Received: 20 September 2024 | Revised: 19 November 2024 | Accepted: 28 November 2024 | Published online: 16 January 2025

RESEARCH ARTICLE

Journal of Computational and Cognitive Engineering
2026, Vol. 5(2) 341–353
DOI: 10.47852/bonviewJCCE52024394

BON VIEW PUBLISHING

GRU-ASPP U-Net: Enhanced Multiple Sclerosis Lesion Segmentation Through Temporal-Spatial Feature Integration in Longitudinal MRI

Ramya Palaniappan[1,*] and Siva Rathinavelayutham[1]

[1]*Department of Computational Intelligence, SRM Institute of Science and Technology–Kattankulathur Campus, India*

Abstract: Identifying the different stages of Multiple Sclerosis (MS) is essential for accurately predicting disease progression, often requiring the capabilities of deep learning networks. The spatial and temporal dependencies within image-based biomarkers frequently present irregular capture challenges across MRI scans taken at different time points, primarily due to unsynchronized multi-scale features and their corresponding network channels. Consequently, this may lead to false predictions of disease progression and inaccurate MS lesion segmentation. In this paper, we propose a novel deep learning pipeline that integrates advanced modules, namely Gated Recurrent Unit (GRU), Atrous Spatial Pyramid Pooling (ASPP), and Squeeze and Excitation (SE) blocks, within a U-Net architecture to effectively manage multi-scale features and channel relationships. It is simply abbreviated as GRU-ASPP U-Net. The GRU modules at skip connections enable temporal feature capture, the ASPP module in the bottleneck layer facilitates multi-scale feature extraction, and SE blocks perform channel-wise feature recalibration. Evaluation metrics, including the Jaccard Coefficient (IoU) and Dice Coefficient, indicate that the proposed pipeline achieves promising prediction outcomes, with a Dice score of 0.87 and IoU of 0.79 (Fold 3). Moreover, this approach combines spatial and temporal analysis, providing a more stringent measure of segmentation accuracy and enhanced sensitivity to subtle overlaps in MS lesion segmentation.

Keywords: multiple sclerosis lesion segmentation, deep learning, Gated Recurrent Unit (GRU), Atrous Spatial Pyramid Pooling (ASPP), longitudinal MRI analysis

1. Introduction

Multiple Sclerosis (MS) is an autoimmune disease that affects the central part of the nervous system and worsens through time mainly due to the damage of the brain and spinal cord [1]. The disease results from lesion formation in specific areas of the brainstem and spinal cord that can be demonstrated by Magnetic Resonance Imaging (MRI). Accumulation of these lesions over time is termed multiple sclerosis; therefore, accurate identification and segmentation of lesions are important for disease monitoring and therapeutic planning [2, 3].

Clinical assessment of MS progression heavily relies on the analysis of longitudinal MRI scans, where the identification and quantification of new or enlarging lesions serve as critical biomarkers [4]. Traditional manual segmentation of these lesions by expert radiologists is not only time-consuming but also subject to inter-rater variability, creating a pressing need for automated segmentation approaches [5]. Multiple MRI sequences, including T1-weighted, T2-weighted, Fluid Attenuated Inversion Recovery (FLAIR), and Proton Density (PD) images, are typically employed for comprehensive lesion analysis, with each modality providing unique insights into lesion characteristics [6].

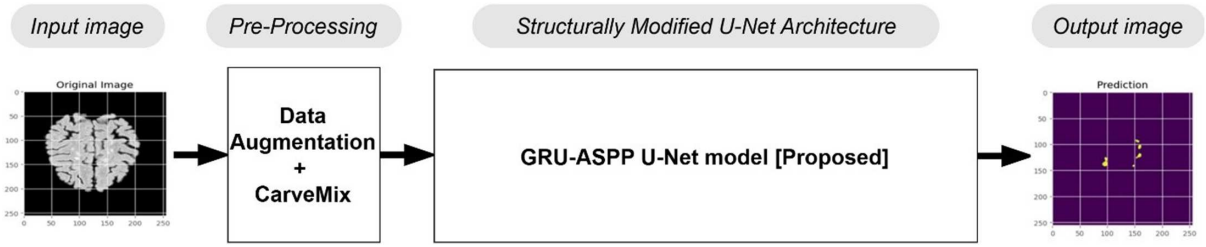
Recent advances in deep learning, particularly in medical image analysis, have shown promising results in automated lesion segmentation. The U-Net architecture and its variants have emerged as powerful tools for biomedical image segmentation [7]. However, existing approaches face several challenges, including the effective capture of temporal dependencies between sequential MRI scans, the integration of multi-scale features, and the management of channel-wise information [8].

Several studies have attempted to address these challenges through various architectural modifications. For instance, squeeze and excitation blocks have been employed to enhance channel-wise feature recalibration [9], while attention mechanisms have been integrated to focus on relevant spatial features [10]. Additionally, approaches incorporating temporal information through recurrent neural networks have shown improved performance in longitudinal analysis [11].

