

Sequential LSTM-Based MRI Preprocessing and Attention-Enhanced EfficientNet U-Net for Brain Tumor Segmentation

K. Santhi¹, Dr. S. Radha Krishna²

¹Research Scholar, Department of CSE, JNTUA, ATP, AP, India, Assistant Professor (On Contract), YSRAFU Kadapa, Andhra Pradesh, India

²Associate Professor, Department of CSE, JNTU College of Engineering, Pulivendula, Andhra Pradesh, India

Emails: kethasanthi79@gmail.com¹

Abstract

Manual delineation of brain tumors in magnetic resonance imaging (MRI) is time-consuming and subject to inter-reader variability. Automated segmentation can support reproducible tumor delineation, but slice-wise processing does not explicitly represent contextual relationships among adjacent MRI slices. This study investigates a sequential framework that first processes ordered MRI slices with a long short-term memory (LSTM) stage and then performs binary tumor segmentation using an attention-enhanced U-Net with an EfficientNet encoder. Experiments use the public Figshare brain-tumor dataset containing 3,064 T1-weighted contrast-enhanced MRI images from 233 patients with glioma, meningioma, or pituitary tumor. The documented configuration uses RMSprop for the recurrent stage and AdamW for the segmentation network, with batch size 8, a maximum of 500 epochs, dropout 0.5, data augmentation during training, and a Dice-based loss. Performance is assessed primarily with the Dice similarity coefficient (DSC) and Jaccard index (JI). The framework reports a DSC of 92.25% and a JI of 91.41%. The JI value requires verification against the underlying predictions because DSC and JI are mathematically linked for the same binary masks. Literature-based U-Net variants are used only as descriptive reference points because they were not independently retrained under the present protocol. The study therefore provides dataset-specific evidence for the proposed architectural integration while retaining explicit limitations concerning sequence construction, ablation, baseline reproduction, and external validation.

Keywords: Brain tumor segmentation; magnetic resonance imaging; sequential LSTM; ordered inter-slice context; EfficientNet; U-Net; attention mechanism; Figshare dataset.

1. Introduction

Brain tumors vary in size, morphology, location, intensity and boundary characteristics, making automated delineation a challenging medical-image segmentation problem. MRI provides high soft-tissue contrast and is widely used for brain-tumor characterization and quantitative analysis. Deep convolutional networks have become a principal approach for automated tumor segmentation because they learn hierarchical representations directly from annotated images [3,4]. Subsequent work has explored cascaded convolutional models, incremental deep networks, and volumetric architectures to increase spatial context [5,6,8]. A central challenge is the representation of contextual information. A conventional two-dimensional segmentation network processes an input slice

independently, whereas a volumetric model can use neighboring spatial information but requires appropriately organized 3D data and generally greater computational resources. The distinction is important for the Figshare brain-tumor dataset, which contains individual MATLAB image records rather than the multimodal volumetric structure used by benchmarks such as BraTS [1,2]. Dataset provenance, preprocessing, partitioning, and evaluation protocol therefore influence how segmentation results should be interpreted [7]. U-Net remains a foundational encoder-decoder architecture for biomedical segmentation because skip connections combine high-level semantic features with spatial detail [10]. Attention gates can selectively regulate information transferred through

these skip pathways [11], while EfficientNet provides a compound-scaled convolutional encoder [15]. Both components have already been investigated for brain-tumor segmentation, including on the Figshare dataset [16,17]. Recurrent networks provide another route to contextual representation. LSTM models maintain information across ordered inputs [18], and ConvLSTM extends recurrent processing to spatially structured sequences [19]. Brain-tumor segmentation studies have incorporated recurrent modules within U-Net-type networks [20]. The research gap addressed here is narrower than the general development of U-Net, attention, EfficientNet, or LSTM. The study investigates the placement of a sequential LSTM stage before an attention-enhanced EfficientNet U-Net, with the recurrent index representing ordered inter-slice position rather than biological time. This arrangement is intended to produce an enhanced MRI representation before hierarchical segmentation features are extracted. The objective is to examine the feasibility of this architectural integration on the Figshare dataset and to report its segmentation performance while clearly distinguishing experimentally obtained values from literature-derived comparator values. The main contributions are: (i) investigation of sequential LSTM-based preprocessing for ordered inter-slice representation; (ii) integration of that representation with an EfficientNet-based U-Net using additive attention gates; (iii) evaluation with overlap-based segmentation measures; and (iv) a cautious interpretation that separates the present result from uncontrolled literature comparisons and identifies the need for component-wise ablation and external validation.

