IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151090

RSAA-UNet: A Residual Squeeze-and-Excitation Attention U-Net with Atrous Context and Focal-Tversky Supervision for Lower-Grade Glioma Segmentation on FLAIR MRI under a Patient-Wise Evaluation Protocol

G. Appasami[1*], A. Selvakumar[2]

Associate Professor, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamil Nadu, India[1]

ORCID: 0000-0002-4186-3166[1]

Professor, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamil Nadu, India[2]

ORCID: 0000-0002-4779-0973[2]

*Corresponding Author

Abstract: Outlining lower-grade glioma (LGG) on FLAIR MRI is a prerequisite for volumetric follow-up and radiogenomic work, yet many published U-Net pipelines for the public TCGA-LGG collection do not state, or do not enforce, a clean separation of patients between training and testing. This paper describes RSAA-UNet, a 2-D encoder-decoder that combines residual double-convolution blocks, squeeze-and-excitation (SE) recalibration, additive attention gates on all four skip connections, an atrous spatial pyramid pooling (ASPP) module at the bottleneck, and deep supervision at three decoder levels, trained with a Focal-Tversky plus binary cross-entropy objective. Ten networks (three baselines, two RSAA-UNet variants and five single-component ablations) were trained under one protocol on a patient-wise 78/16/16 split of 110 patients (3,929 slices) and evaluated on 16 unseen patients (498 slices, 182 containing tumour). Flip-based test-time augmentation and small-component removal were examined as inference options. On tumour-containing slices the full RSAA-UNet with both inference steps reached a Dice coefficient of 0.715 (patient-bootstrap 95% CI 0.651–0.771), against 0.714 for U-Net and 0.743 for the strongest baseline, a residual U-Net; no difference was significant at the patient level after Holm correction. The proposed objective lowered the false-positive rate on tumour-free slices to 2.5%, compared with 4.4–6.0% for the baselines, and the ablation pointed to residual connections and SE recalibration as the most useful components, while attention gates and ASPP gave no measurable benefit here. Training of the full model also diverged after epoch 20, so its weights come from an early checkpoint. The Discussion explains why a strict patient-wise protocol gives lower Dice values than those usually quoted for this dataset, compares the results with five published studies, and lists the changes needed before the architecture can be judged conclusively.

Keywords: lower-grade glioma; brain MRI segmentation; U-Net; attention gate; squeeze-and-excitation; atrous spatial pyramid pooling; Focal-Tversky loss; patient-wise evaluation

1. INTRODUCTION

Lower-grade gliomas (LGG), the WHO grade 2 and grade 3 diffuse gliomas, are infiltrative tumours that tend to recur and can evolve into higher-grade lesions (Buda et al., 2019). Their classification now rests largely on molecular markers rather than histology alone (Cancer Genome Atlas Research Network, 2015; Louis et al., 2021), but imaging remains the first and often the only source of information available before any tissue is taken. In LGG, contrast enhancement is uncommon, so the extent of the disease is usually read from the hyperintense abnormality on fluid-attenuated inversion recovery (FLAIR) images (Buda et al., 2019).

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151090

Delineating that abnormality is more than a convenience. Volume and shape descriptors computed from the outline have been linked to genomic subtypes and to patient outcome (Buda et al., 2019; Mazurowski et al., 2017), and serial volumes are a natural way to follow a slowly growing lesion. Manual contouring is slow and varies between readers; for whole-tumour LGG the Dice overlap between two experts has been put at roughly 0.84 Menze et al. (2015), as quoted by Buda et al. (2019). An automatic method therefore has to be judged against a ceiling that is itself below one.
Convolutional encoder-decoder networks, and the U-Net in particular (Ronneberger et al., 2015), have become the default tool for this kind of task (Akkus et al., 2017; Litjens et al., 2017; Mazurowski et al., 2019; Siddique et al., 2021). Most of the brain-tumour literature is built on the multi-sequence, largely glioblastoma-oriented BraTS benchmark (Bakas et al., 2017; Ghaffari et al., 2020; Menze et al., 2015). The TCGA-LGG FLAIR collection released by Buda et al. (2019) is smaller, two-dimensional in practice, and dominated by diffuse low-grade disease, which is exactly why it is popular for quick experiments and why results obtained on it deserve a careful reading.

Two weaknesses recur in work on this collection. The first concerns architecture. Residual connections, channel attention, attention gates, multi-scale context and deep supervision have each been shown to help in some setting, but they are usually introduced one at a time, on different data and with different training recipes. It is seldom clear how much each one contributes once the others are present. The second concerns evaluation. Neighbouring slices of one patient are almost copies of each other, so a random split at slice level places near-duplicates on both sides of the train-test boundary and inflates the score (Varoquaux & Cheplygina, 2022). In addition, about two thirds of the slices in this dataset contain no tumour at all, and whether such slices are scored, skipped or counted as perfect changes the headline number. Test sets are small, and confidence intervals are rarely reported.

This study addresses both points with a single, deliberately controlled experiment. We assemble the components listed above into one configurable network, RSAA-UNet (Residual, Squeeze-and-excitation, Attention, ASPP U-Net), and train it together with three baselines and five single-component ablations under identical settings. All models are scored on 16 test patients who never appear in training or validation. The contributions are as follows.

(i) A configurable 2-D architecture in which residual connections, SE recalibration, attention gates, ASPP and deep supervision can be switched on and off independently, so that baselines and ablations share one code path.
(ii) A Focal-Tversky plus cross-entropy objective, applied at the output and at three auxiliary decoder levels, whose effect on precision, recall and false positives on tumour-free slices is measured directly.
(iii) A leakage-free evaluation protocol with patient-level bootstrap intervals, paired non-parametric tests with Holm correction, tumour-size strata and a separate account of tumour-free slices.
(iv) An open report of what did not work: the full model does not beat the best baseline, its training became unstable after epoch 20, and several components fail to earn their parameters. We think such a report is more useful to later users of the dataset than another uncontrolled improvement.

The remainder of the paper is organised as follows. Section 2 reviews related work. Section 3 describes the data, the network, the losses and the evaluation protocol. Section 4 presents the results, Section 5 places them next to published figures, Section 6 discusses them and their limits, and Section 7 concludes.

2. RELATED WORK

2.1 Encoder-decoder networks for medical images

Fully convolutional networks (Long et al., 2015) and SegNet (Badrinarayanan et al., 2017) established dense, pixel-wise prediction with learned up-sampling. The U-Net (Ronneberger et al., 2015) added skip connections that carry high-resolution encoder features straight to the decoder, which proved decisive for small training sets; the importance of those long skips was later analysed explicitly by Drozdzal et al. (2016). Volumetric versions followed (Milletari et al., 2016; Çiçek et al., 2016), as did nested and densely connected skip pathways (Zhou et al., 2020) and the self-configuring nnU-Net pipeline, whose strength lies mostly in careful automatic choice of pre-processing, training and post-processing (Isensee et al., 2021). Replacing plain convolutions by residual units (He et al., 2016) gave the residual U-Net (Zhang et al., 2018), a variant that remains a strong baseline in our experiments. More recently, transformer components have been placed in the encoder, the decoder or both (Cao et al., 2023; Chen et al., 2024; Hatamizadeh et al., 2022a; Hatamizadeh et al., 2022b); a broad overview of U-Net variants is given by Siddique et al. (2021).

For brain tumours specifically, early convolutional networks worked patch by patch (Havaei et al., 2017; Pereira et al., 2016), multi-scale 3-D networks with conditional random fields followed (Kamnitsas et al., 2017), and U-Net derivatives soon took over (Dong et al., 2017). Cross-dataset performance evaluation and secure federated learning(Appasami, G. et al., 2025) shows recent advancement. The best BraTS entries rely on large, regularised encoder-decoders such as the autoencoder-regularised network of Myronenko (2019). Those systems use four MRI

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151090

sequences, three-dimensional context and several tumour sub-regions, which makes them hard to compare with a 2-D, single-region problem such as the LGG FLAIR task.

2.2 Attention, multi-scale context and deep supervision

Squeeze-and-excitation (SE) blocks re-weight feature channels with a global descriptor and add little computation (Hu et al., 2020); spatial and channel variants were adapted to segmentation by Roy et al. (2018). Attention gates compute a spatial coefficient from a skip feature and a coarser gating signal and suppress irrelevant regions before concatenation (Oktay et al., 2018; Schlemper et al., 2019). Atrous spatial pyramid pooling (ASPP) applies parallel dilated convolutions to capture several receptive-field sizes at once (Chen et al., 2018a; Chen et al., 2018b), which is attractive when lesions range from a few millimetres to a large part of a hemisphere. Auxiliary outputs at intermediate decoder levels, known as deep supervision, shorten the gradient path and were used in the nested U-Net of Zhou et al. (2020). A novel lightweight CNN design (Appasami, G. et al., 2025) for MRI brain tumor image classification with performance-driven optimization makes platform for mobile device computation. Each of these ideas has a track record, but a track record in one task does not guarantee a gain in another, and that is the question tested here.

2.3 Objectives for imbalanced segmentation

Lesions occupy a small fraction of the image, so a plain pixel-wise cross-entropy tends to favour the background; the effect of imbalance in convolutional networks has been studied systematically by Buda et al. (2018). Overlap-based losses derived from the Dice coefficient (Dice, 1945; Milletari et al., 2016; Sudre et al., 2017) counter it directly. The Tversky index weights false negatives and false positives separately (Salehi et al., 2017), and raising the resulting loss to a power gives the focal Tversky loss (Abraham & Khan, 2019), in analogy with the focal loss for detection (Lin et al., 2020). Comparative studies find that no single loss wins everywhere and that compound losses are often the safest choice (Jadon, 2020; Ma et al., 2021). We follow that advice by pairing the Focal-Tversky term with cross-entropy.

2.4 Work on the TCGA-LGG FLAIR collection and on evaluation practice

Buda et al. (2019) released the collection and reported a mean Dice of 0.82 under 22-fold patient-wise cross-validation with a U-Net, skull stripping and three-dimensional post-processing. Later studies reported 0.84 with a U-Net and VGG16 transfer learning (Naser & Deen, 2020), 0.83 for a U-Net/SegNet hybrid (Jlassi et al., 2023), and values above 0.91 for a UNet++ with a pretrained encoder (Yadav et al., 2024) and for a “Super Vision” U-Net variant (Güzel & Aydın, 2025). These numbers are not directly comparable, because the splitting unit, the treatment of empty slices and the pre-processing differ or are not fully stated. Reviews of medical-imaging machine learning list leakage between training and test data as one of the commonest reasons for over-optimistic results (Varoquaux & Cheplygina, 2022), and recent guidance on metrics asks authors to report the unit of analysis, uncertainty and the handling of empty cases (Maier-Hein et al., 2024; Taha & Hanbury, 2015). Our protocol is built to meet those recommendations.

3. MATERIALS AND METHODS

Fig. 1 summarises the experimental pipeline. The same code path produces every network, so differences between rows of the result tables come from the stated architectural or loss change and from nothing else.