To address the limitations of existing methods, we propose a novel GRU-ASPP U-Net architecture that combines three key innovations: (i) Gated Recurrent Unit (GRU) blocks at skip connections to capture temporal dependencies between consecutive

***Corresponding author:** Ramya Palaniappan, Department of Computational Intelligence, SRM Institute of Science and Technology–Kattankulathur Campus, India. Email: rp3516@srmist.edu.in

Figure 1
Pipeline for MS lesion segmentation using the proposed GRU-ASPP U-Net architecture



MRI scans, (ii) Atrous Spatial Pyramid Pooling (ASPP) module for multi-scale feature extraction, and (iii) Squeeze and Excitation (SE) blocks for dynamic channel recalibration. Furthermore, we introduce CarveMix, a lesion-aware data augmentation technique that enhances the model's generalization capabilities by intelligently combining features from different images based on lesion characteristics. Figure 1 illustrates the pipeline for MS lesion segmentation using proposed GRU-ASPP U-Net architecture.

The main contributions of this work can be summarized as follows:

1) Development of a novel U-Net variant that effectively integrates temporal and spatial information for MS lesion segmentation
2) Implementation of GRU blocks at skip connections to capture sequential dependencies in longitudinal MRI data
3) Integration of ASPP and SE modules to enhance multi-scale feature extraction and channel-wise feature recalibration

The rest of the paper is organized as follows: Section 2 provides a review of existing U-Net variant models, including hybrid U-Net models and structurally modified U-Net architectures, in detail. Section 3 describes the proposed GRU-ASPP U-Net model, with a detailed explanation of its mathematical representation and the key inclusion blocks. Section 4 presents the results and discussion, comparing the outcomes of the proposed model with existing state-of-the-art models. Finally, the conclusion and future scope of the research study are outlined.

2. Related Works

Accurately detecting and segmenting MS lesions in MRI images using deep learning networks is challenging due to the need to train on features with rapidly varying spatial and temporal dependencies. These elements are crucial for medical imaging, aiding in diagnosis and treatment planning. The U-Net architecture and its variants have been extensively used in this field due to their flexible structure, which can be adjusted to enhance the effectiveness of biomedical image segmentation. This section reviews several existing models that enhance segmentation accuracy through hybrid U-Net models and structurally modified U-Net architectures, as discussed below.

2.1. Hybrid U-Net models

Previous works have suggested various modifications to the U-Net architecture to enhance segmentation performance. These include deeper networks, residual units, and attention mechanisms to improve feature learning and ensure stable training [4, 7]. Integrating backbone networks like VGG-16 and ResNet50 with U-Net has shown improved segmentation results by optimizing feature extraction effectiveness and minimizing feature information loss [5, 6]. The novel Path Aggregation U-Net (PAU-Net) architecture introduces bottom-up path aggregation encoders and efficient feature pyramids, reducing noise and refining mask predictions, thereby achieving better results on the BraTS dataset [1]. U-Net++, which incorporates a Deep Supervision Mechanism (DSM), improves segmentation accuracy by 2.8% compared to the basic U-Net. It also utilizes a dilation operator to enhance feature extraction even with limited labeled data [3, 10]. Similarly, models like Attention U-Net and R2 Attention U-Net have been introduced to focus on the most relevant features, making segmentation more accurate and efficient [2]. Depthwise-separable convolutions combined with shuffle attention further enhance the nnU-Net model's performance without adding significant computational overhead, making it suitable for practical applications with lower computing resources [8]. Lastly, hybrid segmentation networks, such as those combining 3D CNN with U-Net, can yield more accurate predictions due to the fine-tuned strengths of each model [9]. Likewise, Transformer UNet++ with the MobileNetv3 module [12], and Hybrid Dilated Convolution based Adaptive MobileNet (HDC-AMNet) module [13–16] are introduced to achieve higher performance when compared to the baseline U-Net model, respectively.

2.2. Models of modified U-Net architectures

The customized U-Net architecture captures both spatial and temporal features in FLAIR MRI samples, focusing on the spatial correlation between sequential axial slices by incorporating a Convolutional LSTM at the bottleneck of the U-Net architecture [9]. Similarly, the 3D U-Net model employs an Online Hard Example Mining (OHEM) strategy to address class imbalance during training and testing on the ISBI 2015 dataset, achieving a Dice score of 90.1% [10]. Furthermore, the nnU-Net model benefits from lesion-aware data augmentation procedures, such as axial subsampling and CarveMix. This approach supports the contraction and expansion paths of four resolution levels through sequences of 3D convolutions, instance normalization, and LeakyReLU activation. As a result, it achieved an average Dice score of 0.510 and an F1 score of 0.552 for patients with new lesions [11]. A two-path architecture based on the U-Net model also uses residual blocks in both up-sampling and down-sampling paths, along with Convolutional Gated Recurrent Units (GRUs) at the skip connections, to capture temporal information from MRI samples. This model achieved a lesion-wise true positive rate of 0.82 and a false positive rate of 0.17 [17]. Similarly, a patch-based convolutional neural network (CNN) model with a 3D ResNet block and a spatial-channel attention module was introduced [13]. The attention guidance mechanism masks old lesions to focus solely on segmenting new lesions, resulting in a lesion-wise true positive rate of 74.2% and a false positive rate of 26.4% [18]. The lightweight

implementation of the U-Net model focuses on reducing the data augmentation step, ensuring that minimal training data is sufficient for accurately segmenting newly formed MS lesions. This model demonstrates remarkable results on the BRATS dataset, achieving a mean Intersection over Union (IoU) of 89%, surpassing standard benchmark algorithms [19, 20]. Finally, segmentation of uneven cerebral bleeding lesions is enhanced by incorporating a Residual Octave Convolution (ResOctConv) module within the U-Net model, along with a Mixed Attention Mechanism (MAM), which provides the capability to dynamically handle multi-scale features in CT image slices [21–23]. An overall summary of all the models discussed above is provided in Table 1 below.

