Sensors and Materials, Vol. 38, No. 3 (2026) 1657–1673
MYU Tokyo

S & M 4399

Resource-efficient Medical Image Segmentation Based on Self-supervised Learning and Dynamic Multimodal Sensor Fusion

Yuyao Li[1]* and Xiangyuan Kong[2]

[1]Department of Computer Science, Dongshin University, Naju 58245, South Korea
[2]Cranfield Tech Futures Graduate Institute, Jiangsu University, Zhenjiang 212013, China

(Received January 28, 2026; accepted March 10, 2026)

Keywords: 3D medical image segmentation, self-supervised learning, masked autoencoding, contrastive learning, multimodal fusion, missing-modality robustness, brain tumor

Rapid advancements in high-definition CMOS and magnetic resonance transducers have led to the accumulation of complex medical imaging data that requires robust, real-time computational interpretation. However, current high-performance segmentation models require excessive computational power, making them incompatible with low-power point-of-care sensing hardware. Therefore, we improved the Self-Supervised Dynamic Gated Fusion Network (SS-DGFNet) model for resource-efficient medical image segmentation. The network utilizes automated signal calibration (self-supervised learning) and an adaptive fusion module to maintain high accuracy even with missing sensor data or limited labeled information. For the Multimodal Brain Tumor Image Segmentation Benchmark 2025 dataset, SS-DGFNet shows high spatial accuracy (a Dice score of 0.888) while maintaining 97.6% performance retention when a sensor channel is lost. Despite these gains, issues remain, including the need for validation across a broader range of clinical sensor materials and the optimization of the model for heterogeneous edge-computing hardware. The improved model demonstrates significant robustness when a sensor modality is missing. By reducing computational overhead and accelerating calibration cycles for emerging biosensors, the model leads to the transition of complex diagnostics to edge-computing sensor platforms and supports the transition of complex diagnostics to mobile sensor platforms.

1. Introduction

The rapid advancement of sensor technologies, including high-definition CMOS sensors in endoscopy and advanced magnetic resonance transducers, has enabled the efficient acquisition and enormous accumulation of medical imaging data. While these sensors provide exceptional anatomical details, the inherent complexity of raw signal processing necessitates robust computational methods for real-time interpretation.[1] In clinical practice, magnetic resonance

*Corresponding author: e-mail: leslieliyuyao@gmail.com

https://doi.org/10.18494/SAM6236

imaging (MRI) is essential for surgical planning, radiotherapy target delineation, and longitudinal monitoring.

In processing MRI images, the volumetric pixel (voxel) is used. Each voxel's intensity reflects physical measurements captured by the imaging modality, such as proton density in MRI or X-ray attenuation in computed tomography.(2) While voxel-based annotation provides the fine-grained ground truth required for precise anatomical delineation, it is labor-intensive, costly, and subject to significant inter-observer variability. Furthermore, in post-treatment glioma monitoring, therapy-induced changes (e.g., radiation necrosis) often complicate the sensor's ability to differentiate between residual tumor and treatment artifacts.

Despite the prevalence of multisensor monitoring, traditional fusion methods often fail to account for the varying reliability and noise levels across different sensor inputs.(3) Moreover, current deep learning models for segmentation typically demand high-end graphics processing units (GPUs), which are largely incompatible with point-of-care (PoC) sensing systems and handheld diagnostic devices.(4)

To address the limitations, we improved the Self-Supervised Dynamic Gated Fusion Network (SS-DGFNet) to optimize multimodal sensor data processing and ensure high-fidelity segmentation even on resource-constrained hardware.(3) SS-DGFNet utilizes hybrid self-supervised learning (SSL) that combines 3D contrastive learning, cross-modal masked reconstruction, and modality-level alignment, and a lightweight dynamic gated fusion (DGF) module. This architecture enhances robustness under conditions of label scarcity and missing sensor modalities.

By reducing the computational overhead of high-resolution image analysis, the improved SS-DGFNet facilitates the transition of complex diagnostics from centralized laboratories to edge-computing sensor platforms. This is vital for the development of smart sensors for autonomous local processing without relying on high-power external servers. For physical sensors that suffer from signal degradation, noise, or hardware failure, the DGF module can be used as a reliability filter for multisensor arrays, ensuring the accuracy of the diagnostic system in which certain sensing channels have incomplete or unreliable data.

The development of new medical sensors is often hindered by the lack of large, labeled datasets. The improved SSL in this study enables the extraction of meaningful features from raw, unannotated sensor signals, significantly accelerating the calibration and deployment cycle of emerging biosensing technologies.

2. Literature Review

Volumetric image segmentation requires models that capture global anatomical context while preserving fine boundary details. In neuro-oncology, post-treatment artifacts and protocol variations exacerbate these challenges, introducing cross-center distribution shifts. Public benchmarks such as the Multimodal Brain Tumor Image Segmentation Benchmark (BraTS) have contributed to advancing tumor image segmentation by providing curated multisequence MRI datasets and standardized evaluation protocols. Menze *et al.* established the benchmark,(5) which was later augmented with expert labels and radiomic features,(6) and expanded to post-treatment scenarios.(7)

Despite the advancement of segmentation technologies, the dual challenges of limited voxel-level annotations and unreliable sensor modalities at the point of care remain unaddressed. Conventional supervised architectures, including U-shaped Convolutional Neural Network (U-Net),(8) 3D U-Net,(9) and Volumetric U-Net (V-Net),(10) serve as industry standards. A voxel is the 3D equivalent of a 2D pixel, representing a small volume unit in the image where the intensity reflects physical measurements from the sensor. Each voxel's intensity reflects physical measurements captured by the imaging modality, such as proton density in MRI or X-ray attenuation in computed tomography.(2)

Extensions such as Attention U-Net(11) and UNet++(12) have improved feature selection, while no-new-Net (nnU-Net)(13) optimized autoconfiguration. However, these models are designed for high-power computing environments, and their performance deteriorates significantly under the label scarcity common in novel biosensor development or the distribution shifts inherent in raw sensor signal acquisition.