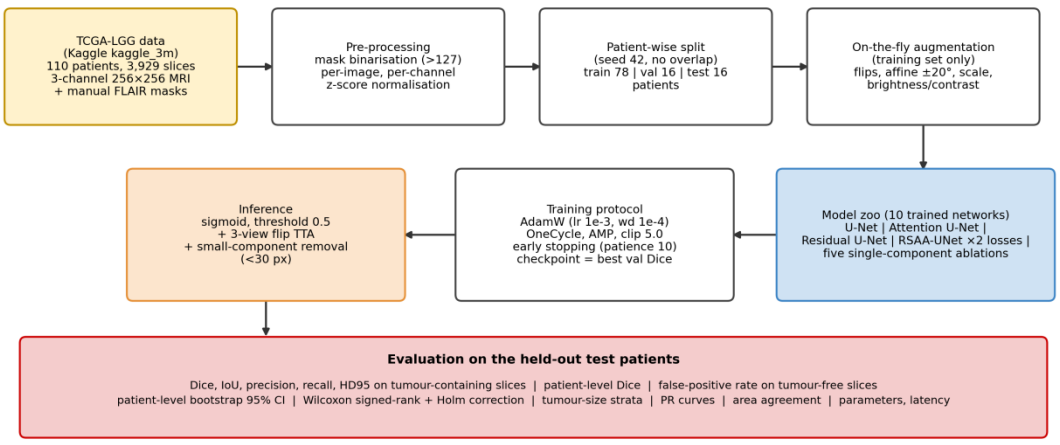


Figure 1. Overall experimental framework: data preparation with a patient-wise split, model zoo, common training protocol, inference options and evaluation.

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151090

3.1 Dataset

The data are the TCGA-LGG FLAIR segmentation collection in the Kaggle release kaggle_3m (Buda et al., 2019), derived from preoperative studies of 110 patients with lower-grade glioma in The Cancer Genome Atlas, whose images are hosted by The Cancer Imaging Archive (Clark et al., 2013; Pedano et al., 2016). The patients come from five institutions. The FLAIR abnormality was outlined slice by slice by a trained researcher and every outline was checked by a board-eligible radiologist (Buda et al., 2019). Each image is stored as a three-channel 256 × 256 slice; where all sequences were available the channels carry the pre-contrast, FLAIR and post-contrast images, and where some were missing neighbouring FLAIR slices fill the gaps (Buda et al., 2019). Image and mask files were paired by file name, giving 3,929 slices. Of these, 1,373 (34.9%) contain tumour, and on those slices the tumour covers on average 2.95% of the image. Fig. 4 shows the distribution of slices per patient, of the share of tumour slices per patient and of tumour area, and Fig. 5 shows example slices with their outlines.

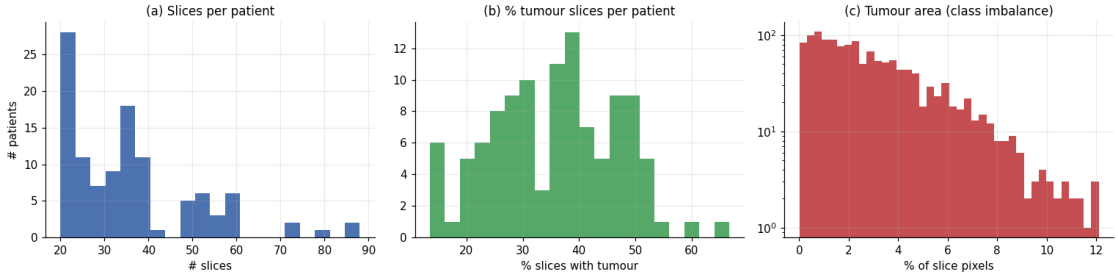


Figure 4. Dataset overview: (a) slices per patient, (b) percentage of tumour-containing slices per patient, (c) tumour area on tumour slices (logarithmic count axis), which illustrates the class imbalance.

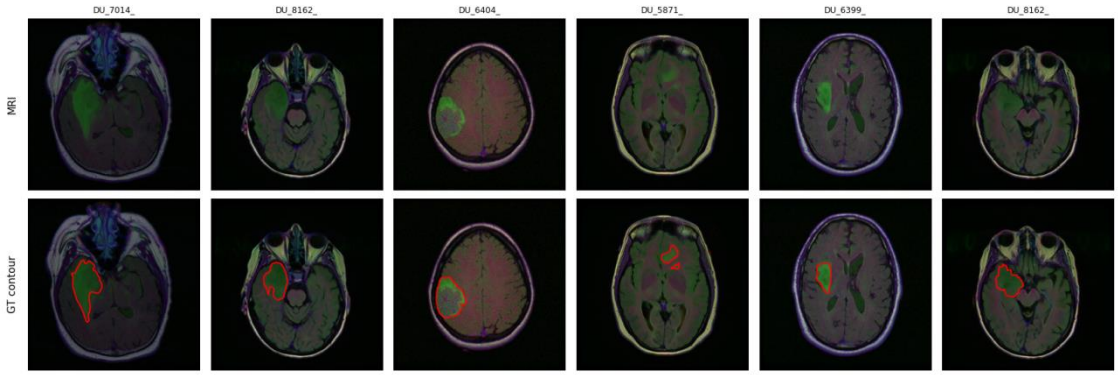


Figure 5. Six randomly drawn tumour slices (top) and the same slices with the manual outline in red (bottom).

3.2 Pre-processing and patient-wise split

Masks were binarised at an intensity of 127. Each image was standardised channel by channel,

$$\tilde{x}_c(i,j) = \frac{x_c(i,j) - \mu_c}{\sigma_c + 10^{-6}}, \qquad c \in \{1,2,3\}, \tag{1}$$

where μ_c and σ_c are the mean and standard deviation of channel c of that image. No skull stripping, bias-field correction or histogram matching was applied beyond what the release already contains.

Because slices of one patient are strongly correlated, the data were divided by patient. The 110 patient identifiers were shuffled with a fixed seed (42) and assigned to a test group (15%), a validation group (15%) and a training group (the rest). After rounding, this gave 16, 16 and 78 patients. A programmatic check confirmed that no patient appears in more than one group. Table 1 lists the resulting numbers. The test patients come from four institutions (four from TCGA-CS, three from DU, three from FG and six from HT, judged from the patient identifiers).

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151090

Table 1. Patient-wise split of the TCGA-LGG collection.

Subset	Patients	Slices	Tumour slices	Tumour-free slices
Training	78	2,774	919	1,855
Validation	16	657	272	385
Test	16	498	182	316
Total	**110**	**3,929**	**1,373**	**2,556**

3.3 Data augmentation

Augmentation was applied on the fly to the training images only. Horizontal and vertical flips were each applied with probability 0.5. With probability 0.7 an affine transform combined a rotation drawn uniformly from ±20°, an isotropic scale between 0.9 and 1.1 and a translation of up to 5% of the image size in each direction (bilinear interpolation for the image, nearest neighbour for the mask). With probability 0.5 the image was multiplied by a factor between 0.85 and 1.15 and shifted by up to ±15 grey levels. These operations are common in medical-image work (Shorten & Khoshgoftaar, 2019) and were chosen to mimic plausible variation in patient position and scanner gain.

3.4 The RSAA-UNet architecture

Fig. 2 shows the complete network and Fig. 3 shows its three building blocks in detail. A single class implements all models; five switches (residual, SE, attention, ASPP, deep supervision) turn the corresponding component on or off. Setting all of them to off yields the plain U-Net used as the first baseline, and the base width is 32 channels.

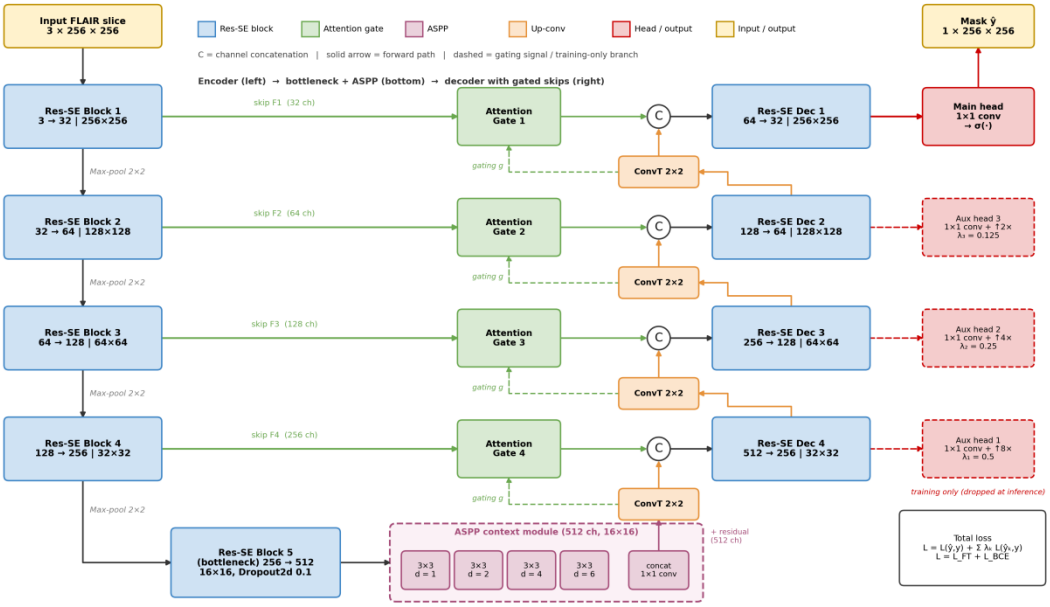


Figure 2. Complete architecture of RSAA-UNet. The encoder has four Res-SE blocks (32–256 channels) joined by max-pooling. A fifth Res-SE block and an ASPP module form the bottleneck at 16 × 16. Each decoder stage up-samples with a 2 × 2 transposed convolution, filters the encoder skip feature with an attention gate that is driven by the up-sampled tensor, concatenates the two, and refines them with a Res-SE block. Three auxiliary heads supply deep supervision during training and are removed at inference.

Encoder and decoder. The encoder has four stages with 32, 64, 128 and 256 channels at 256 × 256, 128 × 128, 64 × 64 and 32 × 32 pixels, separated by 2 × 2 max-pooling. The bottleneck block widens the representation to 512 channels at 16 × 16. The decoder mirrors the encoder with four transposed convolutions (kernel 2, stride 2). Spatial dropout with probability 0.1 is used in the bottleneck block and in the two deepest decoder blocks. A 1 × 1 convolution maps the last decoder feature to one logit per pixel.