3. Proposed Methodology

Figure 2 shows a modified architecture of the U-Net for segmenting newly appearing lesions in multiple sclerosis on the MRI scan. Both spatial and temporal features obtained from 2D MRI sequences at different time points are effectively handled with additional blocks in the proposed GRU-ASPP U-Net model, enhancing segmentation accuracy.

The architecture starts by concatenating baseline and follow-up MRI modalities, (randomly select two different MRI scan at any timepoint) as input to the model. The encoder path comprises of four levels, wherein each of them have time-distributed convolution layers having Squeeze and Excitation (SE) block. These SE blocks are centered on channel-wise adaptively recalibrating feature responses, improving the discriminativeness of the presented model. The SE blocks operate through a two-phase mechanism: first, the feature maps are down-sampled to a single value per feature map through pooling, second, a two-layer convolutional neural network applies specific weights to channels of feature maps for feature recalibration. Gated Recurrent Unit (GRU) blocks is the key modules act as memory units that capture feature evolutions over time between adjacent MRI scans. With their update and reset gates, the information flow can be controlled and histories of important features maintained while new observations are incorporated. This temporal processing capability is especially useful in the ability to detect changes in lesion patterns over time. Moreover, GRU blocks incorporated into the skip connections. The Atrous Spatial Pyramid Pooling (ASPP) module act as a bottleneck layer to capture the multiple scale contextual information successfully. This is achieved by operating both parallel atrous convolutions with different rates are used to capture features at different scales in parallel. This multi-scale processing is crucial for handling lesions of varying sizes and shapes effectively. In the decoder path, the architecture employs a series of up-convolutional blocks. At each level, the output from the corresponding GRU block is concatenated with up-sampled feature maps, effectively combining spatial and temporal information. The final stage includes convolutional operations with sigmoid activation to produce probability maps for lesion segmentation.

3.1. SE block

It is used to function as a channel recalibration module that dynamically changes the corresponding weights of feature channels. Initially, the operation proceeds to the squeeze phase where the global average pooled layer (μ_{global}) reduces each of the feature channel, (X_i), to a single number, (z_i), thereby effectively creating a channel descriptor. This can be expressed as:

$$z_i = \mu_{\text{global}}(X_i) \text{ for } i = 1, 2, 3, \ldots, n \tag{1}$$

Where n represents the number of convolutional channels. The values are collected into a vector $Z = [z_1, z_2, \ldots., z_n]$, representing the global distribution of channel-wise feature responses. The excitation phase then processes this information through a two-layer neural network with weights (W_1, W_2) and biases (b_1, b_2), creating a nonlinear transformation:

$$a = \sigma(W_1 Z + b_1) \tag{2}$$

$$s = \sigma(W_2 a + b_2) \tag{3}$$

Where, σ is the activation function of either sigmoid or ReLu and s is the output vector. The final step applies these learned weights to rescale the original feature channels:

$$\widetilde{X}_i = s_i . X_i \text{ for } i = 1, 2, 3, \ldots, n \tag{4}$$

This channel-wise multiplication allows the network to emphasize informative features while suppressing less useful ones, enhancing the model's representational power.

3.2. ASPP module

The ASPP module enhances the model's ability to capture multi-scale contextual information through parallel atrous convolutions with different dilation rates. The process begins with an input feature map $X \in R^{HWC}$ and applies multiple parallel convolution operations with varying dilation rates:

$$Y_i = Conv_{r_i}(X) \tag{5}$$

Where $Conv_{r_i}$ represents convolution with dilation rate r_i. Simultaneously, global context is captured through:

$$Y_{global} = Conv_{1*1}\left(\mu_{global}(X)\right) \tag{6}$$

Then, these multi-scale features are then concatenated:

$$Y_{Concat} = Concat\left([Y_1, Y_2, \ldots Y_n, Y_{global}]\right) \tag{7}$$

And finally processed through a 1x1 convolution to produce the output:

$$Y_{output} = Conv_{1*1}(Y_{Concat}) \tag{8}$$

This hierarchical feature extraction helps in capturing both fine-grained details and broader contextual information.

