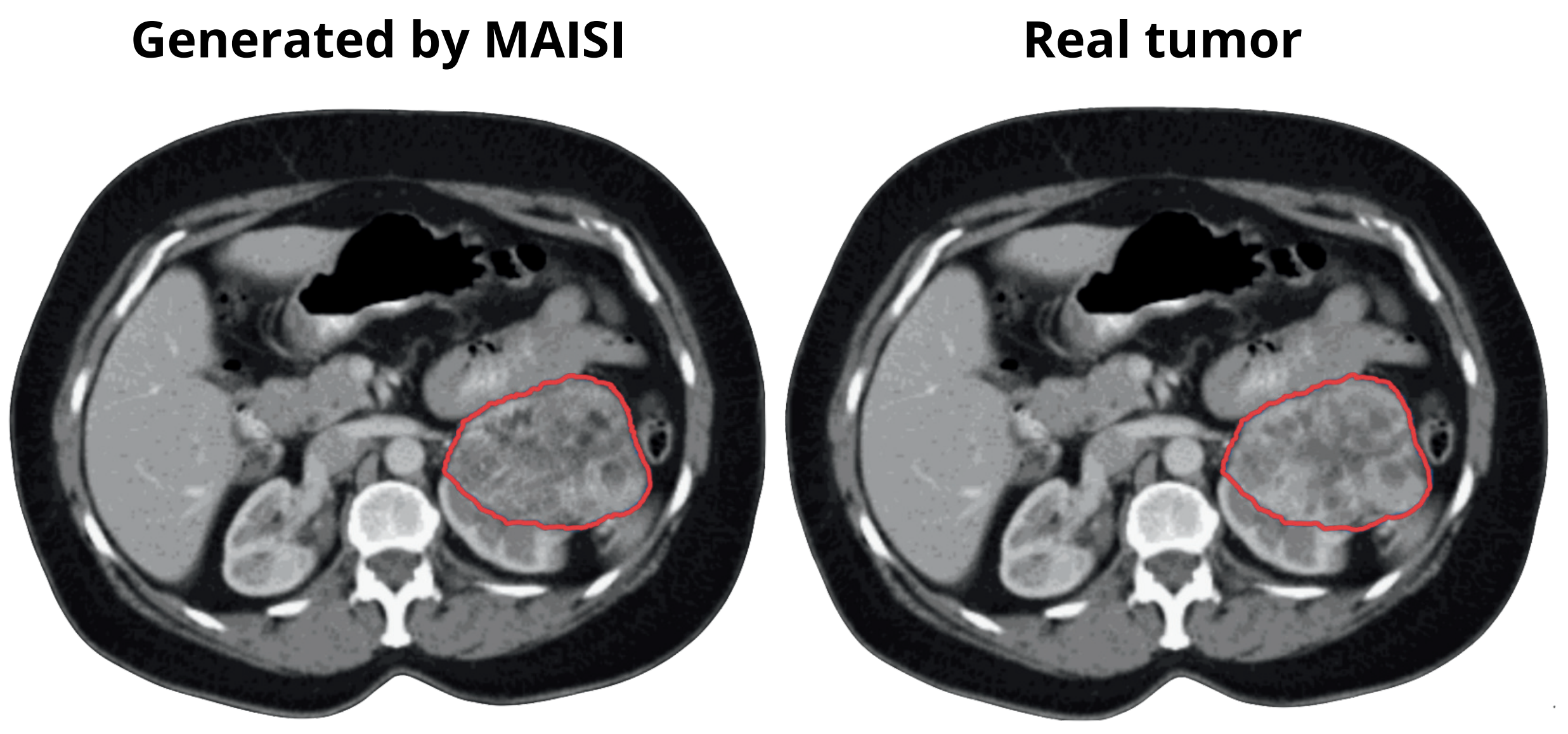

Figure 2. During fine-tuning on a single scan for 10 000 epochs, the modified MAISI framework learns to generate a tumor with appearance close to the real one. The tumor region is outlined in red.

During fine-tuning, we evaluated whether the model could learn to reproduce a specific tumor. As shown in Figure 2, after fine-tuning the modified MAISI framework on a single scan for 10 000 epochs, the model generated a tumor with an appearance similar to the real one.