Residual SE block. Each block contains two 3 × 3 convolutions with batch normalisation (Ioffe & Szegedy, 2015), an SE unit and a shortcut. With $\delta(\cdot)$ denoting ReLU, W_1 and W_2 the two 3 × 3 kernels, $\mathcal{S}(\cdot)$ the shortcut (identity when

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151090

input and output widths agree, otherwise a 1 × 1 convolution followed by batch normalisation) and $\mathrm{Drop}(\cdot)$ the spatial dropout,

$$y = \mathrm{Drop}\left(\delta\left(\mathrm{SE}(\mathcal{F}(x)) + \mathcal{S}(x)\right)\right), \qquad \mathcal{F}(x) = \mathrm{BN}\left(W_2 * \delta\left(\mathrm{BN}(W_1 * x)\right)\right). \tag{2}$$

The SE unit squeezes each channel of $F \in \mathbb{R}^{C\times H\times W}$ to a scalar and then re-weights the channels,

$$z_c = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} F_c(i,j), \qquad s = \sigma\left(W_b\,\delta(W_a z)\right), \qquad \mathrm{SE}(F)_c = s_c\,F_c, \tag{3}$$

with $W_a \in \mathbb{R}^{C_r\times C}$, $W_b \in \mathbb{R}^{C\times C_r}$, $C_r = \max(C/8,4)$ and σ the logistic sigmoid. The reduction ratio of 8 follows Hu et al. (2020).

Attention gate. At decoder stage l the skip feature x from the encoder and the up-sampled decoder tensor g have the same spatial size and the same number of channels. The gate projects both to $C_l/2$ channels with 1 × 1 convolutions and batch normalisation, adds them, applies ReLU, projects to one channel and passes the result through a sigmoid,

$$\alpha = \sigma\left(\mathrm{BN}\left(\psi * \delta\left(\mathrm{BN}(W_g * g) + \mathrm{BN}(W_x * x)\right)\right)\right), \qquad \hat{x} = \alpha \odot x, \tag{4}$$

where $\odot$ broadcasts the single-channel map α over all channels of x. The gated feature $\hat{x}$ is concatenated with g and fed to the decoder block. This is the additive formulation of Oktay et al. (2018), with one difference: the original uses a gating signal from a coarser level, whereas here the already up-sampled tensor is used, which removes the need for resampling inside the gate.

Atrous spatial pyramid pooling. At the 16 × 16 bottleneck, four parallel 3 × 3 convolutions with dilation rates 1, 2, 4 and 6 each produce 128 channels. Their outputs are concatenated, fused by a 1 × 1 convolution and added back to the input,

$$A_r(F) = \delta\left(\mathrm{BN}\left(W_{3\times3}^{(r)} * F\right)\right), \qquad F_{\mathrm{ASPP}} = F + \delta\left(\mathrm{BN}\left(W_p * [A_1(F); A_2(F); A_4(F); A_6(F)]\right)\right), \tag{5}$$

where $[\cdot\,;\cdot]$ is channel concatenation and the superscript (r) marks the dilation. The rates were chosen so that the widest receptive field still fits inside the 16 × 16 map. No image-level pooling branch is used. The module adds about 2.6 million parameters.

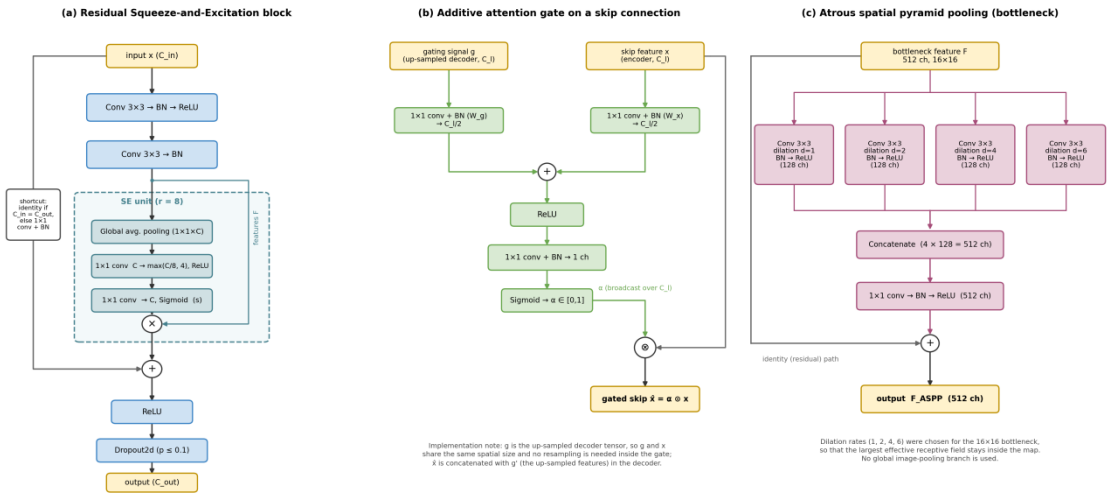


Figure 3. Internal structure of (a) the residual SE block, (b) the additive attention gate and (c) the ASPP module used at the bottleneck.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151090

Deep supervision. During training, the outputs of the three deepest decoder stages (32 × 32, 64 × 64 and 128 × 128 pixels) are mapped to one channel by 1 × 1 convolutions and bilinearly resized to 256 × 256, giving auxiliary predictions $\hat{y}_1, \hat{y}_2, \hat{y}_3$. They are discarded at inference. With all components active the network has 10.94 million trainable parameters.

3.5 Loss functions

Let $p_n = \sigma(\ell_n)$ be the predicted foreground probability of pixel n and $y_n \in \{0,1\}$ its label. The pixel-wise cross-entropy, averaged over all pixels of a mini-batch, is

$$\mathcal{L}_{\mathrm{BCE}} = -\frac{1}{N}\sum_{n=1}^{N}[y_n \log p_n + (1-y_n)\log(1-p_n)]. \tag{6}$$

The baseline objective adds the soft Dice loss computed per image and then averaged over the batch,

$$\mathcal{L}_{\mathrm{Dice}} = 1 - \frac{2\sum_n p_n\, y_n + \epsilon}{\sum_n p_n + \sum_n y_n + \epsilon}, \qquad \mathcal{L}_{\mathrm{BCE+Dice}} = \mathcal{L}_{\mathrm{BCE}} + \mathcal{L}_{\mathrm{Dice}}, \tag{7}$$

with $\epsilon = 1$ for numerical stability and for a defined value on empty slices. The proposed objective replaces the Dice term by a focal Tversky term. With $\mathrm{TP} = \sum_n p_n\, y_n$, $\mathrm{FN} = \sum_n (1-p_n)\, y_n$ and $\mathrm{FP} = \sum_n p_n\, (1-y_n)$ per image,

$$\mathrm{TI} = \frac{\mathrm{TP} + \epsilon}{\mathrm{TP} + \alpha\,\mathrm{FN} + \beta\,\mathrm{FP} + \epsilon}, \qquad \mathcal{L}_{\mathrm{FT}} = (1-\mathrm{TI})^{\gamma}, \qquad \mathcal{L}_{\mathrm{seg}} = \mathcal{L}_{\mathrm{FT}} + \mathcal{L}_{\mathrm{BCE}}, \tag{8}$$

where $\alpha = 0.7$ penalises missed tumour more strongly than $\beta = 0.3$ penalises false alarms, and the exponent is $\gamma = 0.75$. These settings are close to those reported for the focal Tversky loss by Abraham and Khan (2019). Networks with deep supervision minimise the weighted sum

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{seg}}(\hat{y}, y) + \sum_{k=1}^{3} \lambda_k\, \mathcal{L}_{\mathrm{seg}}(\hat{y}_k, y), \qquad (\lambda_1, \lambda_2, \lambda_3) = (0.5,\ 0.25,\ 0.125), \tag{9}$$

where $\hat{y}_1$ is the auxiliary output of the deepest decoder stage. For the baselines and the "BCE+Dice" variant, $\mathcal{L}_{\mathrm{seg}}$ in Eq. (9) is replaced by $\mathcal{L}_{\mathrm{BCE+Dice}}$.

3.6 Training protocol and implementation

All networks were trained with AdamW (Loshchilov & Hutter, 2019), whose update is

$$\theta_{t+1} = \theta_t - \eta_t\left(\frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon} + \lambda_w\,\theta_t\right), \tag{10}$$

where $\hat{m}_t$ and $\hat{v}_t$ are the bias-corrected first and second moment estimates, $\lambda_w = 10^{-4}$ is the decoupled weight decay and η_t follows a one-cycle schedule (Smith & Topin, 2019) with peak 10^{-3} reached after 15% of the planned steps. The schedule is defined for 50 epochs and is not shortened when training stops early. Mixed-precision arithmetic (Micikevicius et al., 2018) and gradient-norm clipping at 5.0 were used. The checkpoint with the highest validation Dice was kept, and training stopped after 10 epochs without improvement. Validation Dice counts a correctly empty prediction on a tumour-free slice as 1. Table 2 lists the remaining settings. Each configuration was trained once with seed 42; no repeated runs or cross-validation folds were performed.

Table 2. Implementation settings shared by all experiments.

Item	Setting
Input size / channels	256 × 256 / 3
Base width; widths per level	32; 32, 64, 128, 256 (bottleneck 512)
SE reduction; ASPP dilation rates	8; 1, 2, 4, 6
Dropout (spatial)	0.1 in bottleneck and two deepest decoder blocks

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151090

Item	Setting
Optimiser	AdamW, weight decay 1e-4
Learning-rate schedule	One-cycle, peak 1e-3, 15% warm-up, 50 epochs planned
Batch size; epochs (max.); patience	16; 50; 10 (validation Dice)
Gradient clipping; precision	Norm 5.0; automatic mixed precision
Focal-Tversky parameters	$\alpha = 0.7$, $\beta = 0.3$, exponent 0.75, $\varepsilon = 1$
Deep-supervision weights	0.5, 0.25, 0.125
Decision threshold	0.5
Random seed	42 (one run per configuration)
Software / hardware	PyTorch 2.11 (CUDA 12.8); single GPU on the Kaggle platform

3.7 Inference: test-time augmentation and post-processing

At test time the logit map is converted to a probability and thresholded at 0.5. Two optional refinements were evaluated on the full model. Test-time augmentation (Wang et al., 2019) averages the predictions for the original image and for its horizontally and vertically flipped copies after flipping them back,

$$\bar{p}(x) = \frac{1}{3}\Big(f(x) + H\big(f(H(x))\big) + V\big(f(V(x))\big)\Big), \tag{11}$$

where f is the network followed by the sigmoid and H, V are the horizontal and vertical flips (each its own inverse). Post-processing removes connected components (8-connectivity) smaller than $\tau = 30$ pixels,

$$\hat{Y}_{\mathrm{PP}} = \bigcup_{k:\,|C_k| \geq \tau} C_k, \tag{12}$$

where C_k are the components of the binary prediction. This is a two-dimensional, per-slice operation; it is not the three-dimensional largest-component rule used by Buda et al. (2019).