3.3. GRU block

The GRU block manages temporal dependencies in the feature space through a sophisticated gating mechanism. The update gate (zt) determines how much previous information should be retained:

$$z_t = \sigma(W_z . [h_{t-1}, x_t] + b_z) \tag{9}$$

While the reset gate (rt) controls the flow of past information:

$$r_t = \sigma(W_r . [h_{t-1}, x_t] + b_r) \tag{10}$$

The candidate activation ($\tilde{h}t$) combines current input with filtered past information:

$$\tilde{h}_t = tanh(W . [r_t \odot h_{t-1}, x_t] + b) \tag{11}$$

Table 1
Summary of related works

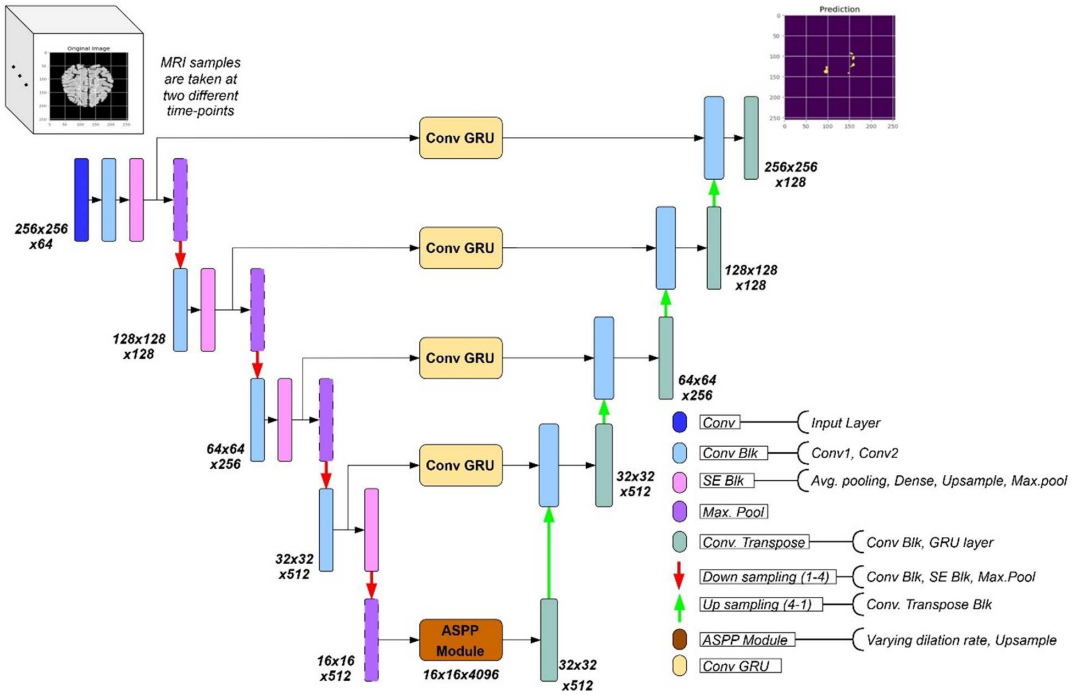
Ref no	Proposed model	Focus	Image modality	Dataset	Evaluation metrics										Outcome values	Research gap
					Accuracy	Resource efficiency	Dice score	Hausdorff distance	Sensitivity	Specificity	IoU	TPR	FPR	DSC		
[1]	PAU-Net	Brain tumor segmentation	Multi-modality MRI	BraTS 2017, 2018	✓	✓									Outperforms SOTA with fewer resources	Traditional methods don't meet clinical standards
[2]	Multiple U-Net variants (3D U-Net, Attention U-Net, R2 Attention U-Net)	Brain tumor segmentation	MRI	BraTS 2020			✓	✓							Varied performance across architectures	Need for automatic, generalized system
[3]	U-Net++DSM	Brain tumor segmentation	MRI	Public databases			✓		✓	✓	✓				Superior to traditional U-Net	Limited labeled data challenges
[4]	Deeper Res U-Net	Brain tumor segmentation	MRI	Test set			✓								Competitive performance	Network degradation in deep architectures
[5]	U-Net-based approach	Brain tumor segmentation	Multi-modal MRI	BraTS-19			✓								92% for entire tumor	Time-intensive manual segmentation
[6]	U-Net with VGG-16	Brain tumor segmentation	MRI	TCGA-LGG			✓				✓				Disc score of 0.9975	Need for improved detection accuracy
[7]	Enhanced U-Net with ResNet50	Brain tumor segmentation	MRI	Test set			✓				✓				3.13% higher IoU than original U-Net	Feature information loss during training
[8]	Efficient nnU-Net	Brain tumor segmentation	MRI	BraTS 2020			✓								Competitive with 2.51 M parameters	High computational complexity
[9]	Ensemble of 3D CNN and U-Net	Brain tumor segmentation	Multi-modal MRI	BraTS-19			✓								Favorable comparison to SOTA. Dice Score:0.47	Need for improved accuracy
[10]	Modified U-Net++	Brain tumor segmentation	MRI	BraTS 2019			✓								Comparable to peer-reviewed work	Need for lightweight efficient models
[11]	nn-UNet with lesion-aware augmentation	MS lesion segmentation	MRI	MICCAI 2021 MSSEG-2			✓		✓						Dice: 0.510, F1: 0.552	Limited data diversity

(*Continued*)

Table 1
(*Continued*)