2. Related Work

2.1.CNN/U-Net and Attention-Based Segmentation

CNN-based segmentation established learned feature extraction as an effective alternative to conventional image processing for brain tumors [3–6]. U-Net subsequently became a standard biomedical segmentation framework by combining an encoder, decoder, and skip connections [10]. Three-dimensional U-Net ensembles can exploit volumetric context, but their data and computational requirements differ from the slice-based Figshare

setting [8]. Attention U-Net introduced additive attention gates that suppress less relevant encoder responses before skip-feature fusion [11]. CBAM provides a different channel-spatial attention formulation [12]; it is therefore not equivalent to the attention-gate mechanism used here.

2.2.Recurrent and Context-Aware Segmentation

LSTM was designed to retain useful information across sequential inputs through gated memory operations [18], while ConvLSTM incorporates convolutional transformations to preserve spatial structure [19]. Recurrent processing has been used in brain-tumor segmentation, including ConvLSTM-based U-Net architectures [20]. The present design differs in architectural placement: the LSTM stage precedes the segmentation network rather than being embedded in the encoder, decoder, or skip pathways. The sequence index represents ordered slice position. Consequently, the method should be interpreted as inter-slice contextual representation rather than temporal modeling of dynamic MRI.

2.3.Research Gap

EfficientNet-based attention segmentation has already been reported on the Figshare dataset [16,17], and recurrent segmentation has also been investigated [20]. The unresolved issue considered here is whether a preceding sequential LSTM representation can be integrated with an attention-enhanced EfficientNet U-Net for the same binary segmentation task. Because prior studies differ in preprocessing, data partitions, image dimensions, optimization, and evaluation protocols, published metrics are treated as contextual evidence rather than controlled baselines. This distinction is especially important because the closest same-dataset EfficientNet-attention study reported a Dice value of 0.9339 [16], which is not directly comparable without reproducing its protocol.

3. Materials and Methods

3.1.Dataset

The experimental dataset is the public Figshare brain-tumor dataset introduced by Cheng et al. [1]. It contains 3,064 T1-weighted contrast-enhanced MRI images from 233 patients: 1,426 glioma images from 89 patients, 930 pituitary-tumor images from 62 patients, and 708 meningioma images from 82

patients. Images are represented in axial, coronal, and sagittal planes and are stored as MATLAB records. Each image is 512×512 pixels and is associated with a manually delineated binary tumor mask. BraTS was not used for training or evaluation; it is discussed only to distinguish the dataset organization and segmentation task [2].

Table 1 Summary of the Figshare Brain-Tumor Dataset and Experimental Partition

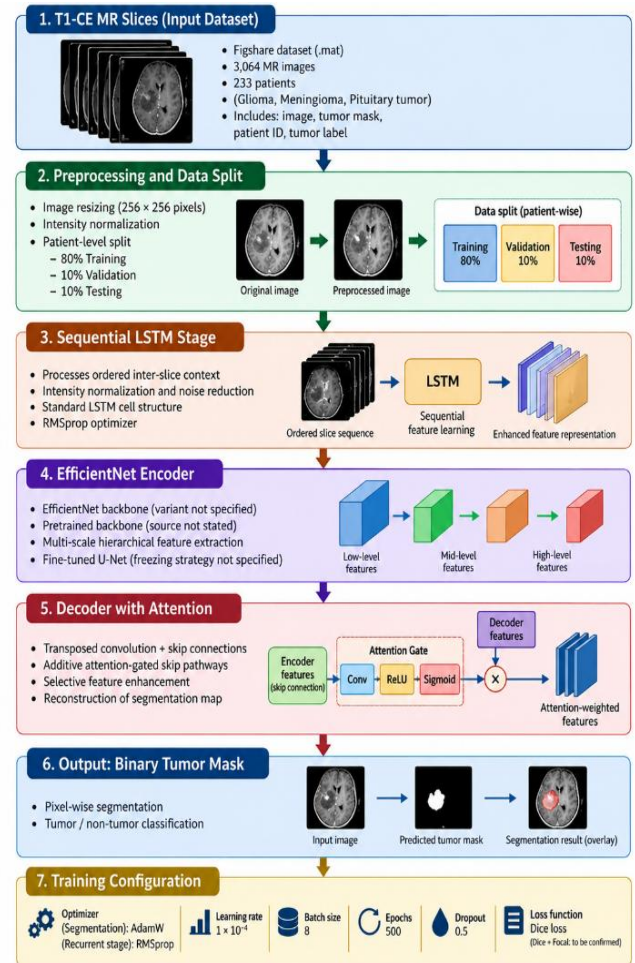
Characteristic	Value
Source	Cheng et al.; Figshare [1]
Patients / images	233 / 3,064
Modality	T1-weighted contrast-enhanced MRI
Tumor categories	Glioma, meningioma, pituitary
Images per category	1,426 / 708 / 930
Original image size	512×512 pixels
Input size	256×256 pixels
Planes	Axial, coronal, sagittal
Ground truth	Manual binary tumor mask
Reported partition	Patient-level 80/10/10