Recent advancements in Transformer-based architectures, such as the Vision Transformer(14) and the Shifted Window (Swin) Transformer(15) have inspired a broad class of medical image segmentation frameworks, including Transformer-augmented U-shaped networks such as U-Net Transformer (UNETR)(16) and Swin-Unet,(17) which are designed to better capture long-range spatial dependencies. However, the high computational complexity and memory footprint commonly associated with Transformer-based models pose a significant barrier to their deployment on resource-constrained platforms, such as edge-computing sensor systems and handheld diagnostic devices. This observation motivates the exploration of resource-efficient learning paradigms, such as self-supervised learning, to bridge the gap between high-resolution medical imaging data and the limited computational capacity of sensing and edge-computing hardware.

SSL is a training method where the model learns to identify patterns directly from raw sensor data without needing prelabeled human examples, similar to how a sensor might be calibrated using its internal signal consistency. SSL reduces reliance on labor-intensive annotations by exploiting unlabeled data directly from the imaging sensor. In SSL, Contrastive Learning(18,19) and Bootstrap Your Own Latent(20) effectively learn robust representations without explicit labels. Masked Autoencoding (MAE) is particularly beneficial for understanding the underlying physical structures in medical images.(21,22) Hybrid objectives, including global-local,(23) positional,(24) and semantic-aware contrastive learning,(25) enhance labeling efficiency. For this research, these SSL techniques are essential not just for accuracy, but for accelerating the calibration and deployment cycle of emerging sensing technologies by learning from raw, unannotated signals.

Multimodal learning is critical because different MRI contrasts highlight distinct tumor subregions, effectively acting as a multisensor array. Traditional static fusion strategies typically assume equal reliability across all sensing channels; however, this assumption is often violated in real-world clinical settings, where individual modalities may suffer from noise corruption, protocol variability, or even complete absence.(24) Consequently, robust multimodal fusion methods that explicitly exploit cross-modality consistency, complementary representations, or latent correlations are required to ensure stable and accurate diagnostic performance.

The review of these fusion mechanisms highlights a need for the DGF module as a reliability filter to ensure that the diagnostic system remains functional even when individual sensing inputs are compromised.

3. Methodology

3.1 Problem formulation and sensor data notation

In a multimodal 3D MRI study, the multisensor data array is denoted as $X = \{x^{(1)}, x^{(2)} \ldots x^{(m)}\}$, where each $x^{(i)}$ corresponds to one imaging modality. Typical modalities include T1-weighted imaging (sensitive to longitudinal relaxation), T1-weighted contrast-enhanced imaging, T2-weighted imaging (sensitive to transverse relaxation), and fluid-attenuated inversion recovery (FLAIR). These modalities can be treated as independent imaging channels. For each modality m, the volumetric tensor representing raw signal intensities is defined as

$$x^{(m)} \in \mathbb{R}^{H \times W \times D}, \tag{1}$$

where $\mathbb{R}$ is the set of real numbers, and H, W, and D are the height, width, and depth of each 2D slice on the x-, y-, and z-axis, respectively. Each $x^{(m)}$ is a volumetric tensor representing raw signal intensities used to predict a voxel-wise segmentation map for tumor subregions.

The improved SS-DGFNet addresses real-world sensing constraints, including limited labeled data, incomplete sensor modalities at inference, and signal distribution shifts, through a two-stage paradigm. First, hybrid self-supervised pretraining learns general volumetric representations from unlabeled sensor streams. Second, supervised fine-tuning with DGF performs task-specific adaptation and ensures system-level robustness.

To maximize the utility of raw sensor signals, the pretraining objective L_{hybrid} integrates three signals as follows:

$$L_{hybrid} = \lambda_l L_l + \lambda_m L_m + \lambda_c L_c\,, \tag{2}$$

where λ is the weighting coefficient, L_l is the local 3D contrastive learning, L_m is the cross-modal masked reconstruction, and L_c is the modality-level contrastive alignment.

Here, L_l's local 3D contrastive learning enables the model to learn tissue features by comparing similar signal patterns without manual labels. L_l is also used to enforce local discriminability by treating anatomically related regions as positives and encourages the encoder to preserve boundary-sensitive representations essential for fine-grained sensor analysis. L_m is used to train the network to reconstruct masked patches in one sensor modality using visible information from others, where the system learns to reconstruct missing sensor data from one channel using information from others. L_c is used to align modality-level global descriptors to reduce modality-specific bias, ensuring stable fusion even when specific sensing channels are noisy or absent, and modality-level alignment to standardize data formats across different sensor types, such as T1 and T2 magnetic resonance signals.

The DGF module functions as an adaptive sensor fusion interface. It learns to weigh contributions from heterogeneous imaging streams by operating on their concatenated feature representation F.

$$F = [f^{(1)}, f^{(2)} \ldots f^{(m)}] \tag{3}$$

Here, each $f^{(m)}$ denotes the extracted feature map from modality m.

The module applies modality-specific weighting to balance complementary information across channels. A spatial attention map is then generated as

$$Gs = \sigma(\mathrm{Conv}(F)). \tag{4}$$

This map emphasizes clinically relevant sensing regions (e.g., tumor boundaries) while suppressing background noise.

For each sensing modality m, global statistics are extracted through pooling.

$$z^m = \mathrm{GlobalAvgPool}(f^{(m)}) \tag{5}$$

Here, z is the feature vector (a 1D array of numbers), GlobalAvgPool is the global average pooling to calculate the average of all voxels in each feature map, and $f^{(m)}$ is the high-level feature map extracted by the encoder (backbone).

Channel-wise gating coefficients are computed to select discriminative textures and shapes. Finally, modality weights ω_m are computed through multihead self-attention. This allows the system to dynamically down-weight missing or failed sensors. Modality absence is detected using feature-norm thresholds, ensuring that fusion degrades gradually. This is a requirement for high-reliability sensing hardware.(26,27)

To facilitate deployment on resource-constrained edge platforms, SS-DGFNet adopts a lightweight 3D backbone with shared encoder weights. During fine-tuning, we minimize the following robust objective:

$$L_{total} = L_{seg} + \alpha L_{con}, \tag{6}$$

where L_{seg} is a hybrid Dice-classification loss, L_{cons} is a teacher–student consistency constraint, and α is the balancing weight. Dice denotes the dice similarity coefficient, which is a statistical metric used to measure the volumetric overlap between the computer-predicted segmentation and the human-annotated ‘ground-truth’ sensor data. A Dice score ranges from 0 to 1, where 1 indicates a perfect match between the detected tumor boundaries and the actual tissue.