3.8 Evaluation metrics and statistical analysis

Metrics were computed per slice on the original 256 × 256 grid. With $\hat{Y}$ the predicted and Y the reference mask,

$$\mathrm{Dice} = \frac{2\left|\hat{Y} \cap Y\right|}{\left|\hat{Y}\right| + |Y|} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}, \qquad \mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}, \tag{13}$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \qquad \mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{14}$$

The 95th-percentile Hausdorff distance (Huttenlocher et al., 1993; Taha & Hanbury, 2015) is computed on the mask boundaries $\partial\hat{Y}$ and ∂Y, with $d(a, B)$ the Euclidean distance from point a to set B and P_{95} the 95th percentile of the pooled distances,

$$\mathrm{HD95} = P_{95}\big(\{d(a, \partial Y) : a \in \partial\hat{Y}\} \cup \{d(b, \partial\hat{Y}) : b \in \partial Y\}\big). \tag{15}$$

The distance is given in pixels, because the images are distributed without pixel-spacing information that we used. All headline figures are averages over the tumour-containing slices of the test set, which keeps the scores from being diluted by the many easy tumour-free slices. Precision is undefined when nothing is predicted, and HD95 is undefined when either mask is empty; those slices are left out of the respective averages, which means HD95 does not penalise missed tumours. In addition, three complementary quantities are reported: the patient-level Dice (the mean of per-patient mean Dice values), the false-positive rate on tumour-free slices,

$$\mathrm{FPR}_{\mathrm{empty}} = \frac{|\{k : Y_k = \varnothing \ \wedge\ \hat{Y}_k \neq \varnothing\}|}{|\{k : Y_k = \varnothing\}|}, \tag{16}$$

and Dice within terciles of tumour area (small, medium, large), defined from the test-set distribution of tumour area on tumour slices.

Uncertainty was estimated by resampling patients, not slices, with replacement. For bootstrap sample b with patient multiset B_b, S_i the sum of slice Dice values of patient i and n_i the number of its tumour slices,

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151090

$$\mathrm{Dice}^{(b)} = \frac{\sum_{i \in B_b} S_i}{\sum_{i \in B_b} n_i}, \qquad \mathrm{CI}_{95\%} = \left[P_{2.5}\{\mathrm{Dice}^{(b)}\},\ P_{97.5}\{\mathrm{Dice}^{(b)}\}\right], \tag{17}$$

using 2,000 resamples (Efron, 1979). The full model was compared with every other configuration by the paired Wilcoxon signed-rank test (Wilcoxon, 1945), once on slice-level Dice and once on patient-level mean Dice. Because 11 comparisons were made, the p-values were adjusted by the Holm step-down procedure (Holm, 1979): with m tests and ordered raw values $p_{(1)} \le \cdots \le p_{(m)}$,

$$\tilde{p}_{(i)} = \min\left(1,\ \max_{j \le i} (m - j + 1)\, p_{(j)}\right). \tag{18}$$

The slice-level test treats slices of one patient as independent and is therefore anti-conservative; the patient-level test is the one on which conclusions are based. Pixel-level precision-recall curves and average precision were computed on a random sample of three million test pixels, and the agreement between predicted and reference tumour area was summarised by the Pearson correlation. Parameters and the latency of a single 256 × 256 forward pass (batch size one, full precision, mean of 50 runs after 10 warm-up runs, GPU synchronised) were measured for every network. The latency of the test-time-augmented model is three times the single-pass value and was not timed separately.

4. RESULTS

4.1 Training behaviour

Table 3 and Fig. 6 summarise training. The residual U-Net and the variant without residual connections used all 50 epochs, the U-Net stopped after 44, and the remaining networks stopped between epoch 15 and epoch 30. The two complete RSAA-UNet versions and the variants without SE, ASPP or attention reached a validation Dice of 0.883–0.891, above the baselines (0.860–0.878). The variant without deep supervision (0.847, best epoch 5) and the variant without residual connections (0.867) did not.

The full model behaved well until epoch 20, where it reached its best validation Dice of 0.8836. The training loss then rose from 0.296 (epoch 20) to 0.932 and 2.028 in the next two epochs and was undefined (NaN) from epoch 23 onwards. The constant validation Dice of 0.586 after that point equals the share of tumour-free validation slices (385 of 657), which is what an empty prediction everywhere would score. Training stopped after the patience of ten epochs was exhausted, and the checkpoint of epoch 20 was used for all test results. At that epoch the scheduled learning rate was still roughly four-fifths of its peak value, so the saved weights are a mid-training snapshot and not a converged model. No other configuration showed this behaviour. We did not isolate the cause (see Section 6).

Table 3. Training summary. Validation Dice counts correctly empty predictions on tumour-free slices as 1.

Model	Epochs run	Best epoch	Best val. Dice	Best val. IoU	Note
U-Net	44	34	0.8702	0.8309	Early stopping
Attention U-Net	28	18	0.8597	0.8199	Early stopping
Residual U-Net	50	47	0.8777	0.8411	Full schedule
RSAA-UNet (BCE+Dice)	26	16	0.8857	0.8462	Early stopping
RSAA-UNet (Full)	30	20	0.8836	0.8433	Loss NaN from epoch 23
w/o SE	20	10	0.8827	0.8461	Early stopping
w/o ASPP	30	20	0.8905	0.8548	Early stopping
w/o Attention	28	18	0.8873	0.8498	Early stopping
w/o Deep supervision	15	5	0.8466	0.8051	Early stopping
w/o Residual	50	46	0.8672	0.8326	Full schedule

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151090

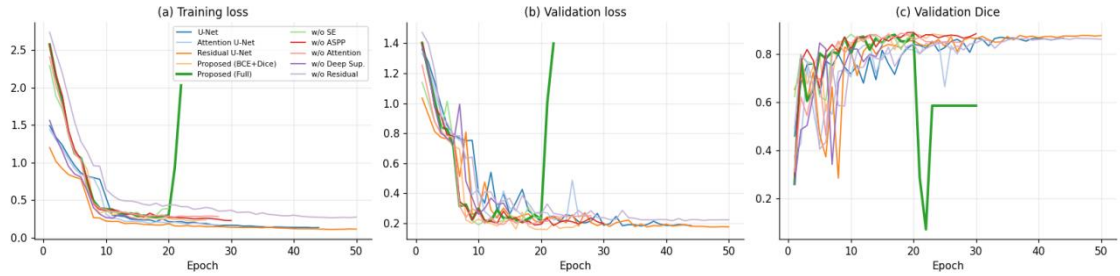


Figure 6. Training loss, validation loss and validation Dice for all ten networks. The thick green line is the full RSAA-UNet; its divergence after epoch 20 is visible in all three panels.

4.2 Comparison with the baselines

Table 4 reports the test results on the 182 tumour-containing slices of the 16 unseen patients, and Figs. 7 and 8 show the same information graphically. The residual U-Net was the strongest model on most overlap measures, with a Dice of 0.7434, an IoU of 0.6424, a recall of 0.8055, a patient-level Dice of 0.7422 and the lowest HD95 (13.43 px). The U-Net reached 0.7137 and the Attention U-Net 0.6993. The RSAA-UNet trained with BCE and Dice loss reached 0.7206, which is 0.7 Dice points above the U-Net, whereas the same network trained with the Focal-Tversky objective gave 0.7120. Flip-based test-time augmentation raised that value to 0.7146, and adding small-component removal gave 0.7151, a total gain of 0.31 points over the single pass. The same two steps raised precision from 0.8110 to 0.8194 and lowered HD95 from 16.73 to 15.93 px, while recall slipped from 0.7452 to 0.7404.

The Dice standard deviations are large (0.25–0.33), because the slice-level distribution has a heavy lower tail: tumours that are missed or badly fragmented score close to zero, which Fig. 8 shows clearly. The patient-bootstrap intervals are correspondingly wide, between about 0.12 and 0.20 Dice points (0.12 for the full model with both inference steps; Table 5), and the intervals of all models overlap almost completely.

The Focal-Tversky objective changed the operating point of the network in an unexpected direction. Even though the false-negative weight ($\alpha = 0.7$) exceeds the false-positive weight, the full model had the highest precision (0.811–0.819) and the lowest recall (0.740–0.745) of all twelve evaluated configurations, and its false-positive rate on tumour-free slices was 0.025, against 0.044 for the U-Net, 0.060 for the Attention U-Net, 0.044 for the residual U-Net and 0.057 for the RSAA-UNet trained with BCE and Dice. With 316 tumour-free test slices, a rate of 0.025 corresponds to roughly eight slices with any predicted tumour pixel, against about 14 to 19 for the other networks.

Table 4. Test-set results on tumour-containing slices (16 patients, 182 slices). Best value of each overlap measure in bold. FPR = false-positive rate on 316 tumour-free slices; TTA = test-time augmentation; PP = small-component removal.

Model	Dice (mean ± SD)	IoU	Prec.	Recall	HD95 (px)	Patient Dice	FPR (empty)	Params (M)	Latency (ms)
U-Net	0.7137 ± 0.2881	0.6149	0.8047	0.7611	13.95	0.7193	0.044	7.76	6.8
Attention U-Net	0.6993 ± 0.2798	0.5942	0.7429	0.7903	17.53	0.6921	0.060	7.85	7.6
Residual U-Net	**0.7434 ± 0.2546**	**0.6424**	0.7960	**0.8055**	**13.43**	**0.7422**	0.044	8.11	7.9
RSAA-UNet (BCE+Dice)	0.7206 ± 0.2746	0.6201	0.7902	0.7765	17.57	0.7193	0.057	10.94	11.2
RSAA-UNet (Full)	0.7120 ± 0.2897	0.6143	0.8110	0.7452	16.73	0.7160	0.025	10.94	11.2
Full + TTA	0.7146 ± 0.2860	0.6163	0.8173	0.7409	15.99	0.7201	0.025	10.94	33.7
Full + TTA + PP	0.7151 ± 0.2866	0.6172	0.8194	0.7404	15.93	0.7208	0.025	10.94	33.7

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151090

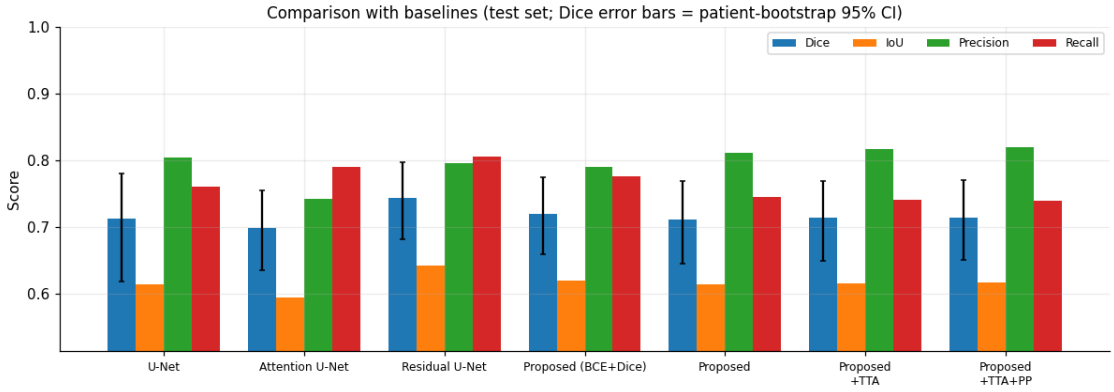


Figure 7. Dice, IoU, precision and recall on tumour slices of the test set. Error bars on Dice are patient-bootstrap 95% confidence intervals.