Ref no	Proposed model	Focus	Image modality	Dataset	Evaluation metrics										Outcome values	Research gap
					Accuracy	Resource efficiency	Dice score	Hausdorff distance	Sensitivity	Specificity	IoU	TPR	FPR	DSC		
[12]	Transformer UNet++ with MobileNetv3	MS classification	MRI	Clinical data	✓		✓			✓					Higher than baseline	Need for efficient classification
[13]	Patch-based CNN with 3DResNet	MS lesion segmentation	FLAIR	ISIB challenge	✓		✓								Better than SOTA	Feature extraction improvement
[14]	Modified Attention U-Net	MS lesion segmentation	FLAIR & T2	Test dataset			✓							✓	82.30%	Multi-modal integration
[15]	Encoder Decoder	Brain Tumor detection	MRI	Clinical data	✓									✓	Acc: 0.9563%	Temporal-spatial feature capture
[16]	Modified Attention U-Net	MS lesion segmentation	T2 & FLAIR	Clinical data	✓										Accuracy: 82.30%	Multi-sequence fusion
[17]	Two-path U-Net with GRU	MS lesion temporal analysis	MRI volumes	Clinical data								✓	✓		TPR: 0.82, FPR: 0.17	Need for temporal information capture
[18]	Two-path CNN with attention	New lesion segmentation	MRI	Clinical data								✓	✓		TPR: 74.2%, FPR: 26.4%	Old vs new lesion differentiation
[24]	FCN with dual pathways	White matter MS lesion segmentation	FLAIR	Clinical data	✓		✓								Sustain standard outcome	Feature extraction efficiency
[25]	2D-UNet with random forest	MS lesion segmentation & counting	MRI	Clinical data			✓			✓	✓				Sustain standard outcome	Need for automated counting
[26]	U-Net with AG, ECA, ASPP	MS lesion segmentation	MRI	Clinical data			✓			✓		✓		✓	Sustain standard outcome	Feature attention and scale
[27]	ViT-GRU hybrid	Alzheimer's detection	MRI	Clinical data			✓							✓	Sustain standard outcome	Model trustworthiness
[28]	Gated Recurrent Unit	Brain Tumor detection	MRI	BTD Dataset											Accuracy: 99.32%	Need for improved accuracy

Figure 2
Proposed GRU-ASPP U-Net model for MS lesion segmentation



The final hidden state is computed as:

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{12}$$

This gating mechanism allows the GRU to effectively manage long-term dependencies while mitigating the vanishing gradient problem, making it particularly suitable for processing sequential MRI data. These three blocks work in concert within the architecture: Here it is shown that the SE blocks improve feature representation per channel, the ASPP module learns multi-scale context in spatial dimensions, and the GRU blocks maintain dependencies between sequentially acquired MRI scans. The SE blocks work feature-wise to adjust the channel's importance, while the ASPP module breaks down the improved features across different scales. The GRU blocks then combine this rich spatial information at each time step to track lesion changeover time points. Combined, it yields an effective a structure that enables addressing the MS lesion segmentation challenge, with the consideration of spatial and temporal properties of inputs.

4. Results and Discussion

In this section, we present a detailed evaluation of the proposed GRU-ASPP U-Net model for MS lesion segmentation using the ISBI 2015 dataset. The dataset consists of chronological MRI images obtained with MS patients along with different modalities (T1 MPRAGE, T2, PD, FLAIR) and the markups by three experts. Thus, the present study is based on a combination of the numerical indices and both quantitative metrics and qualitative assessments, with the emphasis on the performance of the proposed model in the separate validation folds. The experimental setup involved comprehensive preprocessing steps and data augmentation techniques, including the novel CarveMix approach specifically designed for lesion-aware augmentation. We implemented a 5-fold cross-validation strategy, with each fold containing 896 images distributed across training, validation, and testing sets. The model was trained using a combined loss function of binary cross-entropy and Dice loss, optimized using Adam optimizer over 25 epochs on a Google Colab T4 GPU environment.

4.1. Dataset

In this study, involves ISBI 2015 MS Lesion Segmentation dataset which has created to Longitudinal MS Lesion Segmentation Challenge in the 2015 International Symposium on Biomedical Imaging. It includes T1-weighted MRI scans obtained from MS patients over time, which enables the assessment of the performance of computerized lesion segmentation methods.