3.2. Overall Framework and Data Preparation

The documented pipeline comprises MRI preparation, sequential LSTM-based contextual processing, and attention-enhanced EfficientNet U-Net segmentation. The intended information flow is: T1-CE MRI → preprocessing → ordered inter-slice

LSTM representation → EfficientNet encoder → attention-gated skip pathways → U-Net decoder → binary tumor mask. The recurrent stage and segmentation stage are described as sequential components because the source specifies RMSprop for the recurrent stage and AdamW for the segmentation network but does not establish whether recurrent parameters were subsequently frozen or jointly optimized. As shown in Fig. 1. Overall framework of the study, combining MRI preparation, sequential LSTM representation, EfficientNet encoding, attention-enhanced decoding and binary tumor-mask generation.

Figure 1 Overall framework of the study, combining MRI preparation, sequential LSTM representation, EfficientNet encoding, attention-enhanced decoding and binary tumor-mask generation



The source manuscript reports conversion of MATLAB records to image and mask arrays, resizing from 512×512 to 256×256 pixels, intensity normalization, and noise reduction. Binary masks require label-preserving interpolation during resizing. The available record does not sufficiently document channel conversion, sequence length, padding, recurrent hidden-state size, recurrent directionality, or the exact procedure used to construct ordered sequences. Therefore, the method is described only to the level supported by the source. Bias-field correction, skull stripping, anatomical cropping, and other unreported operations are not added. As shown in Figure 2 Representative T1-weighted contrast-enhanced MRI images and corresponding tumor masks for the three tumor categories and three imaging planes in the Figshare dataset.

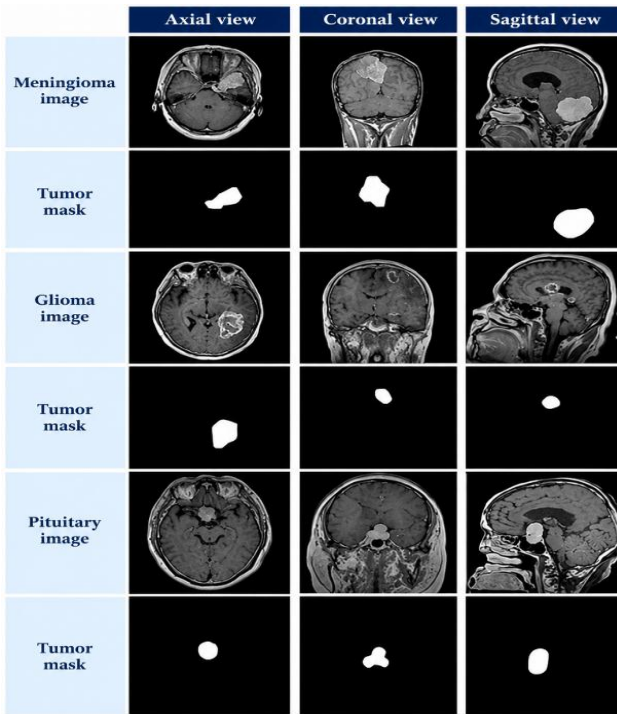


Figure 2 Representative T1-weighted contrast-enhanced MRI images and corresponding tumor masks for the three tumor categories and three imaging planes in the Figshare dataset

3.3.Sequential LSTM Representation

The LSTM stage is intended to process a sequence of

MRI representations ordered by slice position within the same patient and imaging plane. Its purpose is to retain contextual information from preceding sequence elements and produce an enhanced representation for subsequent segmentation. The source manuscript does not establish a complete volumetric reconstruction; the term ordered inter-slice context is therefore used throughout.

