International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(21s) 2026 pp. 1628-1640
https://doi.org/10.70917/ijcisim-2026-5915
Article

Role of Deep Learning in Automated Tumor Detection and Segmentation in MRI and CT Imaging

Ahmed Mudkhre Shareef[1]*

[1] Radiologist, Ninevah University, Medical College, Radiology Department, Mosul, Iraq.
Email: Ahmed.shareef@uoninevah.edu.iq ORCID: 0000-0002-9851-5664

Abstract: Manual delineation of the tumour in magnetic resonance imaging (MRI) or computed tomography (CT) is time consuming, operator dependant and prone to considerable inter-observer variation. Although deep learning (DL) is a potential approach to automatic tumor detection and segmentation, there is still a lack of quantitative comparison between architectures and imaging modalities. In this work, five DL architectures – U-Net, Attention U-Net, ResUNet++, nnU-Net, and Transformer-based Swin-UNETR – were compared and evaluated for the two segmentation tasks: brain tumor segmentation (BraTS 2021 MRI subset) and pulmonary nodule/tumor detection and segmentation (LIDC-IDRI CT subset). The total number of MRI and CT volumes were split into training (70%), validation (15%) and testing (15%). All pre-processing steps included bias-field correction, resampling, intensity normalization and patch extraction. Five-fold cross validation was used to train the models on the Adam optimizer with dice–cross-entropy as the loss function. The Dice similarity coefficient (DSC), Intersection-over-Union (IoU), 95th-percentile Hausdorff distance (HD95), sensitivity, specificity and inference time were used to assess performance. The paired t test and one-way ANOVA models were used at the $\alpha = 0.05$ level of significance. The classical region growing/thresholding methods were significantly outperformed by all DL models ($p < 0.001$). nnU-Net achieved the highest DSC for MRI (0.912 ± 0.021) and CT (0.887 ± 0.026), followed by Swin-UNETR. The transformer and attention models improved the accuracy of the boundaries of heterogeneous and small lesions. In general, nnU-Net achieved the highest accuracy-efficiency tradeoff, which enabled the incorporation of DL into the workflow of radiologists without replacing them in the diagnosis process.
Keywords: Deep Learning; Tumor Segmentation; Medical Image Analysis; Convolutional Neural Network; U-Net; Transformer; MRI; CT; Radiomics; Computer-Aided Diagnosis

1. Introduction

Cancer is on the rise and there is an urgent need for cancer diagnostics to be accurate, timely and automated. The magnetic resonance imaging (MRI) and computed tomography (CT) have been commonly used to detect, characterize and plan treatment for tumours due to their detailed anatomical and pathological information (Spaanderman, D. J., Starmans, M. P., van Erp, G. C., Hanff, D. F., Sluijter, J. H., 2025). Traditional tumor detection and segmentation, however, mainly rely on manual evaluation by the radiologist that is time consuming, subjective and susceptible to inter-observer variability (Rai, 2024). To overcome these drawbacks, automated image analysis has been proposed as a solution and a subdomain of AI that has shown great potential for success is known as deep learning (DL).

The models of deep learning, such as Convolutional Neural Networks (CNNs), U-Net and their advanced versions, can provide automatic learning of complex hierarchical features in large numbers of medical images. These models can help detect suspicious lesions, differentiate between a tumour and healthy tissue in MRI and CT imaging, and create accurate tumour outlines (Krishnapriya, S., & Karuna, Y. 2023). Automated segmentation delivers quantitative information such as tumor size, volume, shape, and spatial distribution that can be beneficial for diagnosis, prognosis, radiotherapy planning, and assessing treatment response (Karim, S., Tong, G., Yu, Y., Laghari, A. A.,

Copyright: © 2026 by the authors

Khan, A. A., 2024). Moreover, the three-dimensional and attention-based deep learning methods have enhanced the ability to include contextual and spatial information in medical images. Though these benefits, however, issues like lack of annotated data, varying imaging techniques, class imbalance, interpretability of models, and clinical validation are of significant concern (Kashyap, M., Wang, X., Panjwani, N., Hasan, M., Zhang, Q., 2025).

1.1 Background and Clinical Motivation

Cancer is still one of the top causes of death in the world, and the correct and timely identification of the location and extent of a tumour, as well as the tumour morphology, are critical to treatment planning, prognosis and longitudinal monitoring. Cross-sectional imaging modalities for tumour characterisation most commonly used are magnetic resonance imaging (MRI) and computed tomography (CT) imaging MRI is the modality of choice for central nervous system, hepatic and musculoskeletal tumors, and CT imaging is used for rapid, high-resolution volumetric imaging useful for pulmonary, hepatic, and abdominal oncologic imaging (Wesdorp, N. J., Zeeuw, J. M., Postma, S. C., Roor, J., van Waesberghe, J. H. T., 2023). Nevertheless, the contours of tumors are still manually or semi-automatically defined during routine clinical practice by radiologists, a time-intensive, highly specialized and inter- and intra-observer dependent task with reported Dice similarity coefficients ranging from 0.74 to 0.85 for gliomas and even lower for infiltrative or low-contrast lesions (Illimoottil, M., & Ginat, D. 2023).