4.1.1. Dataset composition

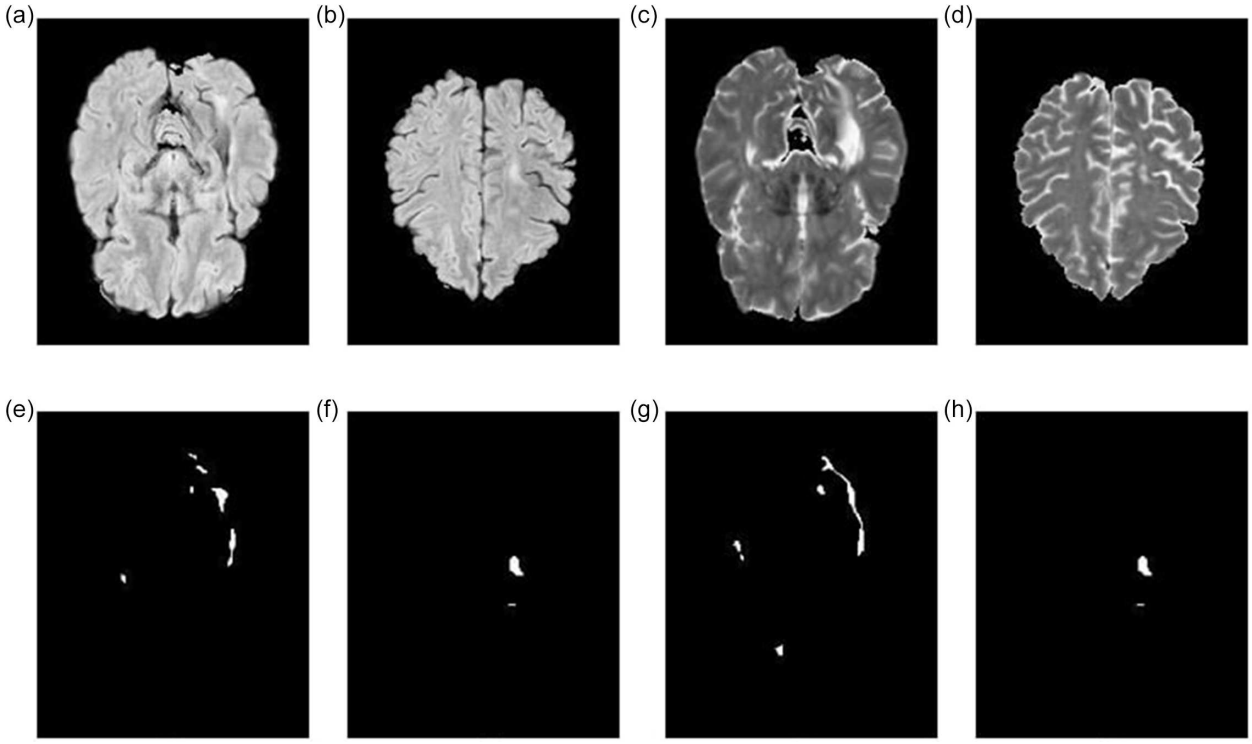
Training Set: Five case subjects are contributed with 22 total time points having an average of 4.4 time points per patient. The mean age at baseline is 43.5 years, and observation period was approximately one year. Ground truth lesion maps produced by expert human annotators are also given for this set.

Test Set: Nine of the 14 patients had global atrophy rates that we consider of borderline statistical significance; the data set includes 57 longitudinal MRI scans, with 4.4 time points per patient on average. Informing this analysis are the mean baseline age of 39.3 years and mean follow-up time of one year. For this set no manual lesion delineations are available for comparison and correlation.

4.1.2. Imaging modalities

For each time point, the dataset includes the following MRI sequences: (i) T1-weighted (T1-w) MPRAGE: Approximately 1 mm^3 isotropic voxel resolution, (ii) T2-weighted (T2-w): 0.82 $\times$ 0.82 $\times$ 2.2 mm^3 voxel size, (iii) Proton Density-weighted (PD-w): 0.82 $\times$ 0.82 $\times$ 2.2 mm^3 voxel size, and (iv) T2-weighted Fluid

Figure 3
Example MRI images from the ISBI-2015 dataset: T2-weighted images (a, b) with corresponding masks annotated by an expert rater (e, f) and FLAIR images (c, d) with equivalent masks marked by rater 1 (g, h)



Attenuated Inversion Recovery (FLAIR): $0.82 \times 0.82 \times 2.2$ mm^3 voxel size. All images were acquired on a 3.0 Tesla MRI scanner.

Figure 3 displays typical MRI samples belonging to the ISBI-2015 dataset to qualitatively demonstrate the T2w and FLAIR modes together with their corresponding masks. T2-weigthed image and its lesion masks are depicted in subfigures (a) and (b) respectively, while subfigures (e) and (f) show the expertise rating of the rater. As with the same location T2w images, subfigures (c) and (d) describe FLAIR images, and the corresponding masks from rater 1 are provided in subfigures (g) and (h). This figure gives an idea of the differences in the image modalities and the expert annotations which provides a basis of understanding the dataset and the segmentation problems to be solved.

4.2. Simulation setup

All experiments were conducted using pre-processed images from the ISBI 2015 dataset. The outer black regions of the MRI (T2 and FLAIR) sequences were cropped, and geometric data augmentations, including random rotations, flipping, and zooming, were applied within the data generator. In addition to these standard augmentation procedures, lesion-aware CarveMix data augmentation was performed. This technique carves and mixes portions of one image into another, based on the location and geometry of the lesions. Examples of augmented samples generated with this approach are shown in Figure 4. Columns C1 and C2 show pairs of T2 MRI images, while Column C3 displays the synthetic augmented sample created by mixing the T2 MRI pairs.

The augmented samples have significantly increased the diversity of the training data, enabling the model to train more effectively and enhancing its robustness, which improves its ability to generalize to new, unseen data. The model was trained using a 5-fold cross-validation strategy as follows:

Fold 1: 896 images total—224 images for validation, 224 images for testing, and 448 images for training.

Fold 2: 896 images total—224 images for validation, 224 images for testing, and 448 images for training.

Fold 3: 896 images total—224 images for validation, 224 images for testing, and 448 images for training.

Fold 4: 896 images total—224 images for validation, 224 images for testing, and 448 images for training.

Fold 5: 896 images total—224 images for validation, 224 images for testing, and 448 images for training.