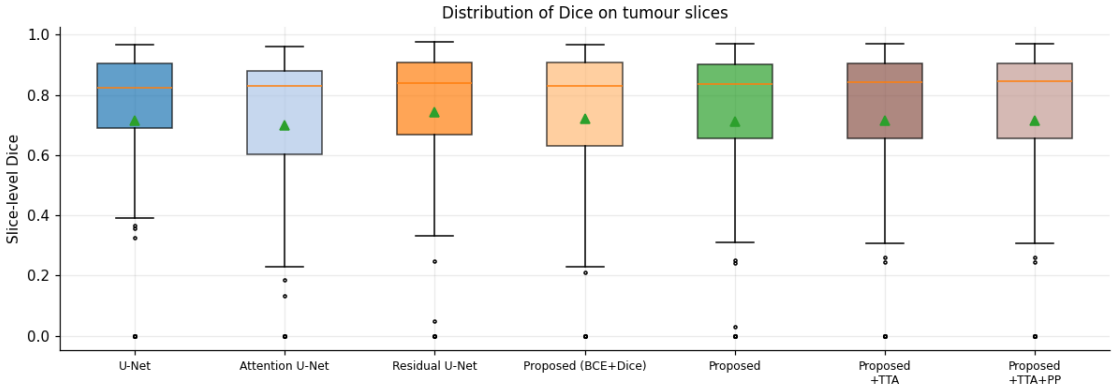


Figure 8. Distribution of slice-level Dice on tumour slices; boxes show quartiles and triangles show means.

4.3 Statistical analysis

Table 5 lists the Dice values with patient-bootstrap intervals and the paired comparisons of the full model with every other configuration. At the patient level, no comparison approached significance: the smallest raw p value was 0.093 (full model against the residual U-Net), and all Holm-adjusted values equal 1.0. At the slice level, three comparisons remained significant after correction: the full model was better than the Attention U-Net (difference +0.0127, $\tilde{p} = 0.006$) and than the variant without SE (+0.0240, $\tilde{p} = 0.022$), and worse than the residual U-Net (-0.0315, $\tilde{p} = 0.040$). Because slices of one patient are not independent, these slice-level results are exploratory; the patient-level results are the ones we rely on. With only 16 test patients, the study has little power to detect differences of one to three Dice points, and the absence of significance should not be read as proof of equivalence.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151090

Table 5. Dice with patient-level bootstrap intervals and paired Wilcoxon tests of the full RSAA-UNet against every other configuration (raw and Holm-adjusted p-values).

Model	Dice	95% CI (patient bootstrap)	Δ Dice (Full − model)	p slice	p patient	Holm p slice	Holm p patient
U-Net	0.7137	.6183–.7804	−0.0017	0.395	0.744	1.000	1.000
Attention U-Net	0.6993	.6358–.7557	+0.0127	0.0006	0.211	0.006	1.000
Residual U-Net	0.7434	.6820–.7978	−0.0315	0.0044	0.093	0.040	1.000
RSAA-UNet (BCE+Dice)	0.7206	.6595–.7748	−0.0086	0.549	0.782	1.000	1.000
RSAA-UNet (Full)	0.7120	.6452–.7690	–	–	–	–	–
w/o SE	0.6880	.5933–.7623	+0.0240	0.0022	0.274	0.022	1.000
w/o ASPP	0.7064	.6370–.7610	+0.0056	0.261	0.464	1.000	1.000
w/o Attention	0.7282	.6586–.7894	−0.0163	0.427	0.495	1.000	1.000
w/o Deep supervision	0.7075	.6361–.7649	+0.0044	0.0087	0.706	0.069	1.000
w/o Residual	0.6812	.5703–.7657	+0.0308	0.385	0.144	1.000	1.000
Full + TTA	0.7146	.6502–.7701	−0.0026	0.711	0.433	1.000	1.000
Full + TTA + PP	0.7151	.6506–.7707	−0.0031	0.480	0.323	1.000	1.000

4.4 Ablation study

Table 6 and Fig. 9 show the effect of removing one component at a time from the full model. Removing residual connections cost the most (3.08 Dice points), followed by SE recalibration (2.40 points), ASPP (0.56 points) and deep supervision (0.44 points). Removing the attention gates went the other way: Dice rose by 1.63 points to 0.7282, with a higher recall (0.8233) and a lower precision (0.7579). None of these changes is significant at the patient level (Table 5). Two details deserve a caution. The variant without deep supervision had by far the highest false-positive rate on tumour-free slices (0.193 against 0.025), but it also stopped at epoch 15 with its best epoch at 5, so the difference mixes the effect of the component with an under-trained network. The variant without residual connections had the lowest HD95 (12.94 px) despite the lowest Dice, which reflects the way HD95 is averaged only over slices where something was predicted. ASPP is the largest single contributor to model size: removing it reduces the parameter count from 10.94 M to 8.31 M, for a Dice change of 0.56 points.

Table 6. Ablation of the full RSAA-UNet (single-pass inference). Δ Dice is the change in Dice points when the component is removed.

Configuration	Dice (mean ± SD)	IoU	Prec.	Recall	HD95 (px)	Patient Dice	FPR (empty)	Params (M)	Δ Dice (points)
RSAA-UNet (Full)	0.7120 ± 0.2897	0.6143	0.8110	0.7452	16.73	0.7160	0.025	10.94	–
w/o SE	0.6880 ± 0.2995	0.5895	0.7417	0.7925	17.77	0.6879	0.041	10.83	−2.40
w/o ASPP	0.7064 ± 0.2935	0.6099	0.7617	0.7985	15.21	0.6996	0.025	8.31	−0.56
w/o Attention	0.7282 ± 0.2640	0.6256	0.7579	0.8233	15.06	0.7298	0.041	10.85	+1.63
w/o Deep supervision	0.7075 ± 0.2582	0.5984	0.7446	0.7921	16.94	0.7028	0.193	10.94	−0.44
w/o Residual	0.6812 ± 0.3313	0.5939	0.7725	0.7676	12.94	0.6777	0.047	10.59	−3.08

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151090

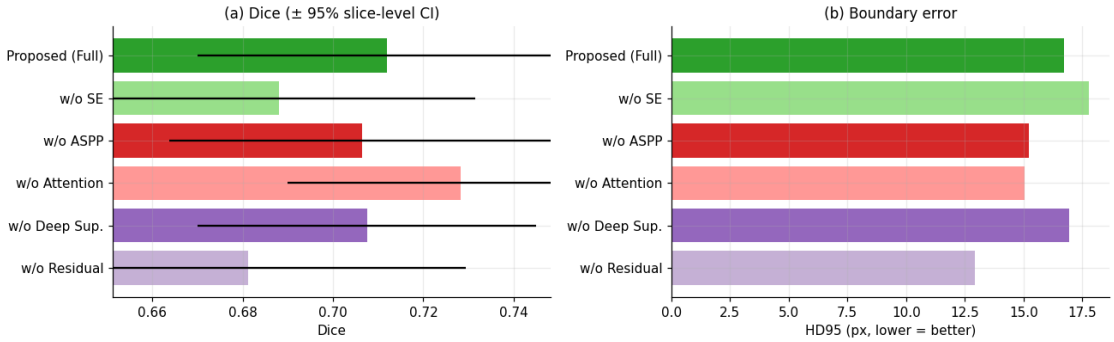


Figure 9. Ablation results: (a) Dice with approximate 95% slice-level intervals and (b) HD95 (lower is better).

4.5 Stratification by tumour size

Tumour area on the test slices was divided into terciles (small below 1.76% of the slice, large above 3.24%). Table 7 and Fig. 10 show that all models are weak on small tumours (Dice 0.44–0.53) and considerably better on medium and large ones (0.79–0.86). The ranking depends on the stratum. On small tumours the residual U-Net (0.5332) and the U-Net (0.5199) came first and the full model with both inference steps reached 0.4996. On medium tumours the residual U-Net led again (0.8436). On large tumours the Attention U-Net (0.8575) and the residual U-Net (0.8552) led, and the plain U-Net was the weakest of all (0.7854), 4.8 points below the full model with test-time augmentation and post-processing (0.8333). The multi-scale context of the proposed network therefore seems to help most with large lesions, and the expected recall advantage of the Tversky weighting for small lesions did not materialise.

Table 7. Mean Dice by tumour-size tercile (share of slice area) on the test set.

Model	Small (< 1.76%)	Medium	Large (> 3.24%)
U-Net	0.5199	0.8378	0.7854
Attention U-Net	0.4388	0.8032	0.8575
Residual U-Net	0.5332	0.8436	0.8552
RSAA-UNet (BCE+Dice)	0.4895	0.8273	0.8467
RSAA-Unet (Full)	0.4848	0.8160	0.8369
Full + TTA	0.4983	0.8138	0.8334
Full + TTA + PP	0.4996	0.8141	0.8333

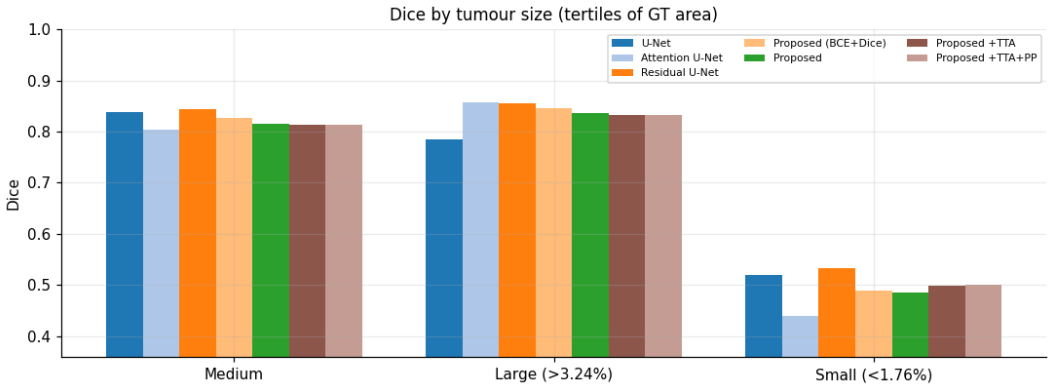


Figure 10. Dice by tumour-size tercile on the test set.