The problem is alleviated by automated tumor detection and segmentation, where the presence, location and precise voxel-wise extent of tumor tissue is algorithmically identified. The traditional methods such as thresholding, region growing, active contours, level-set methods and atlas-based registration are based on handcrafted intensity or texture features, and often do not work when the appearance of the tumor is heterogeneous, the boundaries are not well defined, the partial volume effects are present or the inter-scanner variation exists (Hussain, D., Al-Masni, M. A., Aslam, M., Sadeghi-Niaraki, A., Hussain, J., 2024). Deep learning (DL), and convolutional neural networks (CNNs) especially, have revolutionized over the last decade the field of medical image analysis, which automatically learns hierarchical, task-specific feature representations from data without the need for manual feature engineering (Ahsan, R., Shahzadi, I., Najeeb, F., & Omer, H. 2025).

1.2 Problem Statement

Although a large number of study's have achieved high segmentation accuracy on single datasets with single deep learning models, the following three gaps still exist. First, the results published in the literature are not directly comparable due to the differences in the pre-processing pipelines, data split and evaluation metrics used in the studies. Second, there is a lack of studies that compare classical CNN encoder–decoder designs (such as U-Net variants) with newer self-configuring pipelines (such as nnU-Net) and new Transformer-based designs within the same experimental setup and both MRI and CT modalities. Third, a crucial factor for clinical deployment, namely the compromise between segmentation accuracy and computational costs, seldom goes hand in hand with the accuracy metrics. This study attempts to overcome these shortcomings by performing a head-to-head experimental comparison and control.

1.3 Research Objectives

The specific goals of this study are (i) to evaluate, quantitatively, the performance of five representative deep learning architectures for tumor detection and segmentation on MRI and CT imaging; (ii) to compare the performance of the deep learning architectures to a classical image-processing baseline, also quantitatively; (iii) to analyse the accuracy–efficiency trade-off among these architectures and inform decisions for deployment; and (iv) to statistically test whether the observed performance differences are significant across models and modalities.

1.4 Research Questions and Hypotheses

RQ1: Can deep learning surpass classical segmentation techniques in the task of tumor delineation in MRI and CT?

RQ2: What is the best architectural family (plain CNN encoder–decoder, attention-augmented CNN, self-configuring CNN pipeline or CNN–Transformer hybrid) for each modality in terms of segmentation accuracy?

RQ3: Do the architectures differ significantly in terms of trade-off between segmentation accuracy and inference time?

The null (H0) and alternative (H1) hypotheses were written for each model/baseline and used to test for the significance of the difference at $\alpha = 0.05$.

1.5 Significance of the study

In addition to the comparative results, this study is methodologically relevant as it sets up a framework with identical pre-processing, cross-validation structure, loss formulation, and evaluation that can be applied to other architectures and modalities and to other organ systems as the field continues to grow. The results provide clinical evidence to help radiology departments and software vendors for medical imaging select an architecture that matches their throughput and accuracy needs, without having to rely on individual, non-comparable published benchmarks. From an academic point of view, the study responds to an opportunity in the medical-imaging AI literature to rigorously and reproducibly benchmark the accuracy of an AI model, as well as its computational cost, often ignored in architecture-proposing studys, which naturally focus on the improvement of accuracy and spend less time on feasibility of deployment.

2. Related Work

This part places the current work in the context of the previous research on medical image segmentation with deep learning.

2.1 Classical Approaches to Tumor Segmentation

The initial computer-aided detection methods were based on an intensity thresholding, morphological operations, region growing, watershed transforms and active-contour/level-set models (Han, J. H., Kim, B. W., Kim, T. M., Ko, J. Y., Choi, S. J., 2025). Handcrafted radiomic features were fed into feature-based machine learning methods (support vector machines, random forests) which enhanced robustness of these algorithms, but were still sensitive to feature selection and required intensive pre-processing. The methods typically failed to handle the high inter-patient variability in tumor shape, size, and intensity, driving the development of representation-learning methods (Ronneberger, O., Fischer, P., & Brox, T. 2015).

2.2 Convolutional Neural Networks for Segmentation

The fully convolutional networks and, more recently, the U-Net encoder–decoder architecture, with skip connections, set the new standard for biomedical image segmentation, allowing for accurate localization and the retention of context information (Kashyap, M., Wang, X., Panjwani, N., 2025). Many extensions were developed that extended the framework to volumetric (3D U-Net, V-Net) data, attention gates to suppress irrelevant background features and highlight salient tumor regions (Attention U-Net), and residual and dense connections to enhance the gradient flow and feature reuse in deeper networks (Residual U-Net, Residual U-Net++) (Jha, D., Smedsrud, P. H., Riegler, M. A., 2019). The nnU-Net framework was then shown to be able to match or even surpass hand-designed architectures in a large variety of biomedical segmentation benchmarks, with no manual tuning required, by adapting pre-processing, network topology, and training schedule automatically to the characteristics of the datasets (Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., 2018).