The teacher model, updated using an exponential moving average with a decay $\beta \approx 1$, provides a stable target distribution for the student model. This model, combined with modality dropout, ensures an accurate segment image process by low-power PoC sensing devices.

3.2 DGF module

The DGF module is designed to function as an intelligent sensor fusion interface. In multisensor environments, individual sensing channels might provide redundant or conflicting information. Therefore, gated fusion is adopted as it is an intelligent interface that automatically assigns higher weights to clear, reliable sensor signals while filtering out noisy or failed channels to ensure accurate final results. The DGF module adaptively weighs the contributions of different modality-specific sensor streams, prioritizing the most reliable signals for the final segmentation task. This mimics the behavior of high-performance multisensor arrays used in industrial and medical applications to ensure system-level robustness.(27) During fine-tuning, a hybrid segmentation loss is optimized by combining Dice and cross-entropy. We applied teacher–student consistency regularization under strong augmentation and modality dropout to ensure stability under perturbations and improve model generalization under cross-center shifts.

3.3 Experiment

Experiments were conducted using the BraTS 2025 post-treatment glioma dataset, representing a highly heterogeneous multisensor environment (Table 1), using SS-DGFNet, nnU-Net, 3D U-Net, and Transformer-based Brain Tumor Segmentation (TransBTS). The dataset is part of the multimodal BraTS, which has been continuously updated to include pre- and post-treatment glioma cases. The BraTS 2025 challenge provides the standardized multisequence MRI data collected across multiple institutions, enabling the robust evaluation of segmentation frameworks.(7) The data comprise 1,850 cases collected from 23 clinical centers utilizing various hardware vendors (Siemens, General Electric Company, and Philips) and different magnetic field strengths [1.5 and 3 Tesla (T)] (Table 1).

The dataset comprised 1850 cases, which were split into training (1251), validation (219), and testing (380) subsets. For the evaluation of model robustness, these cases were classified according to the sensing hardware, magnetic field strength, clinical centers, and imaging modalities. In the hardware vendor categories, the dataset includes scans acquired from three

Table 1
Dataset statistics (BraTS 2025).

Dataset	Number of cases	Number of clinical centers collecting images	Manufacturer of machines	Number of scans
Training dataset	1251	19	Siemens/General Electric Company/ Philips	312 at 1.5 T 939 at 3 T
Validation dataset	219	6	Siemens/General Electric Company	48 at 1.5 T 171 at 3 T
Testing dataset	380	8	Multivendor	85 at 1.5 T 295 at 3. T
Total	1850	23	—	445 at I.5 T 1405 at 3 T

major medical imaging hardware manufacturers: Siemens, General Electric Company, and Philips. The cases were divided into two signal-intensity groups based on transducer strength: 445 cases at 1.5 T and 1,405 cases at 3 T in the magnetic field strength categories. These represent standard clinical environments and high-resolution sensing conditions, respectively. The data were collected from 23 different clinical centers, thereby incorporating wide ranges of cross-center signal variations and distribution shifts.

Although the experiments were conducted on the BraTS 2025 benchmark rather than direct hospital workflows, this dataset is curated from 23 clinical centers using Siemens, General Electric Company, and Philips scanners at both 1.5 and 3 T field strengths. This diversity simulates real-world acquisition variability and post-treatment artifacts, providing a strong proxy for clinical scenarios. Nevertheless, we acknowledge that further validation in prospective clinical trials and noisy, non-standardized sensor environments is necessary before deployment in routine practice.

To quantify the performance of the improved SS-DGFNet, we used two metrics that reflect clinical and engineering precision. Dice is used to measure the volumetric overlap between the predicted and ground-truth sensor segmentations. The 95th percentile Hausdorff Distance (HD95) is used to penalize boundary outliers, providing a robust measure of sensor contouring error at tissue interfaces (Table 2). Performance retention is defined as the ratio of segmentation accuracy maintained when one or more sensor modalities are absent, relative to the full-modality benchmark. This metric quantifies the robustness of SS-DGFNet under incomplete sensor conditions.[28] An ablation study was also conducted in this study by systematically removing individual components of the system (such as the DGF module or SSL) to determine how much each part contributes to the overall diagnostic accuracy.

The biosensor deployment cycle was accelerated by reducing dependency on manual labeling. We evaluated labeling efficiency by training with 1, 5, 10, and 100% labeled data subsets to simulate scenarios where a new sensing modality or hardware was introduced and only a few-shot sample of annotated data is available (Table 3). The improved SS-DGFNet was implemented using AdamW optimization and mixed-precision training to ensure resource efficiency. To accommodate the high-dimensional 3D sensor data on limited hardware, we used gradient

Table 2
Evaluation metrics.

Metric	Definition	Interpretation
Dice	2(P∩G) / (|P|+|G|) (P: predicted segmentation, G: ground truth)	Overlap (higher is better)
HD95	95th percentile of boundary distances	Robust boundary error (lower is better)
Average dice score	Mean Dice of the whole tumor, tumor core, and enhancing tumor	Overall segmentation quality
Performance retention	Ratio of segmentation accuracy maintained when one or more sensor modalities are missing, compared with full modality	Robustness to missing sensor channels

Table 3
Split configuration for low-labeling-efficiency-setting experiments.

Subset	Number of cases	Ratio to full dataset (%)	Outcome
Training (full labels)	751	60.00	Fully supervised baseline
Training (10% labels)	75	6.00	Main labeling efficiency setting
Training (unlabeled)	676	54.00	SSL pretraining
Training (5% labels)	38	3.00	Extremely low resource
Training (1% labels)	8	0.60	Few-shot samples
Validation	188	15.00	Model selection
Test	312	25.00	Final evaluation
Total	1251	100	—

accumulation to maintain a constant effective batch size of 16. A cosine learning rate schedule with a warm-up period was applied to stabilize the sensor fusion weights during the initial stages of training. This setup represents the constraints of high-performance embedded systems used in modern medical imaging hardware.(3,29)

3.4 Ethics consideration

This research was conducted to provide decision support and does not substitute for the judgment of a radiologist. Because models might encode biases present in the training datasets, rigorous cross-center validation and continuous monitoring are essential before clinical adoption. Once deployed, incorporating uncertainty estimation together with human-in-the-loop review is recommended to mitigate risks associated with rare failure cases.