4.6 Pixel-level precision-recall and area agreement

Average precision on the pooled test pixels was 0.854 for the residual U-Net, 0.832 for the Attention U-Net, 0.822 for the U-Net and 0.820 for the full model with test-time augmentation (Fig. 11). The curve of the full model follows the others up to a recall of roughly 0.8–0.85 and then falls away steeply, so that lowering the decision threshold would add many false pixels for little extra recall. The predicted tumour area correlated with the reference area at r = 0.873 for the U-Net and r = 0.879 for the full model with both inference steps (Fig. 12). Both models tend to under-estimate the largest tumours, and both leave a cluster of tumour slices whose predicted area is zero.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151090

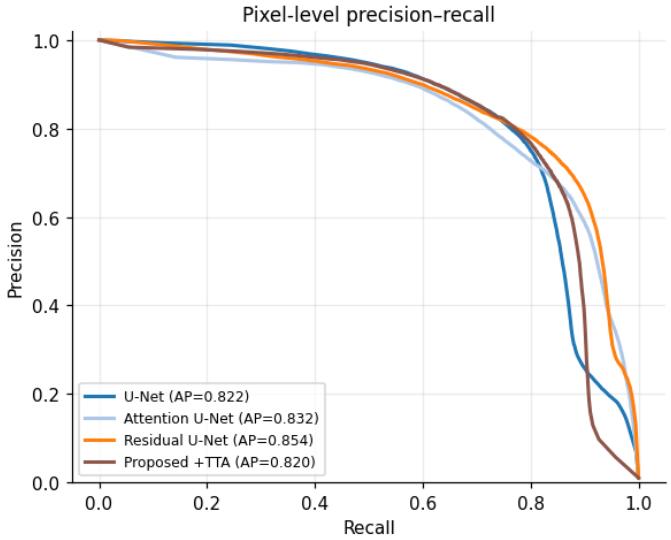


Figure 11. Pixel-level precision-recall curves for the three baselines and the full model with test-time augmentation (random sample of three million test pixels).

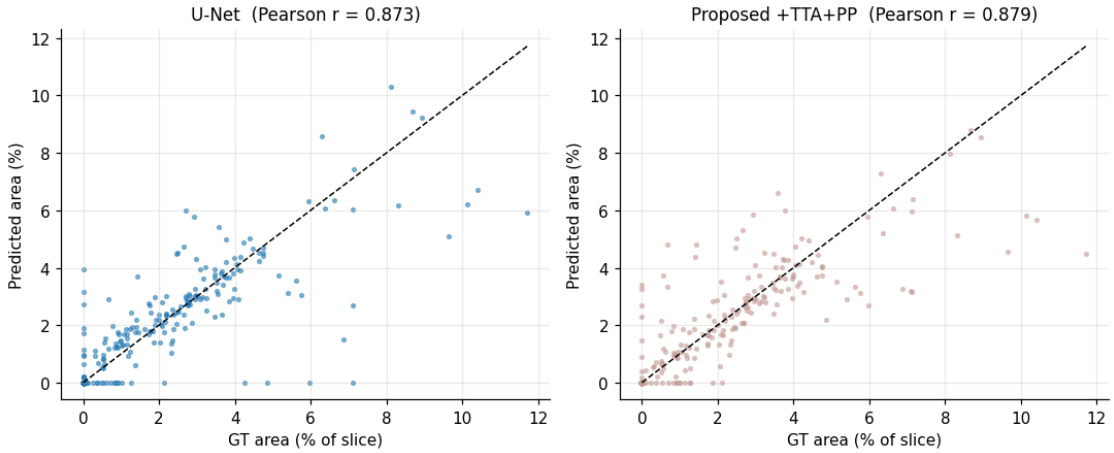


Figure 12. Predicted versus reference tumour area for the U-Net and for the full model with test-time augmentation and post-processing. The dashed line marks perfect agreement.

4.7 Per-patient and qualitative results

Performance varies far more between patients than between models (Fig. 13). Patient CS_6290 reached a mean Dice of roughly 0.9 with several models, whereas CS_6665 and DU_5852 stayed at roughly 0.4–0.6 for most of them, and HT_7608 fell to roughly 0.2–0.3 for the U-Net and for the variants without SE and without residual connections. Differences of one or two Dice points between architectures are small next to these spreads, which explains the wide confidence intervals. The qualitative examples in Fig. 14, arranged from the 3rd to the 97th percentile of the full model's Dice, show the typical pattern. The worst case is a small lesion near the skull base that all four models miss completely (Dice 0.00). For a diffuse lesion in the second row, all models spill into adjacent hyperintense tissue (Dice 0.56–0.70). Compact tumours with clear borders are delineated well (Dice 0.84–0.96 in the lower three rows for the full model).

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151090

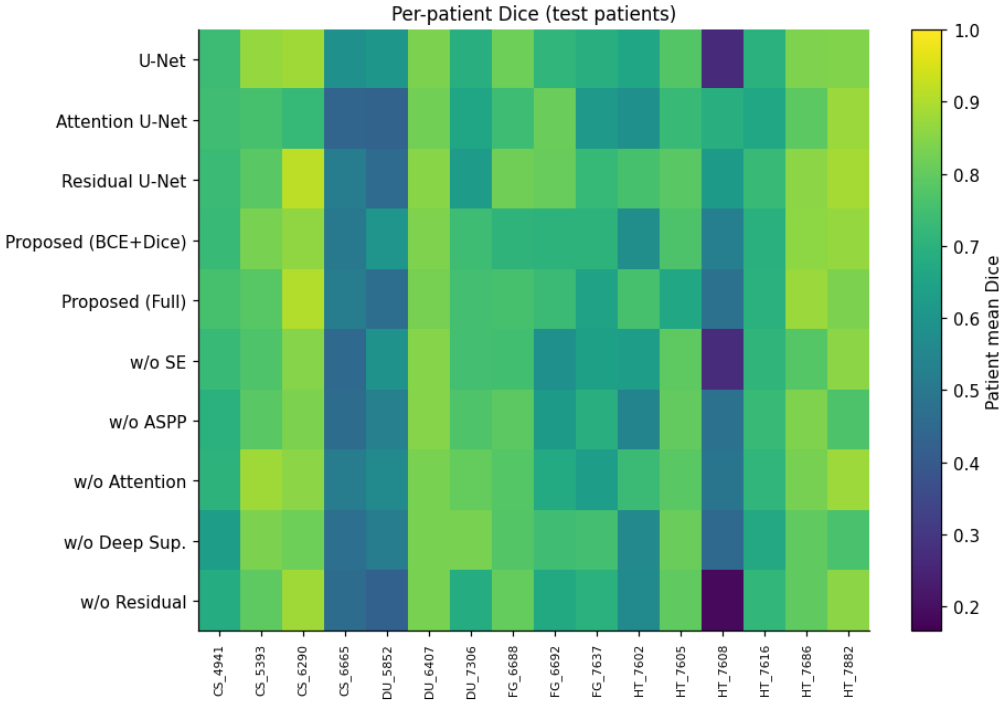


Figure 13. Mean Dice per test patient and model.

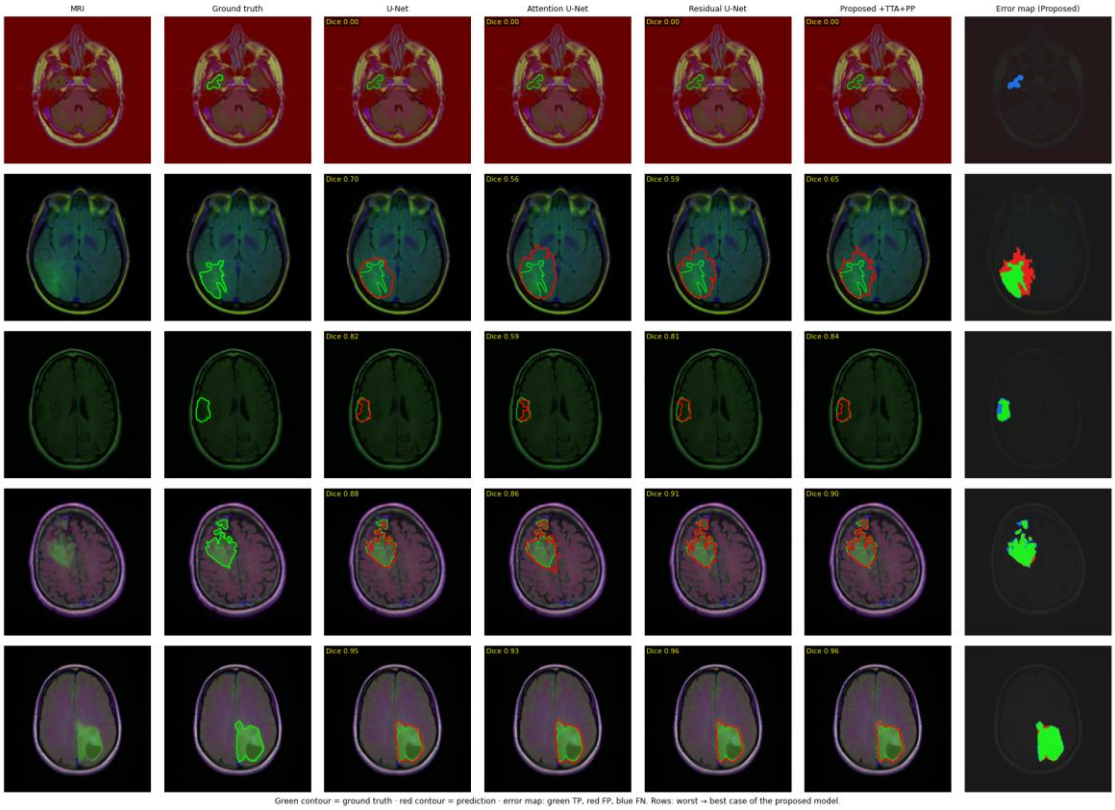


Figure 14. Qualitative comparison on five test slices ranging from the worst to the best case of the full model. Green contours show the reference, red contours show predictions, and the last column is an error map (green true positive, red false positive, blue false negative).

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151090

4.8 Model size and speed

The full network has 10.94 M parameters, 41% more than the U-Net (7.76 M), and a single-slice latency of 11.2 ms against 6.8 ms (+65%). The residual U-Net has 8.11 M parameters and needs 7.9 ms. The three-view test-time augmentation triples the cost to about 34 ms per slice. For a segmentation of one slice, all of these times are small in absolute terms, so the practical cost of the proposed design is determined by training effort and not by inference.

5. DISCUSSION

Principal findings. The experiment gives a clear but modest answer. Under a patient-wise protocol, the five components assembled in RSAA-UNet do not add up to a better segmenter than a residual U-Net with 8.1 M parameters. The ranking by Dice (residual U-Net, RSAA-UNet with BCE and Dice, U-Net, full RSAA-UNet, Attention U-Net) is within the noise of a 16-patient test set. What the proposed objective did change is the balance between error types: the full model made fewer false alarms on tumour-free slices and had higher precision, at the price of recall. Residual connections were the single most useful ingredient in both the baseline and the ablation experiments, and SE recalibration came second.