2.3 Transformer-Based and Hybrid Architectures

In recent years, models based on Vision Transformers (ViTs) and architectures combining CNN–Transformer (e.g., TransUNet, Swin-UNet, UNetr) that can effectively model long-range spatial dependencies have been introduced, which CNN–conv. kernels with a small receptive field may not be able to properly model

(Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H. R., & Xu, D. 2022). Such architectures have either a shifted-window Transformer or a hierarchical Transformer encoder and generally a convolutional decoder, and are competitive or superior on brain tumor segmentation and organ segmentation in the abdomen in general, but often at higher computational expense and higher data requirements, especially for large or irregularly shaped organs or tumors (Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., 2021).

2.4 Applications in MRI and CT Tumor Imaging

In the field of neuro-oncology, segmentation of gliomas on multi-parametric MRI (T1, T1-contrast, T2, FLAIR) has been widely studied on the BraTS challenge series, which offers a standardised benchmark for segmentation of four sub-regions of the tumour: whole tumor, tumor core, enhancing tumor, and non-enhancing tumor (Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J.,2015). For thoracic oncology, the datasets used for studying pulmonary nodule detection and segmentation include LIDC-IDRI and LUNA16, and DL pipelines take advantage of candidate-region detection (such as 3D region-proposal networks) and of segmentation refinement. In related applications, such as hepatic tumor segmentation on contrast-enhanced CT (LiTS challenge), breast lesion segmentation on MRI and prostate tumor delineation, and significant improvements over classical segmentation baselines have been obtained (Rai, 2024).

2.5 Identified Research Gap

This is because although individual architectures have been validated separately, there are relatively few controlled studies across plain CNN, attention-augmented CNN, self-configuring, and Transformer-hybrid architectures that were studied under one unified protocol on both MRI and CT. Furthermore, there are inconsistencies between accuracy and computational efficiency (inference time, number of parameters) reporting, which hampers evidence-based recommendations for clinical use. The aim of this study is to fill that gap by conducting a simple experimental comparison.

3. Materials and Methods

3.1 Datasets

The datasets used were two publicly available and anonymised imaging datasets. For this study, a dataset of 660 multi-parametric brain MR volumes (T1-contrast and FLAIR sequences) from a BraTS-2021-derived cohort, each supplemented with an expert segmentation mask of the whole tumor, was used. Dataset B included 620 thoracic CT volumes, each with one or more radiologist-annotated pulmonary nodules/tumors and consensus segmentation masks (nodules ≥ 3 mm, from at least three of four radiologists). The characteristics of the data are presented in Table 1.

Table 1. Summary of imaging datasets used in this study

Characteristic	Dataset A (MRI – Brain Tumor)	Dataset B (CT – Pulmonary Tumor/Nodule)
Source cohort	BraTS-2021–derived	LIDC-IDRI–derived
Number of volumes	660	620
Modality / sequence	T1-contrast, FLAIR	Contrast-enhanced chest CT
In-plane resolution	240 x 240 (resampled to 1 mm3)	512 x 512 (resampled to 1 mm3)
Slice thickness (native)	1.0-1.5 mm	1.0-2.5 mm

Annotation	Expert whole-tumor voxel mask	Consensus nodule/tumor voxel mask
Train / Val / Test split	462 / 99 / 99 (70:15:15)	434 / 93 / 93 (70:15:15)
Positive lesion prevalence	100% (tumor present)	100% (nodule/tumor present)

3.2 Image Pre-processing

Within the brain mask, all MRI volumes were transformed into N4 bias-field corrected, skull-stripped, and resampled to isotropic 1mm 3 voxels and z-score intensity normalised. The volumes were resampled to 1 mm3 isotropic voxels, clipped to a Hounsfield-unit window to highlight lung and soft-tissue contrast (-1000, 400), and min/max normalised (0, 1). In both modalities 3-D patches of 128 x 128 x 128 voxels were extracted with 50% overlap, which was oversampled with foreground voxels (67% of patches centred at tumor voxels) to mitigate class imbalance. Data augmentation techniques employed were random rotation (15 deg), scaling (0.9-1.1x), elastic deformation, gamma intensity variation, and mirroring, applied during training on-the-fly.

3.3 Deep Learning Architectures Evaluated

Five segmentation pipelines were developed and trained following a similar protocol:

(1) U-Net (3-D baseline): a standard encoder-decoder architecture with 4 downsampling stages, skip connections and instance normalisation. (2) Attention U-Net: U-Net with additive attention gates added to skip connections to focus on features related to the tumor and to reduce background activations. (3) ResUNet++: Residual blocks, squeeze and excitation channel attention, atrous spatial pyramid pooling and an attention based decoder. (4) nnU-Net: a self-configuring pipeline which automatically infers pre-processing parameters, network depth/width, patch size and training schedule from a dataset fingerprinting; this 3-D full-resolution configuration was employed. (5) Swin-UNETR: a hybrid model with a hierarchical shifted-window Vision Transformer (ViT) as encoder and a CNN-based decoder with skip connections to provide long-range volumetric context.

A classical pipeline (Otsu thresholding, morphological region growing and 3-D connected component analysis) was also deployed to quantify the improvement coming from deep learning.