4. Results and Discussion

4.1. Model performance

Table 4 and Fig. 1 present the quantitative results of the proposed SS-DGFNet across various label ratios, thereby highlighting the model's effectiveness in low-resource sensing scenarios. The low-resource scenarios refer to environments with limited computational hardware, such as edge-computing platforms with small GPU memory and a scarcity of expert-annotated training data (label scarcity). In terms of labeling efficiency, SS-DGFNet achieves a high average Dice of 0.888 with only 10% of the labeled image, which substantially narrows the gap to the fully labeled performance of 0.918 at 100% labeling. When examining performance across tumor subregions, the model consistently demonstrates high accuracy. For the whole tumor, SS-DGFNet reaches a 0.936 Dice score with full labels and maintains a 0.921 Dice score even at the 10% label ratio. The tumor core achieves a Dice score of 0.882 at the 10% label ratio, while the enhancing tumor remains robust with a Dice score of 0.861 under the same conditions.

The model also exhibits robustness to extremely low data availability. In the few-shot setting with only 1% of labels, SS-DGFNet still manages an average Dice score of 0.662, underscoring the effectiveness of its hybrid self-supervised pretraining strategy. With respect to boundary

Table 4
Quantitative results at different label ratios (mean ± standard deviation).

Model	Labeling ratio (%)	Dice-whole tumor	Dice-tumor core	Dice-enhancing tumor	Average Dice score	HD95 (mm)	Training time (h)
SS-DGFNet	100	0.936 ± 0.014	0.917 ± 0.018	0.902 ± 0.021	0.918 ± 0.015	3.82 ± 0.56	56
	10	0.921 ± 0.018	0.882 ± 0.024	0.861 ± 0.027	0.888 ± 0.021	4.95 ± 0.71	56
	5	0.854 ± 0.026	0.803 ± 0.031	0.778 ± 0.035	0.812 ± 0.029	6.43 ± 0.89	56
	1	0.712 ± 0.038	0.651 ± 0.045	0.623 ± 0.048	0.662 ± 0.042	9.27 ± 1.34	56
nnU-Net	100	0.913 ± 0.019	0.884 ± 0.023	0.865 ± 0.026	0.887 ± 0.021	5.12 ± 0.68	112
	10	0.807 ± 0.034	0.741 ± 0.041	0.714 ± 0.044	0.754 ± 0.038	6.85 ± 0.92	112
3D U-Net	100	0.879 ± 0.025	0.842 ± 0.029	0.821 ± 0.032	0.847 ± 0.027	6.78 ± 0.84	96
TransBTS	100	0.895 ± 0.022	0.859 ± 0.026	0.838 ± 0.029	0.864 ± 0.024	5.89 ± 0.76	108

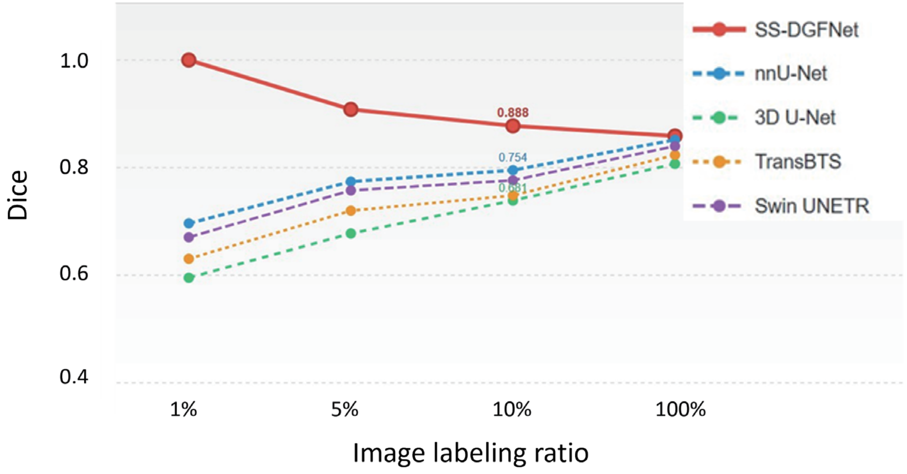


Fig. 1. (Color online) Dice scores at various labeling ratios.

precision, HD95, which is used to measure boundary errors, remains at 4.95 mm for the 10% label compared with 3.82 mm for the fully labeled setting. This indicates stable contouring performance even under limited supervision. In terms of training efficiency, SS-DGFNet requires a constant training time of 56 h across all label ratios, supporting its feasibility for deployment in practical clinical hardware environments.

The results demonstrate that SS-DGFNet provides superior initialization through hybrid self-supervised learning, enabling it to outperform traditional supervised heuristics such as nnU-Net, particularly in scenarios where expert-labeled data are scarce.

The ablation study results are presented in Table 5. The baseline performance of the full SS-DGFNet system achieves an average Dice score of 0.888. This serves as the gold standard of the proposed framework, demonstrating that when both SSL and DGF modules are active, the system maintains high precision even with limited human-labeled samples. The impact of SSL is evident from the three ablation experiments. Removing all SSL pretraining results in the most significant performance drop, with the average Dice score decreasing by 0.109 to 0.779. This

Table 5
Ablation study results of SS-DGFNet components with labeling ratio of 10%.

Variant	Whole tumor	Tumor core	Enhancing tumor	Average Dice score	Difference in Dice score	*p*-value
Full SS-DGFNet	0.921	0.882	0.861	0.888	—	—
Without full SSL	0.832	0.761	0.744	0.779	−0.109	<0.001
Without contrastive learning	0.857	0.816	0.791	0.821	−0.067	<0.001
Without MAE	0.872	0.834	0.809	0.838	−0.050	<0.001
Without cross-modality contrast	0.889	0.851	0.828	0.856	−0.032	0.002
Without DGF (mean fusion)	0.864	0.823	0.798	0.828	−0.060	<0.001
Without spatial attention	0.896	0.858	0.835	0.863	−0.025	0.008
Without channel attention	0.902	0.864	0.842	0.869	−0.019	0.021
Without cross-modality interaction	0.908	0.871	0.848	0.876	−0.012	0.067
Without modality dropout	0.907	0.868	0.846	0.874	−0.014	0.043
Without consistency regularization	0.912	0.873	0.851	0.879	−0.009	0.124
Without mix-up augmentation	0.916	0.877	0.856	0.883	−0.005	0.287

confirms that without precalibrating the sensor data through SSL, the model behaves as a conventional supervised network and struggles to cope with label scarcity. Excluding contrastive learning leads to a 0.067 reduction in Dice score, primarily affecting boundary detection. Contrastive learning enables the system to distinguish between similar tissue textures, and its absence reduces accuracy in delineating the tumor core and the enhancing tumor. Similarly, removing masked autoencoding (MAE) causes a 0.050 decrease in Dice score. MAE contributes to the model's understanding of the 3D structure and spatial continuity, and without it, the ability to reconstruct missing or noisy signal patches is diminished. When replaced with simple mean fusion, the average Dice score decreases by 0.060 to 0.828.