Why did the larger network not win? Several explanations are compatible with the data, and we cannot separate them without further runs. First, the training set contains only 78 patients, and a 10.9 M-parameter model with an ASPP block of 2.6 M parameters at a 16 × 16 resolution, where little spatial context remains, may simply have more capacity than the data can constrain. Second, the full model was trained for only 20 useful epochs, with the learning rate still high, and its best checkpoint was chosen on a validation criterion that gives full credit for empty predictions on empty slices. That criterion rewards false-positive suppression, which the Focal-Tversky models do well, and this may be why the RSAA-UNet variants lead on validation Dice (0.883–0.891) but not on test Dice computed over tumour slices only. Third, the attention gates take a same-resolution gating signal and add two batch-normalisation layers, which differs from the coarse-to-fine design of the original; in a task where the FLAIR abnormality is bright and diffuse and the encoder already localises it, they may have little left to filter. This remains a conjecture, since we did not inspect attention maps.

The unexpected behaviour of the Focal-Tversky term. With $\alpha > \beta$ one would expect higher recall, and we observed the opposite. One contributing factor may be that a per-image Tversky term on tumour-free slices reduces to a penalty on any predicted foreground, and that raising it to a power below one amplifies small residual errors; this would push the network towards empty predictions on the many slices without tumour. A second, separate concern is numerical. The derivative of $(1 - \mathrm{TI})^{\gamma}$ with $\gamma < 1$ grows without bound as $\mathrm{TI} \to 1$, so slices that are predicted almost perfectly can produce very large gradients. Together with a high learning rate and mixed-precision arithmetic, this is one plausible route to the divergence seen after epoch 20. We stress that this is a hypothesis: the experiment did not include a run that isolates it. Reasonable remedies are to clamp the base of the power, to use a larger ϵ, to lower the peak learning rate, or to apply the focal term only to slices that contain tumour.

Validation versus test. The ordering of models on the validation set (including empty slices) and on the test set (tumour slices only) differs. This is not a contradiction; the two quantities measure different things. It does show that checkpoint selection on all-slice Dice will favour conservative models, and that reporting only one of the two views would give a misleading picture.

Practical reading. If a single 2-D network is needed for FLAIR LGG segmentation on this dataset, the residual U-Net is the better choice on current evidence: it is smaller, faster, and at least as accurate. If false alarms on tumour-free slices matter more than missed small lesions, for example when the output feeds a volume estimate that should not include spurious islands, the Focal-Tversky model with test-time augmentation and post-processing is the more conservative option. Test-time augmentation and component removal helped consistently but only slightly (+0.31 Dice points, +0.84 precision points, −0.8 px HD95), at three times the inference cost.

Limitations. Each configuration was trained once, on one split, so run-to-run variance is not measured and is probably of the same order as some of the differences reported. The test set has 16 patients, 182 of whose slices contain tumour. The networks are two-dimensional and see no through-plane context, and the images come from a single public collection with single-reader outlines, so generalisation to other scanners or to BraTS-style data is untested. HD95 is expressed in pixels and excludes slices without a prediction. The full model did not train to completion. The comparison with published studies is indicative only. Finally, the validation set influenced checkpoint choice and early stopping for every model, but not the test set.



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026
DOI: 10.17148/IJARCCE.2026.151090

Future work. The most informative next steps are patient-wise five-fold cross-validation with several seeds, a stabilised version of the Focal-Tversky term with a full-length, annealed training run, a lower-capacity ASPP (or none), inspection of attention maps, the 2.5-D input of adjacent slices, three-dimensional post-processing as in Buda et al. (2019), and external evaluation on the lower-grade cases of BraTS (Bakas et al., 2017; Menze et al., 2015).

6. CONCLUSION

We trained and compared ten 2-D segmentation networks for lower-grade glioma on FLAIR MRI under one protocol and a patient-wise 78/16/16 split. The proposed RSAA-UNet, which combines residual SE blocks, attention gates, ASPP, deep supervision and a Focal-Tversky plus cross-entropy loss, reached a Dice of 0.715 on tumour slices of 16 unseen patients with test-time augmentation and small-component removal. This is level with a plain U-Net (0.714) and below a residual U-Net (0.743), and none of the differences was significant at the patient level. The proposed design lowered false positives on tumour-free slices (2.5% against 4.4–6.0%) and raised precision, but it did not improve recall for small tumours, and its training diverged after epoch 20. Residual connections and SE recalibration were the components that clearly mattered; attention gates and ASPP were not justified by their cost in this setting. We hope the protocol, the ablation and the account of failure are useful to those who use this dataset, and we regard multi-seed cross-validation and a stabilised loss as the necessary next experiments.

Declarations

Data availability: The imaging data and masks are publicly available as the LGG MRI segmentation collection (Kaggle: mateuszbuda/lgg-mri-segmentation) and through The Cancer Imaging Archive (Pedano et al., 2016). The analysis notebook that produced all tables and figures is available from the corresponding author on reasonable request. The patient split used here is reproducible from the seed (42) stated in Section 3.2.

Ethics statement: Only previously published, de-identified data were used. No new data were collected from human participants.

Funding: No specific funding was received for this work.

Competing interests: The authors declare no competing interests.

Author contributions: Both authors contributed equally to this work.

REFERENCES

[1]. Abraham, N., & Khan, N. M. (2019). A novel focal Tversky loss function with improved attention U-Net for lesion segmentation. In *Proceedings of the 16th IEEE International Symposium on Biomedical Imaging (ISBI 2019)* (pp. 683–687). IEEE. https://doi.org/10.1109/ISBI.2019.8759329

[2]. Akkus, Z., Galimzianova, A., Hoogi, A., Rubin, D. L., & Erickson, B. J. (2017). Deep learning for brain MRI segmentation: State of the art and future directions. *Journal of Digital Imaging, 30*(4), 449–459. https://doi.org/10.1007/s10278-017-9983-4

[3]. Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(12), 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615

[4]. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J. S., Freymann, J. B., Farahani, K., & Davatzikos, C. (2017). Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data, 4*, Article 170117. https://doi.org/10.1038/sdata.2017.117

[5]. Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks, 106*, 249–259. https://doi.org/10.1016/j.neunet.2018.07.011

[6]. Buda, M., Saha, A., & Mazurowski, M. A. (2019). Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm. *Computers in Biology and Medicine, 109*, 218–225. https://doi.org/10.1016/j.compbiomed.2019.05.002

[7]. Cancer Genome Atlas Research Network. (2015). Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *New England Journal of Medicine, 372*(26), 2481–2498. https://doi.org/10.1056/NEJMoa1402121

[8]. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., & Wang, M. (2023). Swin-Unet: Unet-like pure transformer for medical image segmentation. In *Computer Vision – ECCV 2022 Workshops* (Lecture Notes in Computer Science, Vol. 13803, pp. 205–218). Springer. https://doi.org/10.1007/978-3-031-25066-8_9

[9]. Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., Luo, X., Xie, Y., Adeli, E., Wang, Y., Lungren, M. P., Zhang, S., Xing, L., Lu, L., Yuille, A. L., & Zhou, Y. (2024). TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers. *Medical Image Analysis, 97*, Article 103280. https://doi.org/10.1016/j.media.2024.103280

DOI: 10.17148/IJARCCE.2026.151090

[10]. Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 40*(4), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184

[11]. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Computer Vision – ECCV 2018* (Lecture Notes in Computer Science, Vol. 11211, pp. 833–851). Springer. https://doi.org/10.1007/978-3-030-01234-2_49

[12]. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., & Prior, F. (2013). The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *Journal of Digital Imaging, 26*(6), 1045–1057. https://doi.org/10.1007/s10278-013-9622-7

[13]. Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology, 26*(3), 297–302. https://doi.org/10.2307/1932409

[14]. Dong, H., Yang, G., Liu, F., Mo, Y., & Guo, Y. (2017). Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks. In *Medical Image Understanding and Analysis (MIUA 2017)* (Communications in Computer and Information Science, Vol. 723, pp. 506–517). Springer. https://doi.org/10.1007/978-3-319-60964-5_44

[15]. Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S., & Pal, C. (2016). The importance of skip connections in biomedical image segmentation. In *Deep Learning and Data Labeling for Medical Applications* (Lecture Notes in Computer Science, Vol. 10008, pp. 179–187). Springer. https://doi.org/10.1007/978-3-319-46976-8_19

[16]. Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics, 7*(1), 1–26. https://doi.org/10.1214/aos/1176344552

[17]. Ghaffari, M., Sowmya, A., & Oliver, R. (2020). Automated brain tumor segmentation using multimodal brain scans: A survey based on models submitted to the BraTS 2012–2018 challenges. *IEEE Reviews in Biomedical Engineering, 13*, 156–168. https://doi.org/10.1109/RBME.2019.2946868

[18]. Güzel, Y., & Aydın, Z. (2025). A comparative study of UNet variants for low-grade glioma segmentation in magnetic resonance imaging. *İnönü Üniversitesi Sağlık Hizmetleri Meslek Yüksek Okulu Dergisi, 13*(2), 344–355. https://doi.org/10.33715/inonusaglik.1677185

[19]. 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. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries (BrainLes 2021)* (Lecture Notes in Computer Science, Vol. 12962, pp. 272–284). Springer. https://doi.org/10.1007/978-3-031-08999-2_22

[20]. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H. R., & Xu, D. (2022). UNETR: Transformers for 3D medical image segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (pp. 574–584). IEEE. https://doi.org/10.1109/WACV51458.2022.00181

[21]. Singh, D. P., Joseph, S. G., Selvi, V. T., Karunakaran, S., Appasami, G., & Jegajothi, B. (2022). Quasi-oppositional Satin Bowerbird with deep learning based content-based image retrieval. In 2022 6th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 998–1005). IEEE. https://doi.org/10.1109/ICCMC53470.2022.9754135

[22]. Appasami, G., Savarimuthu, N., & Kannan, B. (2026). Identification of pneumonia in chest X-ray images using generative and explainable AI methods. In Applied AI and machine learning for real-world systems (pp. 79–108). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-2600-2701-1.ch003

[23]. Parimala, S., Poornima, S. N., Lade, S. G., Venkatesh, N., & Appasami, G. (2025). Deep learning-based prediction models for colorectal disease: Enhancing diagnostic accuracy. International Research Journal of Multidisciplinary Scope, 6(3), 1743–1756. https://doi.org/10.47857/irjms.2025.v06i03.04664

[24]. Appasami, G., & Savarimuthu, N. (2024). Comparative analysis of COVID-19, pneumonia and other lung diseases classifications based on deep learning architectures with product metric score. African Journal of Biomedical Research, 27(4S), 11863–11891. https://doi.org/10.53555/AJBR.v27i4S.5746

[25]. Appasami, G., Kavitha Lakshmi, R., & Savarimuthu, N. (2024). ECOVID-Net: An ensemble deep transfer learning-based neural network for COVID-19 diagnosis using chest X-ray images. African Journal of Biomedical Research, 27(4S), 11831–11844. https://doi.org/10.53555/ajbr.v27i4S.5747