Table 2. Architectural and training configuration summary

Model	Encoder	Parameters (M)	Loss function	Optimiser	Initial LR
Classical baseline	N/A (rule-based)	N/A	N/A	N/A	N/A
U-Net (3D)	Plain CNN	16.3	Dice + Cross-Entropy	Adam	1e-4
Attention U-Net	CNN + attention gates	18.9	Dice + Cross-Entropy	Adam	1e-4
ResUNet++	Residual + SE blocks	24.1	Dice + Focal	Adam	1e-4
nnU-Net (3D full-res)	Self-configured CNN	31.5	Dice + Cross-Entropy	SGD (Nesterov)	1e-2
Swin-UNETR	Swin Transformer + CNN decoder	62.2	Dice + Cross-Entropy	AdamW	1e-4

3.4 Training Protocol

All models were trained with 5-fold cross validation on training/validation partition, with early stopping (patience = 50 epochs) on validation Dice score measure, for a maximum of 500 epochs, batch size 2 (3-D patches), and mixed-precision computation. The Dice and cross-entropy (or focal, for ResUNet++) loss was used to train the model together for overlap and per-voxel classification accuracy. A polynomial schedule (nnU-Net) or cosine annealing (other models) was used to decay the learning rate. All models were trained with identical data splits as well as augmentation policy and hardware (NVIDIA A100 40GB GPUs), which allows for a fair comparison. The model checkpoint with the maximum mean across folds validation Dice score was selected and ensembled by majority voxel voting across the 5 folds for final test-set evaluation.

3.5 Evaluation Metrics

The segmentation performance was evaluated by Dice Similarity Coefficient (DSC), Intersection-over-Union (IoU/Jaccard index), voxel-wise sensitivity (recall) and specificity, and by the 95th-percentile Hausdorff Distance (HD95, mm), as a boundary-accuracy measure. Additionally, detection performance (presence/absence and localisation of tumour) was summarized by the per case sensitivity and the per scan false positive rate. The efficiency of the computation was evaluated by mean inference time per volume (seconds) on a single GPU and the number of trainable parameters. Comparisons of the models were made by paired two-tailed t-tests on per-case Dice score and one-way repeated-measures ANOVA with Bonferroni-corrected post-hoc tests across all five DL models plus baseline, with the significance level set at $\alpha = 0.05$.

3.6 Experimental Environment

The experiments were conducted on a workstation equipped with dual NVIDIA A100 (40GB) GPUs and an AMD EPYC processor, with 128GB RAM.Experiments were performed using a workstation with dual NVIDIA A100 (40 GB) GPUs, 128 GB RAM, and an AMD EPYC processor, running Python 3.10 with PyTorch 2.1 and MONAI 1.3 frameworks. Analyses were done in Python (SciPy, statsmodels) and verified in R 4.3.

3.7 Ethical Considerations and Data Governance

No attempt at contact or re-identification of patients or linkage to external records was made and both data sets included in this study were fully de-identified and publicly released under the original data use agreement of BraTS and LIDC-IDRI, which prohibited the inclusion of patients' protected health information before release. This study used only secondary analysis of existing, pre-collected, de-identified publicly available data sets, and was determined to be exempt from further institutional review board approval, as is typical for retrospective imaging-informatics research; if the same protocol is used with institution-specific data sets, researchers should obtain appropriate ethics approval and patient consent or waiver documentation. In this study, there were no clinical outputs generated or diagnostic outputs.

3.8 Reproducibility Measures

For reproducibility, all data was split in a fixed manner and all random seeds were set when splitting the data and initialising the models within the five-fold cross-validation. All augmentation pipelines, patch-sampling strategies, and normalisation procedures were the same for all architectures. All hyperparameters, such as those that were automatically determined in the self-configuration step of nnU-Net, were logged and archived with all model checkpoints. In the computation of all evaluation metrics (DSC, IoU, HD95, sensitivity and specificity), the same implementation of the metrics across the different models was used to ensure that the evaluation metric was not an implementation confound in the comparison.

4. Results

4.1 Overall Segmentation Performance on MRI (Brain Tumor)

The mean ± standard deviation performance values measured on the MRI test set (n = 99 volumes) for five cross-validation folds are reported in Table 3. All deep learning models yielded significantly higher scores than the classical baseline on all scores (paired t-test, $p < 0.001$). For DL, nnU-Net showed the best mean DSC (0.912 ± 0.021) followed by Swin-UNETR (0.905 ± 0.024) while the plain U-Net baseline achieved the worst DL performance (0.859 ± 0.031).

Table 3. Segmentation performance on MRI test set (mean ± SD, n = 99)

Model	DSC	IoU	HD95 (mm)	Sensitivity	Specificity	Time (s)
Classical baseline	0.671 ± 0.058	0.531 ± 0.061	14.82 ± 3.91	0.702 ± 0.065	0.968 ± 0.012	2.1
U-Net (3D)	0.859 ± 0.031	0.775 ± 0.036	6.47 ± 1.62	0.868 ± 0.033	0.991 ± 0.006	1.4
Attention U-Net	0.891 ± 0.027	0.815 ± 0.031	5.02 ± 1.31	0.897 ± 0.028	0.993 ± 0.005	1.6
ResUNet++	0.884 ± 0.029	0.804 ± 0.033	5.44 ± 1.44	0.889 ± 0.030	0.992 ± 0.005	1.9
nnU-Net	0.912 ± 0.021	0.842 ± 0.026	3.71 ± 0.98	0.918 ± 0.023	0.995 ± 0.004	3.2
Swin-UNETR	0.905 ± 0.024	0.831 ± 0.028	3.98 ± 1.05	0.909 ± 0.025	0.994 ± 0.004	4.6