A compact LSTM formulation is sufficient to describe the recurrent operation. For input x_t and previous hidden state h_{t-1} , the gated state is defined as follows.

$$f_t = \sigma(W_f[x_t, h_{t-1}]) + b_f \quad (1)$$

$$i_t = \sigma(W_i[x_t, h_{t-1}]) + b_i \quad (2)$$

$$\hat{c}_t = \tanh(W_c[x_t, h_{t-1}]) + b_c \quad (3)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \hat{c}_t \quad (4)$$

$$o_t = \sigma(W_o[x_t, h_{t-1}]) + b_o \quad (5)$$

$$h_t = o_t \odot \tanh(c_t) \quad (6)$$

Here, t denotes ordered slice position rather than acquisition time; σ is the logistic sigmoid, $\tanh$ is the hyperbolic tangent, and $\odot$ denotes element-wise multiplication. The formulation summarizes the six gate/state operations in the original manuscript without reproducing elementary derivations.

3.4.EfficientNet Encoder and Attention-Enhanced U-Net

The segmentation network uses an EfficientNet encoder followed by a U-Net decoder. EfficientNet is based on compound scaling of depth, width, and input resolution [15]. In the documented configuration the encoder is identified as EfficientNet-B4, with four resolution stages and a reported base filter count of 24. The source does not specify pretraining, weight initialization, or whether the encoder was frozen or fine-tuned; these details are therefore not inferred. Additive attention gates operate on encoder-decoder skip pathways. For encoder feature x and decoder gating signal g , the attention intermediate representation can be expressed as $q = \psi^T(\text{ReLU}(W_x + U_g + b)) + b_\psi$, followed by $\alpha = \sigma(q)$ and attended feature $\hat{x} = \alpha \odot x$. The gate is used to regulate feature transmission before concatenation with the decoder representation. The mechanism is therefore described as additive attention rather than transformer self-attention or CBAM [11,12].

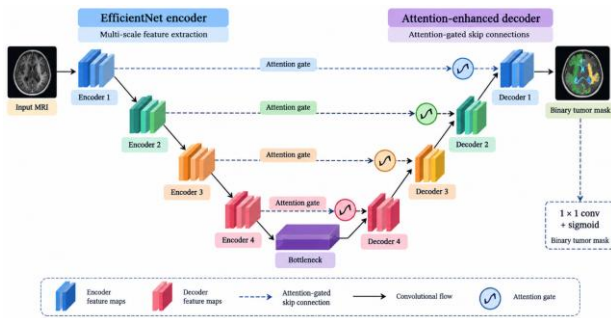


Figure 3 Detailed segmentation architecture showing the EfficientNet encoder, attention-gated skip pathways, U-Net decoder and binary output

3.5.Training and Experimental Protocol

Table 2 Documented Training Configuration and Data-Augmentation Strategy

Parameter	Documented setting
Segmentation optimizer	AdamW
Recurrent-stage optimizer	RMSprop
Learning rate	Not recoverable from source manuscript
Batch size	8
Maximum epochs	500
Dropout	0.5
Encoder	EfficientNet-B4
Encoder/decoder depth	4 resolution stages
Convolutions	3×3; 24 reported base filters
Activation	ReLU in documented hidden layers
Output	1×1 convolution + sigmoid
Loss	Dice-based loss
Attention	Additive attention gates
Augmentation	Rotation, flipping, scaling/zooming, elastic deformation

Augmentation was applied during training and paired transformations were applied consistently to the MRI image and its binary mask. The source does not provide augmentation probabilities or parameter

ranges. The manuscript reports an 80/10/10 patient-level partition and states that all slices from a patient were kept in one subset. However, the available documentation does not provide sufficient implementation detail to independently verify the exact partition file or sequence-construction procedure. Early stopping is stated in the experimental protocol, but the source does not document its stopping criterion. The learning-rate value is also not recoverable from the manuscript text and is therefore not fabricated here.

3.6.Evaluation Metrics

Primary segmentation evaluation uses the Dice similarity coefficient and Jaccard index. For predicted mask P and ground truth G , $DSC = 2|P \cap G| / (|P| + |G|)$, while $JI = |P \cap G| / |P \cup G|$. For identical binary predictions and evaluation units, DSC and JI satisfy $JI = DSC / (2 - DSC)$. Pixel accuracy is supplementary because the large background region can dominate its value. The source manuscript contains inconsistent accuracy values and a U-Net JI that exceeds its DSC; those values are not used as validated primary outcomes.