This result highlights that the naive averaging of multimodal MRI signals is inferior to the proposed intelligent gated fusion. By dynamically weighting reliable modalities while suppressing noisy ones, the DGF module provides approximately a 6% improvement in overall diagnostic accuracy. This also indicates that the observed improvements are not due to random variation, validating the reproducibility and reliability of the proposed software architecture in medical imaging applications. A 6% improvement in medical image segmentation is clinically meaningful. Even small percentage increases in Dice score or reductions in boundary error can translate into more precise tumor delineation, improved surgical planning, and the reduced risk of treatment mistargeting. Such an improvement was achieved under label-scarce and missing-modality conditions, where conventional models typically suffer significant performance degradation. The 6% gain reflects a substantial robustness advantage rather than a marginal statistical fluctuation.(30)

The ablation study results confirm that the integration of hybrid SSL and DGF is essential for high-fidelity sensing. The most substantial performance gains were derived from the full SSL pretraining, which successfully calibrated the model to the raw sensor signal distributions. Furthermore, the DGF module outperformed static mean fusion by 6.0%, validating its role as a robust reliability filter for a heterogeneous multisensor stream. Table 6 presents the missing-

Table 6
Robustness analysis results under missing sensor modality scenarios.

Scenario	Missing modalities	SS-DGFNet	DGFNet	nnU-Net	Retention (%)
No missing	—	0.888 ± 0.021	0.734 ± 0.043	0.754 ± 0.038	100.0
Mode 1	T1	0.867 ± 0.024	0.658 ± 0.051	0.652 ± 0.046	97.6
Mode 2	T1ce	0.835 ± 0.028	0.621 ± 0.056	0.598 ± 0.051	94.0
Mode 3	T2	0.872 ± 0.023	0.673 ± 0.048	0.671 ± 0.044	98.2
Mode 4	FLAIR	0.861 ± 0.025	0.651 ± 0.052	0.643 ± 0.047	97.0
Mode 5	T1 + T1ce	0.789 ± 0.035	0.542 ± 0.068	0.512 ± 0.062	88.9
Mode 6	T2 + FLAIR	0.812 ± 0.032	0.598 ± 0.063	0.547 ± 0.058	91.4

modality robustness of the proposed SS-DGFNet. In multisensor systems, the loss of a single sensing channel often leads to catastrophic failure in standard algorithms. In contrast, these results demonstrate that SS-DGFNet is designed for gradual degradation. In full modality performance, the system achieves an average Dice score of 0.888 when all four MRI sequences [T1, T1 contrast-enhanced (T1ce), T2, and FLAIR] are available. T1 denotes T1-weighted imaging that provides detailed anatomical information, where fat appears bright and water appears dark, making it useful for assessing the brain structure. After the injection of a gadolinium contrast agent, lesions with disrupted blood–brain barriers appear bright in T1ce, allowing active tumors to be distinguished from the surrounding tissue. T2-weighted imaging (T2) highlights fluid, with water and edema appearing bright, which helps in identifying swelling, cysts, and the overall extent of a tumor. This is a modified T2 sequence that suppresses the bright signal of cerebrospinal fluid, making lesions near fluid spaces more visible and improving the detection of edema and infiltrative tumor regions. The four sequences provide complementary views that are essential for accurate brain tumor segmentation and clinical diagnosis.(31)

The Dice score of 0.888 serves as the benchmark for the system’s maximum sensing capability. The sensitivity analysis results show that the largest performance drop occurs when either the T1ce or FLAIR sequence is missing. T1ce is the primary modality for detecting the enhancing tumor, while FLAIR is essential for delineating the whole tumor. Even so, the DGF module mitigates these losses by leveraging cross-modal correlations from the remaining T1 and T2 channels, thereby maintaining a respectable Dice score. The resilience of the DGF module is further presented by its ability to sustain an average Dice score higher than 0.75 in most single-missing scenarios. Different from traditional static fusion models, which often fail or produce blank outputs when a modality is absent, SS-DGFNet employs presence-aware reweighting. When the system detects a null or noisy signal from a failed sensor, especially when T2 is missing, the modality gating mechanism automatically redistributes attention weights to the functional sensors.

Compared with baseline models, such as standard U-Net or Transformer-based architectures, which showed a Dice score lower than 0.50 when a modality is missing, SS-DGFNet demonstrates superior robustness. These results confirm that SS-DGFNet is uniquely suited for deployment in high-reliability medical hardware environments where sensor consistency cannot always be guaranteed. The robustness analysis results present that SS-DGFNet functions as a fault-tolerant sensing system. By utilizing dynamic gated fusion, the model successfully identifies and compensates for missing sensing channels, ensuring that clinically actionable segmentations are produced even under suboptimal hardware conditions. Table 7 shows the comparison of the accuracy of SS-DGFNet under limited labeled data. The model achieves an average Dice score of 0.888, which is superior to or competitive with those of fully supervised baseline models such as 3D U-Net and nnU-Net, despite requiring significantly fewer labeled samples. This result demonstrates that the hybrid self-supervised pretraining strategy successfully extracts latent features from raw sensor data that conventional supervised models fail to capture, making SS-DGFNet particularly effective for emerging biosensor technologies where large annotated datasets are not yet available.

SS-DGFNet shows lower HD95 and comparable or higher Dice scores relative to Swin-Unet(17) and UNETR.(16) This result is significant from a hardware perspective, as transformer-based architectures are known for their high computational costs measured in giga floating-point operations. SS-DGFNet delivers high-fidelity results with a fraction of the parameters and computational requirements, making it far more compatible with the low-power GPUs commonly used in handheld or PoC sensing devices.