The mean DSC improvement of 24 percentage points of nnU-Net over the classical baseline (0.912 vs 0.671) shows that data-driven feature learning is able to better capture the tumor-tissue heterogeneity compared to intensity-threshold based rules, which are very sensitive to partial-volume effects at infiltrative tumor margins. The results of the consistently lower HD95 achieved by nnU-Net (3.71 mm) and Swin-UNETR (3.98 mm) compared to the plain U-Net (6.47 mm) demonstrate that self-configuring optimisation and long-range context modelling both meaningfully improve boundary localisation, which is clinically relevant for radiotherapy target-volume delineation where sub-centimetre boundary errors would influence the margin design.

4.2 Overall Segmentation Performance on CT (Pulmonary Tumor/Nodule)

Table 4 shows similar results for the test set of CT (n = 93 volumes). The relative model ranking was similar to the MRI experiment, however absolute DSC values were slightly reduced in all models, corresponding to lower average lesion size seen in the thoracic CT compared to brain MRI tumors, and higher background tissue heterogeneity in the thoracic CT.

Table 4. Segmentation performance on CT test set (mean ± SD, n = 93)

Model	DSC	IoU	HD95 (mm)	Sensitivity	Specificity	Time (s)
Classical baseline	0.598 ± 0.071	0.463 ± 0.069	18.35 ± 4.87	0.649 ± 0.078	0.951 ± 0.017	1.8
U-Net (3D)	0.821 ± 0.038	0.727 ± 0.041	8.14 ± 2.05	0.834 ± 0.040	0.986 ± 0.008	1.3
Attention U-Net	0.858 ± 0.033	0.769 ± 0.036	6.29 ± 1.72	0.865 ± 0.034	0.989 ± 0.006	1.5
ResUNet++	0.849 ± 0.035	0.756 ± 0.038	6.81 ± 1.85	0.854 ± 0.036	0.988 ± 0.007	1.8

nnU-Net	0.887 ± 0.026	0.801 ± 0.030	4.63 ± 1.28	0.892 ± 0.027	0.992 ± 0.005	3.0
Swin-UNETR	0.881 ± 0.028	0.792 ± 0.032	4.91 ± 1.35	0.887 ± 0.029	0.991 ± 0.005	4.3

The performance gap between the DL and the classical baseline model is larger on CT than on MRI, indicating that the rules based on intensity are more challenging to differentiate between pulmonary nodules and the similar Hounsfield-unit range (HU) of vasculature, bronchial walls, and pleural surfaces, an issue that is much more easily addressed by convolutional and self-attention features. The relatively small absolute difference of 2.9 DSC points between nnU-Net and Attention U-Net on CT versus 2.1 points on MRI indicate the importance of attention mechanisms in cases where the lesions are small and located in complex background anatomy, which is in keeping with the role of the attention-gate mechanism of suppressing irrelevant activations.

4.3 Statistical Significance Testing

The models with the best performance for segmentation accuracy were found to be statistically different across cases for both MRI ($F(5,494) = 187.4$, $p < 0.001$) and CT ($F(5,464) = 201.6$, $p < 0.001$) data using one-way repeated-measures ANOVA. The MRI dataset showed that the two top-performing architectures were statistically comparable as all the DL-vs-baseline comparisons were significant at $p < 0.001$, and nnU-Net was significantly better than all other DL architectures except Swin-UNETR ($p = 0.084$, not significant), which used a larger point estimate. Bonferroni-corrected pairwise post-hoc comparisons are summarised in Table 5.

Table 5. Pairwise post-hoc comparison of DSC (MRI dataset); values are Bonferroni-adjusted p-values

Comparison	Mean DSC difference	Adjusted p-value	Significant (a=0.05)?
nnU-Net vs Baseline	0.241	< 0.001	Yes
nnU-Net vs U-Net	0.053	< 0.001	Yes
nnU-Net vs Attention U-Net	0.021	0.006	Yes
nnU-Net vs ResUNet++	0.028	0.001	Yes
nnU-Net vs Swin-UNETR	0.007	0.084	No
Swin-UNETR vs Attention U-Net	0.014	0.032	Yes
Attention U-Net vs ResUNet++	0.007	0.291	No
U-Net vs Baseline	0.188	< 0.001	Yes

4.4 Accuracy–Efficiency Trade-off

Table 6 provides a summary of the accuracy (mean DSC over both modalities) and of the computational cost (inference time and number of parameters) to choose the deployment. At comparable accuracy, the inference time of Swin-UNETR is more than that of attention U-Net, although both achieve competitive accuracy; Swin-UNETR, however, has a high computational cost due to the quadratic-to-log-linear nature of self-attention over volumetric tokens even in its windowed formulation.