3.7.Literature-Based Baselines

U-Net, Attention U-Net, and Residual Attention U-Net values are quoted from Hussain et al. [11] rather than independently retrained. The reported Residual Attention U-Net values are DSC 91.10%, JI 89.30%, and accuracy 93.35%; Attention U-Net is reported at DSC 86.12% and JI 85.50%. The source also quotes U-Net DSC 83.45%, while its JI value is not retained because the reported pair is mathematically inconsistent. These values are contextual reference points only and are not controlled head-to-head baselines.

3.8.Patient-Level Experimental Protocol

The experimental protocol is described at the patient level because the dataset contains multiple images associated with each patient. Under the reported 80/10/10 arrangement, patient identifiers are intended to determine subset membership so that images originating from one patient do not occur in more than one subset. This is the appropriate unit for leakage control when a single patient contributes multiple slices. The manuscript nevertheless does not provide the actual partition file or a reproducible

patient list, so the revised conference paper reports the patient-level strategy as documented rather than claiming independent verification of its implementation. Preprocessing was intended to be applied consistently across the dataset while preserving the integrity of evaluation subsets. Spatial resizing was performed from 512×512 to 256×256 pixels, followed by intensity preparation. Training augmentation included geometric transformations such as rotation, flipping, scaling or zooming, and elastic deformation. Transformations were paired between image and mask, and the source states that validation and test data were not augmented. The exact probabilities and parameter ranges are not recoverable from the source manuscript and are therefore not introduced here. The recurrent and segmentation components were associated with different optimizers. RMSprop was reported for the sequential recurrent stage and AdamW for the segmentation network. The segmentation configuration used a batch size of 8, a maximum of 500 epochs, and dropout of 0.5. Early stopping is mentioned in the experimental protocol, but its monitoring threshold and patience are not documented. The learning-rate field is blank in the available manuscript text; consequently, no numerical learning rate is inserted into the conference version. Similarly, no claim is made about learning-rate scheduling, weight-decay magnitude, hardware, software versions, or random seed because these details are not documented. The reported experimental design therefore supports a concise reproducibility statement: the dataset, image size, broad preprocessing operations, patient-level split, optimizer assignments, batch size, maximum epochs, dropout, loss family, and augmentation categories are specified; several lower-level implementation parameters remain unresolved. These missing details should be supplied by the authors if available before final camera-ready submission, particularly if the conference requires complete computational reproducibility.

3.9. Architectural Interpretation

The proposed architecture should be understood as a sequential composition rather than as a single newly invented neural-network primitive. The LSTM stage provides a recurrent representation of an ordered

image sequence. The EfficientNet encoder then performs hierarchical convolutional feature extraction, and the U-Net decoder reconstructs a spatially resolved prediction. Attention gates intervene on the skip pathways to modulate encoder information before fusion with decoder features. This organization allows the paper to distinguish the purpose of each component without attributing the complete segmentation result to any single block. The recurrent index requires particular care in interpretation. The MRI data are static images, and the source does not establish biological temporal acquisition. The sequence therefore represents spatially ordered slices. If the ordering is incorrect or slices from different patients or planes are mixed, the recurrent representation would not correspond to meaningful inter-slice context. For this reason, sequence construction is a methodological dependency rather than a minor implementation detail. The present paper explicitly limits its claim to an intended ordered-slice representation because the exact construction procedure is incompletely documented. The EfficientNet encoder and attention decoder are likewise treated as established components. EfficientNet was introduced as a family of compound-scaled convolutional networks [15], while attention U-Net demonstrated additive attention gates for selective skip-feature transmission [11]. The use of these components provides a coherent segmentation backbone but should not be presented as independent algorithmic novelty. The architectural contribution is the integration of the preceding sequential representation stage with this segmentation backbone and the empirical examination of the resulting pipeline on the Figshare dataset.

4. Results

4.1. Dataset and Experimental Configuration

The study evaluates the proposed framework on the 3,064-image, 233-patient Figshare dataset using the documented 80/10/10 patient-level partition. The experimental configuration combines the sequential recurrent stage with the EfficientNet-B4 attention U-Net described above. No external dataset was used.