B. Visual comparison across methods

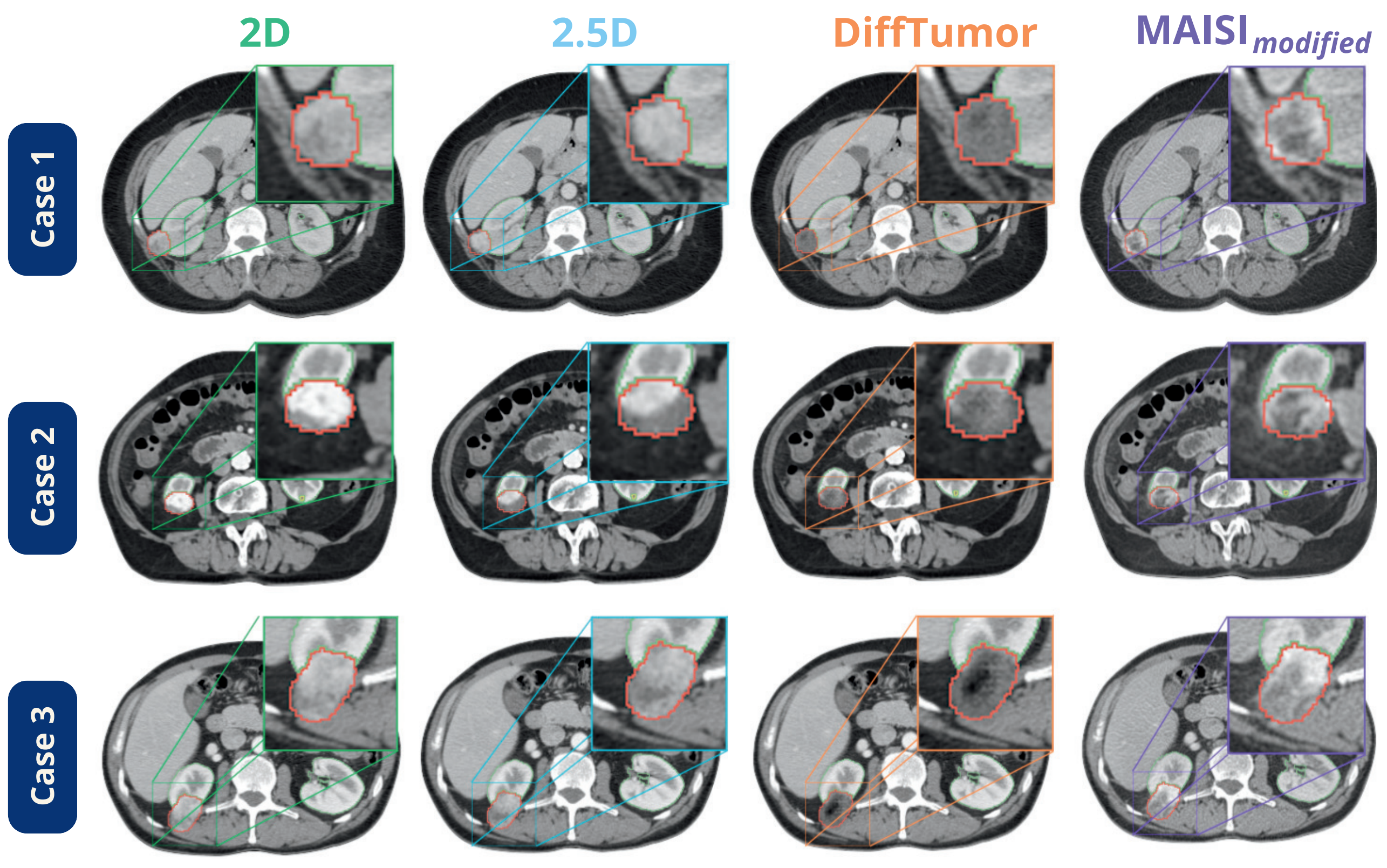

Figure 3. Visual comparison of synthetic tumor generation methods across three cases. Zoomed regions highlight tumor appearance.

Figure 3 shows that all methods generated tumors within the provided masks, but visual quality varied. DiffTumor mainly produced overly dark tumors, while the 2D method occasionally generated overly bright regions. The 2.5D method and modified MAISI produced the most realistic tumors, with more plausible texture, contrast, and intensity. However, faded tumor boundaries were occasionally observed across all methods.

Best visual quality: Modified MAISI produced the most realistic and consistent tumors across cases.

8 RESULTS - QUANTITATIVE

Quantitative results showed that the modified MAISI was the strongest overall method across the evaluated metrics and dataset settings. It achieved the highest Dice score in 4 out of 7 dataset settings and showed the largest advantage in the synthetic-only and 5% real-data settings. This suggests that MAISI-generated synthetic data is especially useful when annotated real data is limited.