Table 6. Composite accuracy-efficiency ranking (averaged across MRI and CT test sets)

Model	Mean DSC (both modalities)	Mean inference time (s)	Parameters (M)	Efficiency index (DSC / s)

Classical baseline	0.635	1.95	N/A	0.326
U-Net (3D)	0.840	1.35	16.3	0.622
Attention U-Net	0.875	1.55	18.9	0.565
ResUNet++	0.867	1.85	24.1	0.469
nnU-Net	0.900	3.10	31.5	0.290
Swin-UNETR	0.893	4.45	62.2	0.201

The efficiency index (mean DSC / inference time), which emphasizes that for high-volume screening, U-Net or Attention U-Net would be the most desirable trade-off, while for tasks like radiotherapy planning, where near-maximal accuracy for each case is more important than turnaround time, nnU-Net and Swin-UNETR may be the more appropriate choices.

4.5 Detection Performance and False-Positive Analysis

For tumor/nodule detection, the sensitivity for nnU-Net on both modalities was above 96%, and the spurious detections per scan (false-positive results that do not correspond to an annotated lesion) monotonically decreased in nnU-Net (0.21 FP/scan) compared to the classical baseline models (1.42 FP/scan). The detection level results are summarised in Table 7.

Table 7. Case-level detection performance summary

Model	MRI Detection Sensitivity	CT Detection Sensitivity	CT False Positives / scan
Classical baseline	89.9%	84.9%	1.42
U-Net (3D)	97.0%	95.7%	0.58
Attention U-Net	98.0%	97.8%	0.39
ResUNet++	97.0%	96.8%	0.44
nnU-Net	99.0%	98.9%	0.21
Swin-UNETR	98.0%	98.9%	0.24

4.6 Qualitative Observations

Qualitatively, the quantitative results were confirmed by visual evaluation of the segmentation masks overlaid on the ground truth (not shown here for space constraints): The masks of plain U-Net showed uneven and sometimes discontinuous borders as well as small islands of false positives adjacent to vasculature, while the masks produced by Attention U-Net and ResUNet++ were more smooth and continuous with fewer false positives; the masks generated by nnU-Net were closer to the expert annotations especially at the tumor border, even for irregularly shaped and infiltrative lesions, and the masks from Swin-UNETR were comparable to those of nnU-Net for large, well-distinguished tumors, and tended to under-segment very small (<5 mm) pulmonary nodules, as is known for patch/window-based Transformer tokenisation for sub-token-sized structures.

5. Discussion

This section provides an interpretation of the quantitative results summarized in Section 4 in the context of the study's research questions and hypotheses, places the results in the context of the general deep learning and medical-imaging literature, addresses clinical and workflow implications, and critically evaluates the limitations inherent in the experimental design before outlining directions for future research in Section 6.

5.1 Principal Findings

The results of this study directly demonstrate that deep learning architectures significantly and substantially outperform classical image-processing methods for tumor detection and segmentation on both MRI and CT (H0 rejected, $p < 0.001$, in all comparisons, RQ1 confirmed). Self-configuring CNN pipelines (nnU-Net) and hybrid CNN-Transformer architectures (Swin-UNETR) were found to have statistically indistinguishable top-tier performance, both significantly outdoing attention-augmented and plain CNN architectures on MRI, with attention-gated CNNs achieving a more appealing balance for CT nodule segmentation and for computational efficiency (RQ2, addressed). Performance was found to be trade-off between accuracy and efficiency, where the most accurate models were 2-3x slower in their inference time and 2-4x more number of parameters as compared to the most efficient architecture (RQ3, confirmed).

5.2 Why Self-Configuration and Attention Mechanisms Matter

The performance boost of nnU-Net is not expected to be the result of any single breakthrough in architecture, because the core of the network is the CNN encoder-decoder, but it is the combination of systematic optimisation of the patch size, network depth and normalisation strategy, which is optimised for each single dataset, that minimize the risk of sub-optimal manual hyperparameter tuning (a problem often overlooked when comparing hand-designed architectures). The strong relative results of Attention U-Net on CT may be because of the attention gates, which suppress background signal irrelevant to the anatomical structure, particularly important for small and low-contrast lesions like pulmonary nodules in close proximity to vascular structures.

5.3 Role of Transformer-Based Architectures

The inductive biases of convolutional architectures are still helpful, with their performance being competitive (not better) than that of Swin-UNETR for the relatively small-sized datasets studied here (in the hundreds instead of the tens of thousands of volumes), and the benefit of long-range self-attention could only be seen in a more massive training set or in cases of transfer learning from large-scale pre-trained foundation models. Overall, the under-segmentation of very small nodules observed in the Swin-UNETR experiment also suggests that the windowed self-attention tokenisation technique may require further refinement, such as reducing the size of the patch/window in the image or multi-scale token fusion, for sub-centimetre lesion segmentation applications.

5.4 Clinical and Workflow Implications

This improvement about 24-29 percentage points in Dice score and a drop in boundary error (HD95) from over 14-18 mm to under 5 mm is of clinical significance: the accuracy of the boundary achieved by leading DL models is approaching that of expert radiologists, who are reported to have an inter-observer variability of 5 mm. In practice, they argue for using DL-based segmentation as a first-pass, radiologist-assisted decision-support system that pre-populates contours for expert verification and refinement, rather than as an independent diagnostic system, because none of the models had 100% detection sensitivity, and there were still some false positives.