4.2. Quantitative Segmentation Results

Table 3 Reported Segmentation Performance and Contextual Literature Values

Model / source	DSC (%)	JI (%)	Interpretation
U-Net [11], literature	83.45	Not verified	Contextual reference
Attention U-Net [11], literature	86.12	85.50	Contextual reference
Residual Attention U-Net [11], literature	91.10	89.30	Contextual reference
Proposed sequential LSTM + attention EfficientNet U-Net	92.25	91.41*	Reported present-study value

The proposed framework reports a DSC of 92.25%. A JI of 91.41% is also reported, but the pair should be checked against the original prediction masks before the value is treated as mathematically validated. The source manuscript also contains inconsistent pixel-accuracy values, so accuracy is omitted from the principal result claim. Because the comparator models were not independently retrained, the table does not establish performance superiority or isolate the contribution of any individual architectural component.

4.3. Qualitative Segmentation

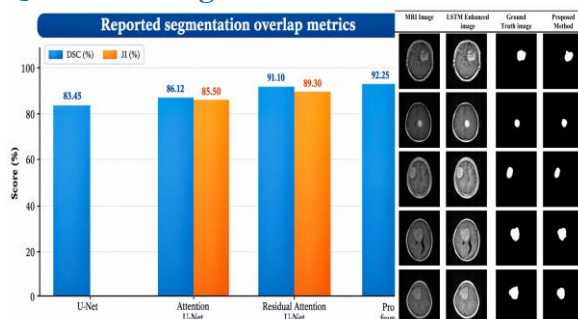


Figure 4 Consolidated quantitative and qualitative evidence: left, reported DSC/JI comparison; right, representative MRI, LSTM-enhanced representation, reference mask and predicted mask

The qualitative examples show predicted tumor regions that broadly correspond to the reference masks, with some boundary differences. The source manuscript does not document an objective case selection rule or independently establish that all displayed examples originate from a held-out test subset. They are therefore treated as illustrative rather than as a statistically representative test-set sample.

4.4. Comparison with Previous Studies

The closest same-dataset comparison identified in the manuscript is the EfficientNetB4-based multi-scale attention U-Net of Preetha et al. [16], which reported a Dice value of 0.9339. That value is higher than the present reported DSC of 0.9225, but the studies differ in architecture and experimental protocol. The present result should therefore be interpreted as dataset-specific evidence for the proposed integration rather than as evidence of state-of-the-art performance.

4.5. Ablation and Computational Evidence

No component-wise ablation was performed. Consequently, the available experiments do not separately quantify the effect of the LSTM stage, EfficientNet encoder, attention gates, or loss formulation. Parameter count, memory consumption, training time, and inference time were also not measured. These omissions prevent attribution of the observed segmentation result to individual components.

5. Discussion

The reported DSC of 92.25% indicates substantial overlap between the predicted and reference tumor regions within the evaluated Figshare setting. The result is consistent with the intended use of overlap-based segmentation metrics, but it should not be interpreted as evidence of clinical performance or generalization beyond the evaluated dataset. The reported JI requires verification against the underlying predictions because the two overlap measures are mathematically related. This verification is particularly important before the numerical pair is used in a final published comparison. The principal methodological idea is the use of sequential recurrent processing before segmentation. Neighboring MRI slices can contain related anatomical and pathological information when they originate from the same patient, imaging

plane, and anatomical order. The LSTM stage is intended to encode this relationship before the EfficientNet encoder extracts hierarchical features. However, the manuscript does not document the exact sequence length, ordering implementation, padding, recurrent dimensionality, or whether the recurrent parameters were frozen or jointly updated. The present evidence therefore supports feasibility of the pipeline but does not demonstrate that the LSTM stage itself improves segmentation. EfficientNet serves as the convolutional encoder rather than as an independent novelty claim. Its compound-scaling strategy provides a structured feature-extraction backbone [15], and EfficientNet-based attention segmentation has already been investigated on the Figshare dataset [16,17]. Similarly, additive attention gates are an established mechanism for selective skip-feature transmission [11]. Their role in the present framework is architectural integration. Without a matched model with the attention gates removed, their independent contribution cannot be quantified. The literature comparison requires particular caution. The reported Residual Attention U-Net values provide a useful contextual reference, but those models were not retrained using the present patient partition, preprocessing, augmentation, optimizer, and training procedure. The same limitation applies to the comparison with Preetha et al. [16]. Differences in data handling can affect segmentation metrics independently of architecture. Therefore, the present paper reports published values descriptively and does not calculate a performance margin table or assign a ranking among methods. Generalization is also unresolved. The experiments use one public dataset, one imaging modality, three tumor categories, and two-dimensional image records. No external scanner, institution, modality, or dataset shift was evaluated. The approach also does not provide uncertainty estimation or explainability. These limitations do not invalidate the reported dataset-specific experiment, but they constrain claims about robustness and clinical deployment. A radiologist would remain responsible for reviewing any segmentation output in a prospective workflow, and no inference-time or memory claim is made because those quantities were not measured.