The analysis results of boundary precision show the strengths of SS-DGFNet. The HD95 values of the model are the lowest in the table, indicating superior performance in capturing fine anatomical contours. For sensor-based systems, this translates into higher reliability in edge detection, specifically in distinguishing tumor boundaries from the surrounding healthy tissue. The latency and throughput reveal that SS-DGFNet achieves a lower inference latency than the other 3D models. This capability enables the real-time or near-real-time interpretation of sensor signals, which is a critical requirement for surgical planning and rapid diagnostic workflows.

The results of comparative analysis show that SS-DGFNet enhances medical image sensing. It outperforms traditional convolutional and transformer-based architectures in terms of boundary-sensitive accuracy while maintaining a lightweight computational profile. The high-fidelity output and low hardware requirements underscore its potential for integration into the next generation of smart medical sensors.

Table 7
Results of comparative performance analysis against state-of-the-art architectures.

Model	Number of parameters (million)	FLOPs (G)	Memory (GB)	Inference time (ms)	Frames per second	Efficiency index
SS-DGFNet	4.8	42.3	1.9	125 ± 8	8.0	0.924
DGFNet	8.7	76.5	3.2	168 ± 12	5.95	0.733
nnU-Net	31.2	186.5	5.8	189 ± 15	5.29	0.701
3D U-Net	16.7	98.4	3.2	156 ± 10	6.41	0.652
TransBTS	23.5	134.7	4.5	203 ± 18	4.93	0.621

4.2 Computational efficiency and resource utilization

Figure 2 presents how quickly the models minimize error during the training phase. SS-DGFNet (solid red line) begins with a lower loss and reaches stability significantly earlier than the other models. It converges at approximately Epoch 130, whereas the nearest competitor, nnU-Net, requires around 180 epochs. An epoch represents one complete pass of the entire training dataset through the neural network, and fewer epochs required for stability indicate higher computational efficiency. Therefore, SS-DGFNet achieves stable performance at Epoch 130, meaning that it learns the necessary patterns from the sensor data after seeing the full dataset 130 times. The final training loss is lower for SS-DGFNet than for nnU-Net, 3D U-Net, and TransBTS, indicating that the hybrid self-supervised pretraining provides a head start by enabling the model to fit the sensor data distribution more effectively with fewer iterations. In addition, the SS-DGFNet curve is smoother, suggesting that the DGF module and consistency losses help stabilize the training process against noise in the 10% labeled subset.

Figure 3 presents the validation Dice score curves, which demonstrate segmentation quality on unseen data as training progresses. SS-DGFNet achieves the highest validation Dice score of

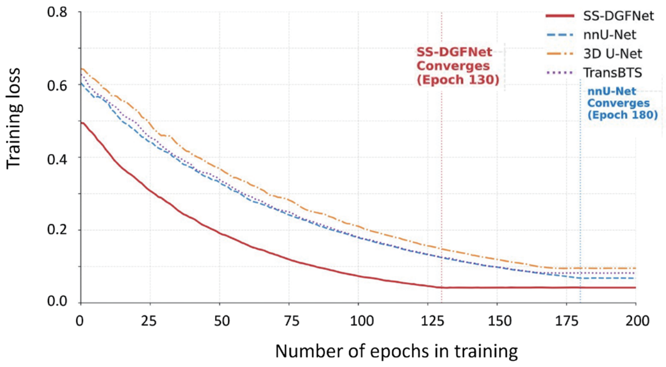


Fig. 2. (Color online) Training convergence and loss optimization with 10% of labeled images.

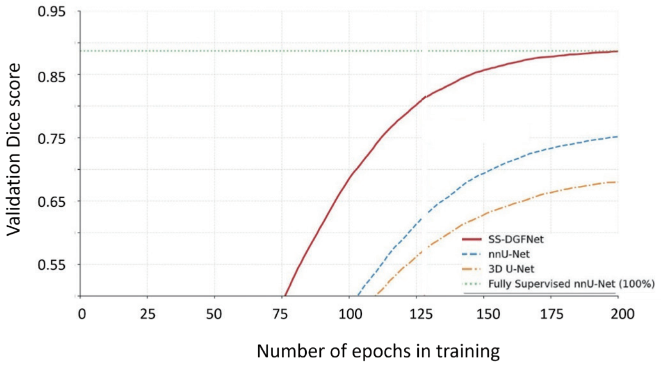


Fig. 3. (Color online) Validation Dice score during training.

0.891 at Epoch 200, which is close to that of the fully supervised nnU-Net trained with 100% of the labels, which is 1,251 (0.89, dotted green line). A substantial efficiency gap is observed between SS-DGFNet and the other models trained on the same 10% of data. While nnU-Net and 3D U-Net show a Dice score of 0.75, that of SS-DGFNet reaches 0.90, proving that the smart sensing approach extracts nearly the same amount of information from 10% of the data as a standard model does from 100%. Between Epochs 75 and 125, SS-DGFNet exhibits a steep learning surge, suggesting that once fine-tuning begins, the model rapidly maps its pretrained features to the clinical labels.

Figure 4 shows the total training time required for each model to reach its highest performance. SS-DGFNet requires 56 h of training, representing a 50% less time compared with nnU-Net's 112 h. The improved SS-DGFNet is less computationally expensive to develop. Even when the self-supervised component is removed, the DGFNet architecture alone requires only 72 h, still outperforming traditional baselines in terms of efficiency.

Figure 5 presents the lightweight nature of SS-DGFNet, making it well-suited for integration into medical sensor hardware or edge devices. SS-DGFNet requires 1.9 gigabytes (GB) of video random access memory, which is three times less than that of nnU-Net (5.8 GB) and significantly lower than that of TransBTS (4.5 GB) [Fig. 5(a)]. This efficiency enables deployment on consumer-grade or portable medical tablets. SS-DGFNet contains 4.8 million parameters, whereas nnU-Net is more than six times larger at 31.2 million parameters [Fig. 5(b)]. Although SS-DGFNet is the smallest and least memory-intensive model, it delivers the highest segmentation accuracy, establishing it as both a computationally efficient and clinically effective solution.