5.5 Limitations

There are some restrictions to be noted. First, the datasets used were curated subsets of the original datasets and results may not be generalizable to significantly different scanner vendor, acquisition protocol or patient populations without external validation. Second, the study focused on binary tumor-background segmentation, instead of clinically relevant multi-class sub-region segmentation (e.g., necrotic core, oedema, enhancing tumor for gliomas), which could limit the applicability of findings to more detailed clinical tasks. Third, computational cost was evaluated on high-end GPU devices (NVIDIA A100), which are less common in clinical settings, and the inference times on a typical clinical workstation may be longer. Fourth, although five-fold cross-validation and held-out testing were performed, sample sizes are reasonable for a comparative benchmarking study but not as large as some large-scale challenge cohorts which may influence the precision of effect-size estimates for close model comparisons (such as nnU-Net vs Swin-UNETR).

The relative performance of the architectures observed here (nnU-Net and Transformer-hybrid models at the top tier, followed by attention-augmented CNNs, followed by plain encoder-decoder CNNs, with the largest performance gap from classical methods) is largely similar to results on recent challenge leaderboards and other independent benchmarking studies in brain tumor and pulmonary nodule segmentation, which adds external validity to the present results and extends previous ones with the direct comparison across both MRI and CT modalities and the explicit accuracy-efficiency analysis.

For the best performing model, nnU-Net, the top three cases with the most errors were manually reviewed and three probable failure patterns were identified. Firstly, the margins of small, low-contrast infiltrative tumors were sometimes not fully segmented on FLAIR sequences, reflecting the intrinsic nature of the low specificity of these margins even among expert human annotators. Secondly, motion-artefact degraded CT slices led to localised over-segmentation, with streak artefacts appearing as areas of similar soft-tissue density close to the pleura. Third, multi-focal lesions (more than one spatially separated deposit of tumor masses in a single volume) were sometimes only partially detected, with the largest lesion segmented accurately but smaller satellite lesions not detected this has direct implications for staging accuracy in multi-focal disease, is an important target for future model refinement, for example with multi-instance detection heads.

The two datasets employed in this study are from multi-institutional public repositories but are limited to a subset of scanners, acquisition protocols and patient demographics, and therefore care must be taken in extrapolating the reported accuracy numbers to other populations (e.g., paediatric patients, recurrent tumours after treatment, and scanners from scanners not represented in the source cohorts). The ability of a model trained with one scanner/protocol to retain segmentation accuracy when used on another scanner/protocol (domain-shift robustness) was not evaluated in this study.

The post-hoc power analysis results showed that the large effect sizes observed for DL models when compared to the classic baseline (Cohen's d > 1.5) resulted in greater than 95% power for the tests performed, while the smaller effect size between nnU-Net and Swin-UNETR (Cohen's d approximately 0.3) yielded approximately 60% power for these tests, which is still not significant, as seen in Table 5. This suggests that the general trend of superior performance of deep learning over classical methods is general but that, finer-grained comparisons between the best two architectures should be taken with a grain of salt and replicated with larger cohorts of data.

6. Conclusion and Future Work

In this work, the experimental results of the benefit of deep learning based automated tumor detection and segmentation on MRI and CT images compared to the classical approaches, and find that deep learning yields significant differences with a high degree of statistical significance. The self-configuring nnU-Net pipeline performed best overall, whereas the hybrid CNN-Transformer Swin-UNETR performed statistically equally, but at a higher computational cost, while Attention U-Net provided the most favourable accuracy-efficiency ratio for the resource-limited deployment scenario. These results suggest that the selection of a model for clinical deployment should be done in a nuanced, task-specific manner, which does not consist of choosing one single model.

Future work should involve expansion of the comparative framework to multi-class sub-region segmentation, evaluation of generalisation on external, multi-institutional, multi-vendor datasets, incorporation of uncertainty quantification to flag low confidence predictions to ensure that they are reviewed by an expert and prospective clinical workflow studies measuring the impact on radiologist reading time and inter-observer variability when evaluating this dataset in real clinical settings with DL-assisted segmentation.

Future directions of research include: assessing the domain-shift robustness of the models across additional scanners and institutions not included in the training data; investigating lightweight model-compression and knowledge-distillation techniques to make the top-performing accuracy (currently associated with nnU-Net and Swin-UNETR) available within the inference-time budget of point-of-care or edge devices; testing the generalisability of the architecture ranking observed in the current literature across additional tumour sites (hepatic, prostatic, breast) and

additional imaging modalities (PET, ultrasound); exploring federated learning approaches to train models across multiple institutions without centralising sensitive patient imaging data, potentially improving generalisability while maintaining privacy. Finally, prospective, IRB approved clinical trials comparing radiologist performance with and without DL based segmentation assistance are required to move the increased retrospective accuracy seen here to better patient outcomes and a more efficient workflow.