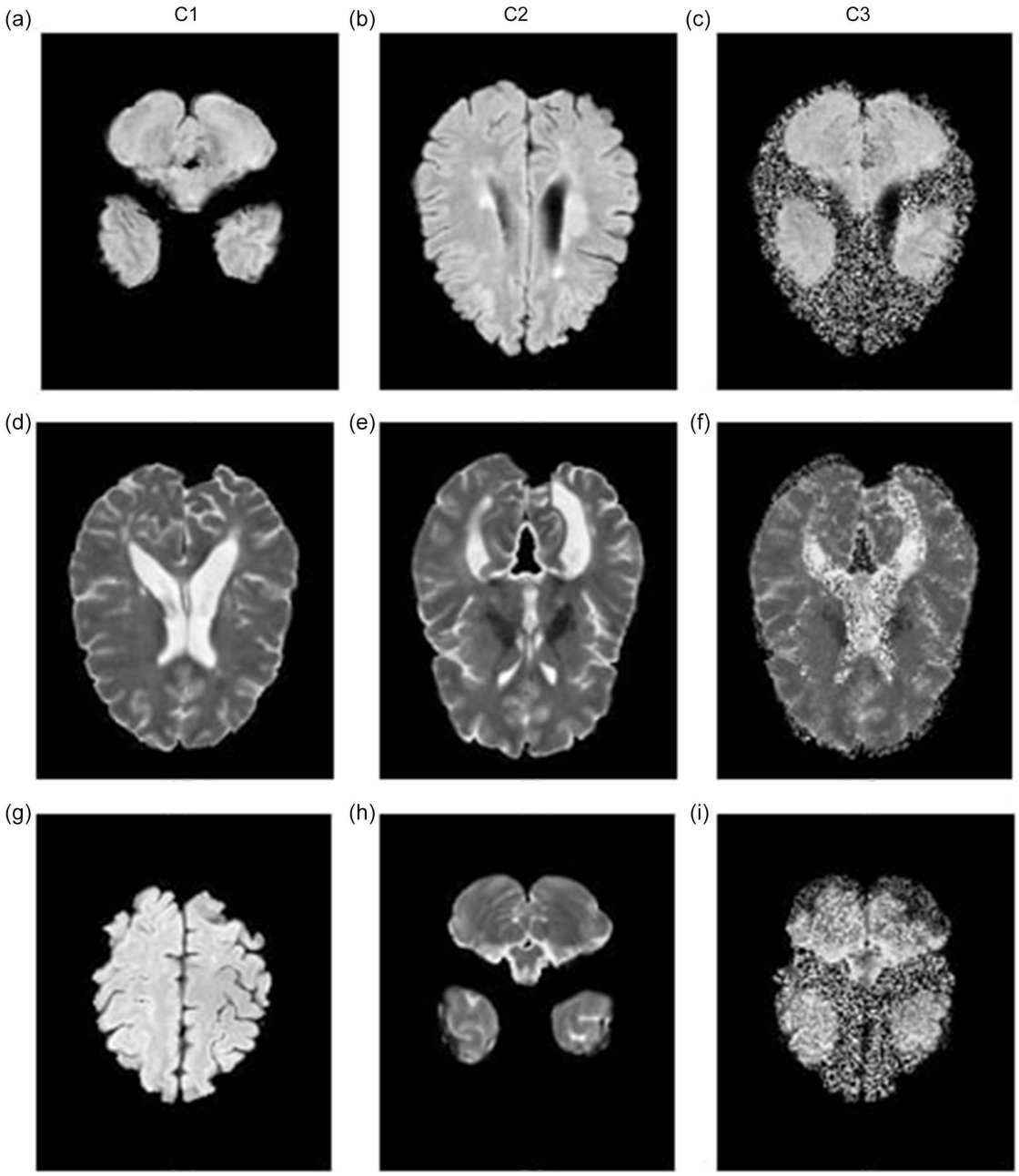
The proposed GRU-ASPP U-Net model is implemented in Python using the Keras library. This model was trained on Google Colab using a T4 GPU. A modified U-Net model with GRU was used, compiled with the Adam optimizer, and trained using a combined loss function of binary cross-entropy and Dice loss. The metrics analyzed for performance evaluation include the Dice coefficient and Jaccard coefficient. Training was conducted over 25 epochs, with the total training time for the proposed model being approximately 14 hours.

4.3. Evaluation metrics

The performance of the proposed model is evaluated by comparing the predicted mask with the ground truth mask provided by expert raters. This evaluation uses the Jaccard coefficient, Dice coefficient, and accuracy as performance metrics.

The Jaccard coefficient, also known as the Intersection over Union (IoU), is calculated by dividing the intersection of the predicted and ground truth masks by their union. The IoU value ranges

Figure 4
CarveMix-augmented MRI samples



from 0 to 1, where 0 indicates no overlap (poorest performance), and 1 indicates perfect overlap (best performance) between the predicted mask and the ground truth mask. The Jaccard coefficient is calculated using Equation (13).

$$Jaccard\ Coeff = \frac{X \cap Y}{X \cup Y} \tag{13}$$

Where X represents a set of pixels in segmented mask and Y represents a set of pixels in a ground truth mask. The Dice coefficient, also known as the Sørensen–Dice coefficient, measures the degree of overlap between the predicted segmentation mask and the ground truth mask. It is calculated using Equation (14).