6. Limitations

Five limitations are central to interpretation. First, evaluation is restricted to one public Figshare dataset and lacks external validation. Second, the exact construction and ordering of inter-slice sequences are not fully documented in the available experimental record. Third, no component-wise ablation isolates the LSTM, EfficientNet, attention, or loss contributions. Fourth, literature comparator models were not independently retrained under a common protocol. Fifth, several reproducibility details including the recoverable learning rate, exact augmentation parameters, recurrent layout, pretraining status, random seed, software versions, and computational cost are incompletely documented. These limitations should be addressed before making claims about generalization, component attribution, or clinical utility.

Conclusion

This conference version presents a compact investigation of a sequential LSTM-based MRI representation stage followed by an attention-enhanced EfficientNet U-Net for binary brain-tumor segmentation on the Figshare dataset. The reported framework achieved a DSC of 92.25%, while the accompanying JI of 91.41% requires verification against the underlying predictions. The evidence supports feasibility of the architectural integration within the evaluated dataset but does not establish component-wise superiority or external clinical generalization. Future evaluation should use transparent patient-level partitioning, controlled ablation, independently reproduced baselines, repeated runs with uncertainty estimates, and external validation.

Declarations

Funding

The authors declare that no financial support was received for the research, authorship, or publication of this work.

Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Ethics Approval and Consent to Participate

Not applicable. This study involved secondary computational analysis of a publicly available

Figshare brain-tumor MRI dataset and did not involve new participant recruitment, clinical intervention, or direct interaction with human participants.

Consent for Publication

Not applicable. No new identifiable participant information was collected or generated as part of this study.

Data Availability

The MRI images and corresponding tumor masks used in this study are from the publicly available Figshare brain-tumor dataset reported by Cheng et al. [1]. Additional information concerning the implementation and experimental configuration is available from the corresponding author upon reasonable request.

Code Availability

The computational implementation used in this study is not publicly deposited. Additional implementation details may be obtained from the corresponding author upon reasonable request.

Author Contributions

K. Santhi: Conceptualization, methodology, investigation, data analysis, manuscript preparation, and revision.

K. Radha Krishna: Supervision, methodological review, interpretation of results, and manuscript review and revision.

Acknowledgements

The authors have no specific acknowledgements to declare.

AI-Assisted Writing Statement

During manuscript preparation, AI-assisted tools were used only to support language improvement, grammar correction, and readability. The authors reviewed and validated the scientific content and remain responsible for the accuracy, interpretation, and integrity of the manuscript.

References

- [1]. Cheng, J., Huang, W., Cao, S., Yang, R., Yang, W., Yun, Z., Wang, Z. and Feng, Q., "Brain tumor dataset," Figshare, 2017. <https://doi.org/10.6084/m9.figshare.1512427>
- [2]. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al., "The

multimodal brain tumor image segmentation benchmark (BRATS)," IEEE Transactions on Medical Imaging, 34(10), 1993–2024, 2015. <https://doi.org/10.1109/TMI.2014.2377694>.

- [3]. Pereira, S., Pinto, A., Alves, V. and Silva, C.A., "Brain tumor segmentation using convolutional neural networks in MRI images," IEEE Transactions on Medical Imaging, 35(5), 1240–1251, 2016. <https://doi.org/10.1109/TMI.2016.2538465>.
- [4]. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M. and Larochelle, H., "Brain tumor segmentation with Deep Neural Networks," Medical Image Analysis, 35, 18–31, 2017. <https://doi.org/10.1016/j.media.2016.05.004>.
- [5]. Wang, G., Li, W., Ourselin, S. and Vercauteren, T., "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks," in Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Springer, pp. 178–190, 2018. https://doi.org/10.1007/978-3-319-75238-9_16.
- [6]. Naceur, M.B., Saouli, R., Akil, M. and Kachouri, R., "Fully automatic brain tumor segmentation using end-to-end incremental deep neural networks in MRI images," Computer Methods and Programs in Biomedicine, 166, 39–49, 2018. <https://doi.org/10.1016/j.cmpb.2018.09.007>.
- [7]. AlBadawy, E.A., Saha, A. and Mazurowski, M.A., "Deep learning for segmentation of brain tumors: Impact of cross-institutional training and testing," Medical Physics, 45(3), 1150–1158, 2018. <https://doi.org/10.1002/mp.12752>.
- [8]. Feng, X., Tustison, N.J., Patel, S.H. and Meyer, C.H., "Brain tumor segmentation using an ensemble of 3D U-Nets and overall survival prediction using radiomic features," Frontiers in Computational Neuroscience, 14, 25, 2020. <https://doi.org/10.3389/fncom.2020.00025>.
- [9]. Magadza, T. and Viriri, S., "Deep learning for