[26]. Vedanarayanan, V., Cholli, N. G., Meleet, M., Maurya, B., Appasami, G., & Khurana, M. (2023). Implementation of the deep learning-based website for pneumonia detection and classification. In AI and IoT-based intelligent health care & sanitation (pp. 129–143). Bentham Science Publishers. https://doi.org/10.2174/9789815136531123010011

[27]. Jagadiswary, D., Appasami, G., & Rajesh, S. (2011). Eye features normalization and face emotion detection for human face recognition. In 2011 International Conference on Electronics, Communication and Computing Technologies (ICECCT). https://doi.org/10.1109/ICECCT.2011.6077071

DOI: 10.17148/IJARCCE.2026.151090

[28]. Appasami, G., & Joseph, K. S. (2011). Optimization of operating systems towards green computing. International Journal of Combinatorial Optimization Problems and Informatics, 2(3), 39–51. https://www.redalyc.org/pdf/2652/265219635005.pdf

[29]. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M., & Larochelle, H. (2017). Brain tumor segmentation with deep neural networks. *Medical Image Analysis, 35*, 18–31. https://doi.org/10.1016/j.media.2016.05.004

[30]. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). IEEE. https://doi.org/10.1109/CVPR.2016.90

[31]. Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70. https://www.jstor.org/stable/4615733

[32]. Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2020). Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 42*(8), 2011–2023. https://doi.org/10.1109/TPAMI.2019.2913372

[33]. Huttenlocher, D. P., Klanderman, G. A., & Rucklidge, W. J. (1993). Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 15*(9), 850–863. https://doi.org/10.1109/34.232073

[34]. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016* (Lecture Notes in Computer Science, Vol. 9901, pp. 424–432). Springer. https://doi.org/10.1007/978-3-319-46723-8_49

[35]. Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv*. https://doi.org/10.48550/arXiv.1502.03167

[36]. Isensee, F., Jaeger, P. F., Kohl, S. A. 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. https://doi.org/10.1038/s41592-020-01008-z

[37]. Jadon, S. (2020). A survey of loss functions for semantic segmentation. In *Proceedings of the 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)* (pp. 1–7). IEEE. https://doi.org/10.1109/CIBCB48159.2020.9277638

[38]. Jlassi, A., ElBedoui, K., & Barhoumi, W. (2023). Brain tumor segmentation of lower-grade glioma across MRI images using hybrid convolutional neural networks. In *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023)*. SCITEPRESS. https://doi.org/10.5220/0011895900003393

[39]. Kamnitsas, K., Ledig, C., Newcombe, V. F. J., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D., & Glocker, B. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis, 36*, 61–78. https://doi.org/10.1016/j.media.2016.10.004

[40]. Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2020). Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 42*(2), 318–327. https://doi.org/10.1109/TPAMI.2018.2858826

[41]. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42*, 60–88. https://doi.org/10.1016/j.media.2017.07.005

[42]. Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3431–3440). IEEE. https://doi.org/10.1109/CVPR.2015.7298965

[43]. Loshchilov, I., & Hutter, F. (2019). Decoupled weight decay regularization. In *International Conference on Learning Representations (ICLR 2019)*. https://doi.org/10.48550/arXiv.1711.05101

[44]. Louis, D. N., Perry, A., Wesseling, P., Brat, D. J., Cree, I. A., Figarella-Branger, D., Hawkins, C., Ng, H. K., Pfister, S. M., Reifenberger, G., Soffietti, R., von Deimling, A., & Ellison, D. W. (2021). The 2021 WHO classification of tumors of the central nervous system: A summary. *Neuro-Oncology, 23*(8), 1231–1251. https://doi.org/10.1093/neuonc/noab106

[45]. Maier-Hein, L., Reinke, A., Godau, P., Tizabi, M. D., Buettner, F., Christodoulou, E., Glocker, B., Isensee, F., Kleesiek, J., Kozubek, M., Reyes, M., Riegler, M. A., Wiesenfarth, M., Kavur, A. E., Sudre, C. H., Baumgartner, M., Eisenmann, M., Heckmann-Nötzel, D., Rädsch, T., … Jäger, P. F. (2024). Metrics reloaded: Recommendations for image analysis validation. *Nature Methods, 21*(2), 195–212. https://doi.org/10.1038/s41592-023-02151-z

[46]. Ma, J., Chen, J., Ng, M., Huang, R., Li, Y., Li, C., Yang, X., & Martel, A. L. (2021). Loss odyssey in medical image segmentation. *Medical Image Analysis, 71*, Article 102035. https://doi.org/10.1016/j.media.2021.102035

DOI: 10.17148/IJARCCE.2026.151090

[47]. Mazurowski, M. A., Buda, M., Saha, A., & Bashir, M. R. (2019). Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI. *Journal of Magnetic Resonance Imaging, 49*(4), 939–954. https://doi.org/10.1002/jmri.26534

[48]. Mazurowski, M. A., Clark, K., Czarnek, N. M., Shamsesfandabadi, P., Peters, K. B., & Saha, A. (2017). Radiogenomics of lower-grade glioma: Algorithmically-assessed tumor shape is associated with tumor genomic subtypes and patient outcomes in a multi-institutional study with The Cancer Genome Atlas data. *Journal of Neuro-Oncology, 133*(1), 27–35. https://doi.org/10.1007/s11060-017-2420-1

[49]. Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.-A., Arbel, T., Avants, B. B., Ayache, N., Buendia, P., Collins, D. L., Cordier, N., … Van Leemput, K. (2015). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging, 34*(10), 1993–2024. https://doi.org/10.1109/TMI.2014.2377694

[50]. Micikevicius, P., Narang, S., Alben, J., Diamos, G., Elsen, E., Garcia, D., Ginsburg, B., Houston, M., Kuchaiev, O., Venkatesh, G., & Wu, H. (2018). Mixed precision training. In *International Conference on Learning Representations (ICLR 2018)*. https://doi.org/10.48550/arXiv.1710.03740

[51]. Milletari, F., Navab, N., & Ahmadi, S.-A. (2016). V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In *Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV)* (pp. 565–571). IEEE. https://doi.org/10.1109/3DV.2016.79

[52]. Myronenko, A. (2019). 3D MRI brain tumor segmentation using autoencoder regularization. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries (BrainLes 2018)* (Lecture Notes in Computer Science, Vol. 11384, pp. 311–320). Springer. https://doi.org/10.1007/978-3-030-11726-9_28

[53]. Naser, M. A., & Deen, M. J. (2020). Brain tumor segmentation and grading of lower-grade glioma using deep learning in MRI images. *Computers in Biology and Medicine, 121*, Article 103758. https://doi.org/10.1016/j.compbiomed.2020.103758

[54]. Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., & Rueckert, D. (2018). Attention U-Net: Learning where to look for the pancreas. *arXiv*. https://doi.org/10.48550/arXiv.1804.03999

[55]. Pedano, N., Flanders, A. E., Scarpace, L., Mikkelsen, T., Eschbacher, J. M., Hermes, B., Sisneros, V., Barnholtz-Sloan, J., & Ostrom, Q. (2016). *The Cancer Genome Atlas Low Grade Glioma collection (TCGA-LGG)* [Data set]. The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2016.L4LTD3TK

[56]. Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Transactions on Medical Imaging, 35*(5), 1240–1251. https://doi.org/10.1109/TMI.2016.2538465

[57]. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (Lecture Notes in Computer Science, Vol. 9351, pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28

[58]. Roy, A. G., Navab, N., & Wachinger, C. (2018). Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2018* (Lecture Notes in Computer Science, Vol. 11070, pp. 421–429). Springer. https://doi.org/10.1007/978-3-030-00928-1_48

[59]. Appasami, G., & Savarimuthu, N. (2025). Cross-dataset performance evaluation and secure federated learning with weighted model aggregation for MRI brain tumor image classification. Journal of Supercomputing, 81, Article 1555. https://doi.org/10.1007/s11227-025-08051-7

[60]. Appasami, G., & Savarimuthu, N. (2025). Federated learning for secure medical MRI brain tumor image classification. European Physical Journal Special Topics, 234, 4813–4827. https://doi.org/10.1140/epjs/s11734-025-01516-z

[61]. Appasami, G., & Savarimuthu, N. (2025). A novel lightweight CNN design for MRI brain tumor image classification with performance-driven optimization. Discover Computing, 28, Article 206. https://doi.org/10.1007/s10791-025-09603-4

[62]. Salehi, S. S. M., Erdogmus, D., & Gholipour, A. (2017). Tversky loss function for image segmentation using 3D fully convolutional deep networks. In *Machine Learning in Medical Imaging (MLMI 2017)* (Lecture Notes in Computer Science, Vol. 10541, pp. 379–387). Springer. https://doi.org/10.1007/978-3-319-67389-9_44

[63]. Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., & Rueckert, D. (2019). Attention gated networks: Learning to leverage salient regions in medical images. *Medical Image Analysis, 53*, 197–207. https://doi.org/10.1016/j.media.2019.01.012

[64]. Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data, 6*, Article 60. https://doi.org/10.1186/s40537-019-0197-0



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151090

[65]. Siddique, N., Paheding, S., Elkin, C. P., & Devabhaktuni, V. (2021). U-Net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access, 9*, 82031–82057. https://doi.org/10.1109/ACCESS.2021.3086020

[66]. Smith, L. N., & Topin, N. (2019). Super-convergence: Very fast training of neural networks using large learning rates. In *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications* (Proceedings of SPIE, Vol. 11006, Article 1100612). SPIE. https://doi.org/10.1117/12.2520589

[67]. Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., & Cardoso, M. J. (2017). Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (Lecture Notes in Computer Science, Vol. 10553, pp. 240–248). Springer. https://doi.org/10.1007/978-3-319-67558-9_28

[68]. Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Medical Imaging, 15*, Article 29. https://doi.org/10.1186/s12880-015-0068-x

[69]. Varoquaux, G., & Cheplygina, V. (2022). Machine learning for medical imaging: Methodological failures and recommendations for the future.*npj Digital Medicine, 5*, Article 48. https://doi.org/10.1038/s41746-022-00592-y

[70]. Wang, G., Li, W., Aertsen, M., Deprest, J., Ourselin, S., & Vercauteren, T. (2019). Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. *Neurocomputing, 338*, 34–45. https://doi.org/10.1016/j.neucom.2019.01.103

[71]. Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin, 1*(6), 80–83. https://doi.org/10.2307/3001968

[72]. Yadav, A. C., Kolekar, M. H., Sonawane, Y., Kadam, G., Tiwarekar, S., & Kalbande, D. R. (2024). EffUNet++: A novel architecture for brain tumor segmentation using FLAIR MRI images. *IEEE Access, 12*, 152430–152443. https://doi.org/10.1109/ACCESS.2024.3480271

[73]. Zhang, Z., Liu, Q., & Wang, Y. (2018). Road extraction by deep residual U-Net. *IEEE Geoscience and Remote Sensing Letters, 15*(5), 749–753. https://doi.org/10.1109/LGRS.2018.2802944

[74]. Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2020). UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging, 39*(6), 1856–1867. https://doi.org/10.1109/TMI.2019.2959609

 This work is licensed under a Creative Commons Attribution 4.0 International License