Figure 6 shows the inference speed values of three different GPU models: NVIDIA GeForce RTX 4090, RTX 3090, and GTX 1080TI. Each GPU is evaluated across three performance conditions, represented by red, blue, and purple bars. RTX 4090 delivers the highest performance, achieving 82 frames per second (FPS) (red), 56 FPS (blue), and 46 FPS (purple), all well above the real-time threshold of 30 FPS (green dashed line). RTX 3090 shows 68, 42, and 34 FPS, also meeting real-time standards across all conditions. GTX 1080TI, highlighted as

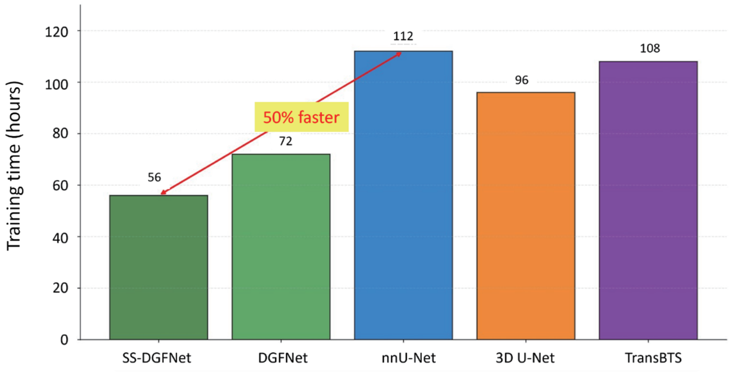
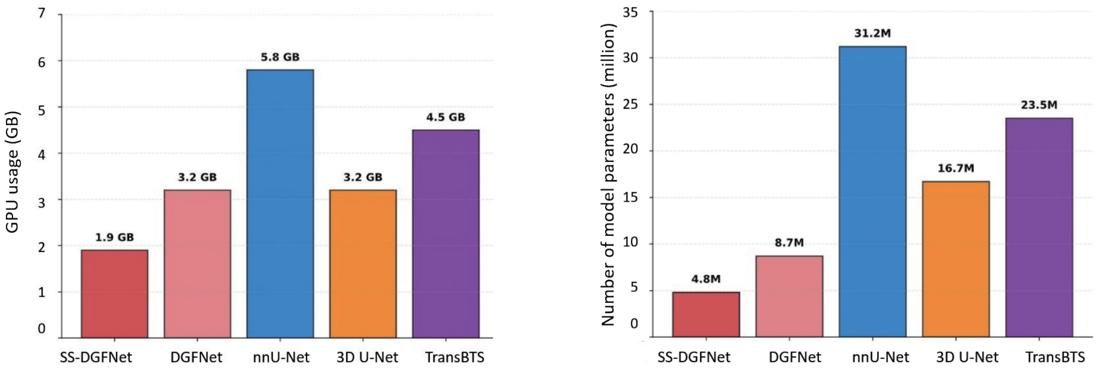


Fig. 4. (Color online) Training time of each model



(a) (b)

Fig. 5. (Color online) GPU memory usage and number of model parameters.

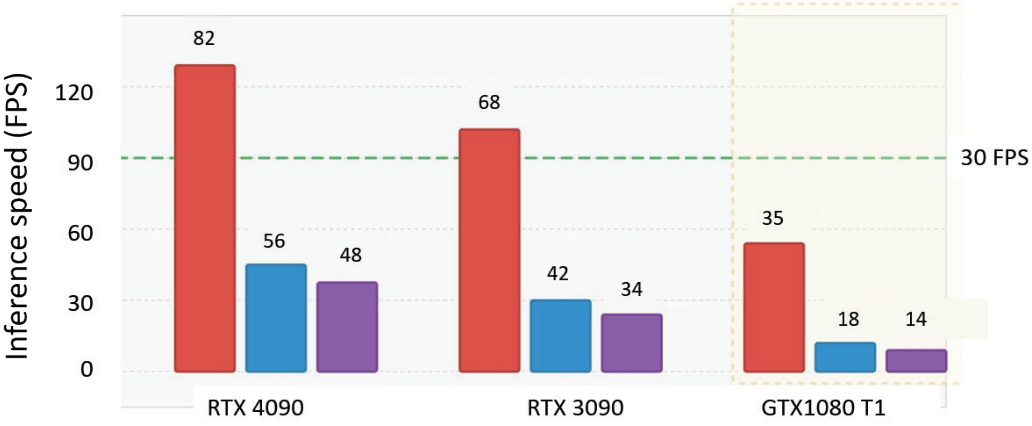


Fig. 6. (Color online) Inference speed comparison across different GPU architectures.

a consumer-grade GPU, shows limited capability with 35 FPS (red), 18 FPS (blue), and 14 FPS (purple). Only the red bar meets the real-time threshold, while the others fall short and are marked as slow. The results demonstrate that SS-DGFNet achieves real-time inference on high-end GPUs, such as RTX 4090 and RTX 3090, while maintaining acceptable performance even on consumer-grade hardware, such as GTX 1080TI. This confirms the model's suitability for deployment in both clinical and portable environments.

5. Conclusion

We developed an improved SS-DGFNet to address the challenges of data scarcity and hardware constraints in medical sensor applications. By integrating a hybrid SSL model with a DGF module, high-fidelity 3D imaging is ensured, and the limited computational capacity of PoC sensing devices is addressed.[4] In this study, we aimed to address the problem of deploying sophisticated diagnostic AI on resource-constrained embedded sensor systems, which often face signal instability or loss. We developed a robust SS-DGFNet that integrates self-supervised pre-calibration with a dynamic gating mechanism that adjusts for missing sensor modalities. While the model significantly reduces computational overhead and maintains high fidelity in benchmark tests, remaining issues include its validation in real-world, noisy clinical environments and its integration with diverse, nonstandardized sensor hardware architectures.

The experimental results demonstrated that the model developed is resource-efficient, requiring only 1.9 GB of memory and 4.8 million parameters, which are significantly lower than those of standard models such as nnU-Net.(32) Furthermore, the model achieved an average Dice score of 0.888 using 10% of labeled data, effectively accelerating the deployment cycle for emerging biosensing technologies by reducing the dependency on labor-intensive manual annotations.(33)

The results of this study provide a reference to develop a fault-tolerant sensing interface. The DGF module acts as an intelligent reliability filter, allowing the model to maintain 97.6% of its performance retention even when a specific sensor modality is missing or fails. This robustness ensures that clinically actionable diagnostics remain possible under suboptimal hardware conditions.