- brain tumor segmentation: A survey of state-of-the-art,” *Journal of Imaging*, 7(2), 19, 2021.
<https://doi.org/10.3390/jimaging7020019>.
- [10]. Ronneberger, O., Fischer, P. and Brox, T., “U-Net: Convolutional networks for biomedical image segmentation,” in *MICCAI 2015*, Springer, pp. 234–241, 2015.
https://doi.org/10.1007/978-3-319-24574-4_28.
- [11]. Oktay, O., Schlemper, J., Le Folgoc, L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B. and Rueckert, D., “Attention U-Net: Learning where to look for the pancreas,” *arXiv:1804.03999*, 2018.
<https://doi.org/10.48550/arXiv.1804.03999>.
- [12]. Woo, S., Park, J., Lee, J.-Y. and Kweon, I.S., “CBAM: Convolutional Block Attention Module,” in *ECCV*, pp. 3–19, 2018.
https://doi.org/10.1007/978-3-030-01234-2_1.
- [13]. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Landman, B.A., Yuille, A.L. and Zhou, Y., “TransUNet: Transformers make strong encoders for medical image segmentation,” *arXiv:2102.04306*, 2021.
<https://doi.org/10.48550/arXiv.2102.04306>.
- [14]. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J. and Maier-Hein, K.H., “nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation,” *Nature Methods*, 18, 203–211, 2021.
<https://doi.org/10.1038/s41592-020-01008-z>.
- [15]. Tan, M. and Le, Q.V., “EfficientNet: Rethinking model scaling for convolutional neural networks,” in *ICML*, PMLR 97, 6105–6114, 2019.
- [16]. Preetha, R., Priyadarsini, M.J.P. and Nisha, J.S., “Brain tumor segmentation using multi-scale attention U-Net with EfficientNetB4 encoder for enhanced MRI analysis,” *Scientific Reports*, 15, 9914, 2025.
<https://doi.org/10.1038/s41598-025-94267-9>.
- [17]. Chen, Y., Quan, L., Long, C., Chen, Y., Zu, L. and Huang, C., “Brain tumor segmentation based on the U-Net++ network with EfficientNet encoder,” *Technology and Health Care*, 32(Suppl. 1), 183–195, 2024.
<https://doi.org/10.3233/THC-248016>.
- [18]. Hochreiter, S. and Schmidhuber, J., “Long short-term memory,” *Neural Computation*, 9(8), 1735–1780, 1997.
<https://doi.org/10.1162/neco.1997.9.8.1735>.
- [19]. Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K. and Woo, W.-C., “Convolutional LSTM network: A machine learning approach for precipitation nowcasting,” *NeurIPS* 28, 2015.
- [20]. Almiahi, O.M., Albu-Salih, A.T. and Alhajim, D., “A novel ConvLSTM-based U-Net for improved brain tumor segmentation,” *IEEE Access*, 12, 157346–157358, 2024.
<https://doi.org/10.1109/ACCESS.2024.3483562>.
- [21]. 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. and Sánchez, C.I., “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, 42, 60–88, 2017.
<https://doi.org/10.1016/j.media.2017.07.005>.
- [22]. Kingma, D.P. and Ba, J., “Adam: A method for stochastic optimization,” *ICLR*, 2015.
- [23]. Loshchilov, I. and Hutter, F., “Decoupled weight decay regularization,” *ICLR*, 2019.
- [24]. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S. and Cardoso, M.J., “Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations,” *Springer*, pp. 240–248, 2017.
https://doi.org/10.1007/978-3-319-67558-9_28.
- [25]. Lin, T.-Y., Goyal, P., Girshick, R., He, K. and Dollár, P., “Focal loss for dense object detection,” *ICCV*, pp. 2980–2988, 2017.
<https://doi.org/10.1109/ICCV.2017.324>.