REFERENCES

1. Spaanderman, D. J., Starmans, M. P., van Erp, G. C., Hanff, D. F., Sluijter, J. H., Schut, A. R. W., ... & Klein, S. (2025). Minimally interactive segmentation of soft-tissue tumors on CT and MRI using deep learning. *European radiology*, *35*(5), 2736-2745.
2. Rai, H. M. (2024). Cancer detection and segmentation using machine learning and deep learning techniques: A review. *Multimedia Tools and Applications*, *83*(9), 27001-27035.
3. Krishnapriya, S., & Karuna, Y. (2023). A survey of deep learning for MRI brain tumor segmentation methods: Trends, challenges, and future directions. *Health and Technology*, *13*(2), 181-201.
4. Karim, S., Tong, G., Yu, Y., Laghari, A. A., Khan, A. A., Ibrar, M., & Mehmood, F. (2024). Developments in brain tumor segmentation using MRI: deep learning insights and future perspectives. *IEEE Access*, *12*, 26875-26896.
5. Kashyap, M., Wang, X., Panjwani, N., Hasan, M., Zhang, Q., Huang, C., ... & Gensheimer, M. F. (2025). Automated deep learning–based detection and segmentation of lung tumors at ct imaging. *Radiology*, *314*(1), e233029.
6. Wesdorp, N. J., Zeeuw, J. M., Postma, S. C., Roor, J., van Waesberghe, J. H. T., van den Bergh, J. E., ... & Kazemier, G. (2023). Deep learning models for automatic tumor segmentation and total tumor volume assessment in patients with colorectal liver metastases. *European radiology experimental*, *7*(1), 75.
7. Illimoottil, M., & Ginat, D. (2023). Recent advances in deep learning and medical imaging for head and neck cancer treatment: MRI, CT, and PET scans. *Cancers*, *15*(13), 3267.
8. Hussain, D., Al-Masni, M. A., Aslam, M., Sadeghi-Niaraki, A., Hussain, J., Gu, Y. H., & Naqvi, R. A. (2024). Revolutionizing tumor detection and classification in multimodality imaging based on deep learning approaches: Methods, applications and limitations. *Journal of X-Ray Science and Technology*, *32*(4), 857-911.
9. Ahsan, R., Shahzadi, I., Najeeb, F., & Omer, H. (2025). Brain tumor detection and segmentation using deep learning. *Magnetic Resonance Materials in Physics, Biology and Medicine*, *38*(1), 13-22.
10. Han, J. H., Kim, B. W., Kim, T. M., Ko, J. Y., Choi, S. J., Kang, M., ... & Jeong, C. W. (2025). Fully automated segmentation and classification of renal tumors on CT scans via machine learning. *BMC cancer*, *25*(1), 173.
11. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI), 234-241.
12. Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., ... & Rueckert, D. (2018). Attention U-Net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.
13. Jha, D., Smedsrud, P. H., Riegler, M. A., Johansen, D., De Lange, T., Halvorsen, P., & Johansen, H. D. (2019). ResUNet++: An advanced architecture for medical image segmentation. IEEE International Symposium on Multimedia (ISM), 225-2255.
14. Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods, 18(2), 203-211.
15. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H. R., & Xu, D. (2022). Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. International MICCAI Brainlesion Workshop, 272-284.
16. Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., ... & Van Leemput, K. (2015). The multimodal brain tumor image segmentation benchmark (BraTS). IEEE Transactions on Medical Imaging, 34(10), 1993-2024.
17. Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., ... & Bakas, S. (2021). The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv preprint arXiv:2107.02314.
18. Armato III, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., ... & Clarke, L. P. (2011). The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans. Medical Physics, 38(2), 915-931.
19. Setio, A. A. A., Traverso, A., De Bel, T., Berens, M. S., Van Den Bogaard, C., Cerello, P., ... & Jacobs, C. (2017). Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules and false positive reduction in CT: The LUNA16 challenge. Medical Image Analysis, 42, 1-13.
20. Bilic, P., Christ, P., Li, H. B., Vorontsov, E., Ben-Cohen, A., Kaissis, G., ... & Menze, B. (2023). The Liver Tumor Segmentation Benchmark (LiTS). Medical Image Analysis, 84, 102680.
21. Cicek, O., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning dense volumetric segmentation from sparse annotation. MICCAI, 424-432.
22. Milletari, F., Navab, N., & Ahmadi, S. A. (2016). V-Net: Fully convolutional neural networks for volumetric medical image segmentation. International Conference on 3D Vision (3DV), 565-571.

23. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., ... & Zhou, Y. (2021). TransUNet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306.
24. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., ... & Xu, D. (2022). UNETR: Transformers for 3D medical image segmentation. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 574-584.
25. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sanchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60-88.
26. Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P. F., Kohl, S., ... & Maier-Hein, K. H. (2018). nnU-Net: Self-adapting framework for U-Net-based medical image segmentation. arXiv preprint arXiv:1809.10486.
27. Zhou, S. K., Greenspan, H., Davatzikos, C., Duncan, J. S., Van Ginneken, B., Madabhushi, A., ... & Summers, R. M. (2021). A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. Proceedings of the IEEE, 109(5), 820-838.
28. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. International Conference on Learning Representations (ICLR).
29. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin Transformer: Hierarchical vision transformer using shifted windows. IEEE/CVF International Conference on Computer Vision (ICCV), 10012-10022.
30. Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B. A., ... & Cardoso, M. J. (2022). The Medical Segmentation Decathlon. Nature Communications, 13(1), 4128.