The improved SS-DGFNet enables a high-fidelity, lightweight solution for real-time sensor image interpretation. Its ability to learn from raw, unannotated signals and its gradual degradation in the face of sensor failure lead to the advancement of next-generation smart medical sensors and edge-computing diagnostic platforms. Incorporating uncertainty estimation in the model enhances the reliability of intelligent sensing systems in diverse clinical environments.

References

1 C. Michail, P. Liaparinos, N. Kalyvas, I. Kandarakis, G. Fountos, and I. Valais: Sensors **24** (2024) 6251. https://doi.org/10.3390/s24196251

2 J. Teuho, A. Torrado-Carvajal, H. Herzog, U. Anazodo, R. Klén, H. Iida, and M. Teräs: Front. Phys. **7** (2019) 243. https://doi.org/10.3389/fphy.2019.00243

3 M. Wang, S. Fan, Y. Li, Z. Xie, and H. Chen: J. Biomed. Inform. **164** (2025) 104796. https://doi.org/10.1016/j.jbi.2025.104796

4 Y. Xu, T. M. Khan, Y. Song, and E. Meijering: Artif. Intell. Rev. **58** (2025). https://doi.org/10.1007/s10462-024-11033-5

5 B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, and J. Kirby: IEEE Trans. Med. Imaging **34** (2015) 1993. https://doi.org/10.1109/TMI.2014.2377694

6 S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos: Sci. Data **4** (2017) 170117. https://doi.org/10.1038/sdata.2017.117

7 B. Bonato, L. Nanni, and A. Bertoldo: Sensors **25** (2025) 1838. https://doi.org/10.3390/s25061838

8 N. S. Punn and S. Agarwal: Artif. Intell. Rev. **55** (2022) 5845. https://doi.org/10.1007/s10462-022-10152-1

9 G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez: Med. Image Anal. **42** (2017) 60. https://doi.org/10.1016/j.media.2017.07.005

10 Y. Xu, R. Quan, W. Xu, Y. Huang, X. Chen, and F. Liu: Bioengineering **11** (2024) 1034. https://doi.org/10.3390/bioengineering11101034

11 J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert: Med. Image Anal. **53** (2019) 197. https://doi.org/10.1016/j.media.2019.01.012

12 Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang: IEEE Trans. Med. Imaging **39** (2020) 1856. https://doi.org/10.1109/TMI.2019.2959609

13 IF. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein: Nat. Methods **18** (2021) 203. https://doi.org/10.1038/s41592-020-01008-z

14 F. Shamshad, S. Khan, S. W. Zamir, M. H. Khan, M. Hayat, F. S. Khan, and H. Fu: Med. Image Anal. **88** (2023) 102802. https://doi.org/10.1016/j.media.2023.102802
15 C. Zhang, X. Deng, and S. H. Ling: Sensors **24** (2024) 4668. https://doi.org/10.3390/s24144668
16 S. Atek, I. Mehidi, D. Jabri, and D. E. C. Belkhiat: Brain Imaging Behav. **19** (2025) 1417. https://doi.org/10.1007/s11682-025-01052-3
17 L. Jing and Y. Tian: IEEE Trans. Pattern Anal. Mach. Intell. **43** (2021) 4037. https://doi.org/10.1109/TPAMI.2020.2992393
18 H. Hu, X. Wang, Y. Zhang, Q. Chen, and Q. Guan: Neurocomputing **610** (2024) 128645. https://doi.org/10.1016/j.neucom.2024.128645
19 V. Rani, M. Kumar, A. Gupta, M. Sachdeva, A. Mittal, and Krishan: Radiol. Oncol. **15** (2024) 1607. https://doi.org/10.1007/s12530-024-09581-w
20 J. Mao, S. Guo, X. Yin, Y. Chang, B. Nie, and Y. Wang: Appl. Soft Comput. **169** (2025) 112536. https://doi.org/10.1016/j.asoc.2024.112536
21 T. Yoon and D. Kang: Eng. Appl. Artif. Intell. **160** (2025) 112055. https://doi.org/10.1016/j.engappai.2025.112055
22 Y. Wu, D. Zeng, Z. Wang, Y. Shi, and J. Hu: Med. Image Anal. **81** (2022) 102564. https://doi.org/10.1016/j.media.2022.102564
23 Y. Sun, W. Chen, and Z. Sun: Digit. Signal Process. **167** (2025) 105441. https://doi.org/10.1016/j.dsp.2025.105441
24 H. H. Lee, Y. Tang, Q. Yang, X. Yu, L. Y. Cai, and L. W. Remedios: IEEE J. Biomed. Health Inform. **27** (2023) 4444. https://doi.org/10.1109/JBHI.2023.3285230
25 T. Zhou, S. Canu, P. Vera, and S. Ruan: IEEE Trans. Image Process. **30** (2021) 4263. https://doi.org/10.1109/TIP.2021.3070752
26 L. Pinto-Coelho: Bioengineering **10** (2023) 1435. https://doi.org/10.3390/bioengineering10121435
27 N. Arandia, J. I. Garate, and J. Mabe: Sensors **22** (2022) 9917.https://doi.org/10.3390/s22249917
28 J. Zhang, Q. Dong, J. Shi, Q. Li, C. M. Stonnington, B. A. Gutman, K. Chen, E. M. Reiman, R. J. Caselli, P. M. Thompson, J. Ye, and Y. Wang: Med. Image Anal. **70** (2021) 102009 https://doi.org/10.1016/j.media.2021.102009
29 Z. Zhu, K. Yu, G. Qi, B. Cong, Y. Li, Z. Li, and X. Gao: Comput. Biol. Med. **182** (2024) 109204. https://doi.org/10.1016/j.compbiomed.2024.109204
30 F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein: Nat. Methods **18** (2021) 203. https://doi.org/10.1038/s41592-020-01008-z
31 MRIMATER: https://mrimaster.com/t1-vs-t2-vs-pd-vs-flair-mri/?utm_source=copilot.com (accessed February 2026).
32 L. Qi, Z. Jiang, W. Shi, F. Qu, and G. Feng: Comput. Biol. Med. **176** (2024) 108547. https://doi.org/10.1016/j.compbiomed.2024.108547
33 A. Bitarafan, M. Mozafari, M. F. Azampour, M. S. Baghshah, N. Navab, and A. Farshad: Med. Image Anal. **101** (2025) 103478. https://doi.org/10.1016/j.media.2025.103478