Across all methods, segmentation performance generally improved as more real data was added, confirming that synthetic data alone cannot fully replace real training data. However, adding more real data did not always improve all metrics, as false positive predictions often increased.

Table 1. Summary of segmentation performance across different training datasets and methods. The reported values are averaged across the three folds. The baseline corresponds to the model trained using only real data.

Method	Synth only	Synth +5% real	Synth +10% real	Synth +25% real	Synth +50% real	Synth +75% real	Synth +100% real	Real only baseline
Tumor Dice score (higher is better)								**0.86**
2D	0.02	0.51	0.67	0.74	0.83	0.83	0.83	
2.5D	0.06	0.60	**0.73**	**0.81**	0.83	0.85	0.86	
DiffTumor	0.03	0.56	0.69	**0.81**	0.83	0.84	**0.87**	
MAISI*modified*	**0.25**	**0.61**	0.68	0.80	**0.86**	**0.87**	0.85	
False positive rate (%) (lower is better)								**37.2**
2D	**9.80**	61.2	35.0	38.8	**27.3**	32.2	**38.3**	
2.5D	26.2	30.1	36.6	36.1	43.7	32.2	43.7	
DiffTumor	21.3	38.8	40.4	39.9	38.8	35.0	42.1	
MAISI*modified*	10.9	**16.9**	**20.8**	**25.7**	**27.3**	**28.4**	47.0	
Sensitivity (higher is better)								**0.95**
2D	0.01	0.61	0.76	0.84	0.91	0.91	0.90	
2.5D	0.06	**0.71**	**0.83**	0.90	**0.92**	0.94	**0.95**	
DiffTumor	0.02	0.62	0.80	**0.92**	**0.92**	0.92	0.94	
MAISI*modified*	**0.26**	0.70	0.80	0.89	**0.92**	**0.95**	**0.95**	
Specificity (higher is better)								**0.62**
2D	**0.92**	0.43	0.73	0.71	**0.79**	0.71	**0.66**	
2.5D	0.72	0.68	0.73	0.67	0.60	0.69	0.58	
DiffTumor	0.81	0.70	0.69	0.67	0.66	0.70	0.61	
MAISI *modified*	**0.92**	**0.86**	**0.80**	**0.74**	0.74	**0.77**	0.57	

The modified MAISI achieved the best overall balance between Dice score, sensitivity, specificity, and false positive rate. The strongest trade-off was observed when MAISI-generated synthetic data was combined with 75% real data, providing competitive segmentation performance while reducing false positive predictions.

Best trade-off: Modified MAISI-generated synthetic data combined with 75% of the provided real dataset achieved the strongest balance between Dice score and false-positive robustness.

9 CONCLUSION

Synthetic kidney tumor CT data **can reduce the need for annotated real data** in low-data settings, but it cannot fully replace real training data.

The **modified MAISI** framework achieved the best overall results among the evaluated methods. **The strongest trade-off was obtained when MAISI-generated synthetic data was combined with 75% real data**, reaching performance comparable to the real-only baseline while reducing false positive predictions.

REFERENCES

[1] Shadekul Islam S. K. M., Nasim M. A., Hossain I., Ullah M. A., Gupta K. D., andBhuiyan M. M. H. Introduction of Medical Imaging Modalities. arXiv (June 2023). DOI: 10.48550/arXiv.2306.01022.
[2] McDonald R., Schwartz K., Eckel L., Diehn F., Hunt C., Bartholmai B., Erickson B., and Kallmes D. The Effects of Changes in Utilization and Technological Advancements Of Cross-Sectional Imaging on Radiologist Workload. Academic radiology 22 (July 2015). DOI: 10.1016/j.acra.2015.05.007.
[3] Isensee F., Jaeger P., Kohl S., Petersen J., and Maier-Hein K. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods 18 (Feb. 2021), pp. 1–9. DOI: 10.1038/s41592-020-01008-z.
[4] Guo P., Zhao C., Yang D., Xu Z., Nath V., Tang Y., Simon B. D., Belue M. J., Harmon S. A., Turkbey B. I., and Xu D. MAISI: Medical AI for Synthetic Imaging. 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) (2024), pp. 4430–4441. https://api.semanticscholar.org/CorpusID:272694143.
[5] Chen Q., Chen X., Song H., Xiong Z., Yuille A., Wei C., and Zhou Z. Towards Generalizable Tumor Synthesis. arXive (June 2024), pp. 11147–11158. DOI: 10.11 09/CVPR52733.2024.01060.

*Developed by the co-supervison of this work