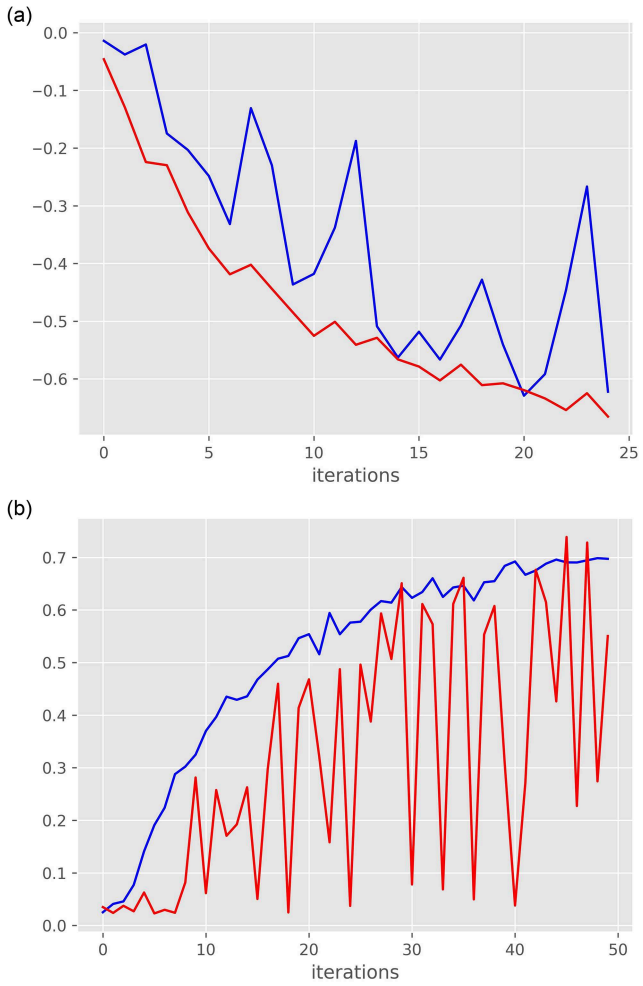
$$Dice\ Coeff = 2 * \frac{X \cap Y}{X + Y} \tag{14}$$

Both the Dice coefficient and the Jaccard coefficient measure the overlap between predicted and actual masks in image segmentation, but they differ in how they handle overlapping regions. The Dice coefficient assigns twice the weight to intersection regions, making it more sensitive to small overlaps, which is advantageous for tasks like medical image segmentation. In contrast, the Jaccard coefficient provides a more stringent measure of overlap, making it suitable for applications like object detection and general image segmentation.

4.4. Quantitative and qualitative analysis

A series of tests were conducted using the entire training data and various combinations of data across different folds to accurately assess the performance of the proposed approach. Figure 5(a) shows

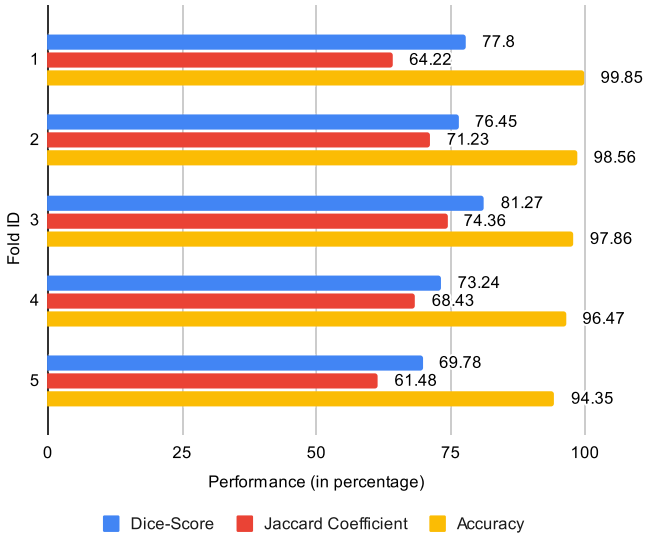
Figure 5
Training performance. (a) Validation loss curve and (b) validation IoU curve



the validation loss curve obtained during training with the full training set, while Figure 5(b) displays the IoU score achieved when training with the entire training set.

The prediction results of the proposed GRU-ASPP U-Net model for MS lesion segmentation are also outstanding and proved to be optimal in Fold 3 of the cross-validation that was performed, having 0.81 (81.27%) Dice score and 0.74 (74.36%) IoU. The cross-fold validation reveals a Dice score of 0.75 (75%) and IoU of 0.67 (67%) in total for all folds. This result can be attributed to the synergistic combination of three main architectural components: the mechanism of GRU blocks located at skip connections that allow to address temporal connections between MRI scans, the ASPP module enabling multi-scale feature extraction, and the SE blocks providing dynamic channel-wise recalibration. This architectural design is supported by effective data processing solutions which include ISBI 2015 appropriate preprocessing, CarveMix augmentation, removal of unnecessary black areas, normalization of images between 0 and 1 and resampling to 160 x 160 pixels. Even more, the training strategy enhances the model's predictive performance through cross-validation, combined loss function including both binary cross-entropy, Dice loss, optimized with Adam optimizer of choice and trained for 25 epoch. The best model from the different folds is selected based on the highest Dice score achieved and is then used for testing on the test data. The average evaluation

Figure 6
Performance metrics for the proposed model in different folds of validation data



metrics obtained from the cross-validation test folds are presented in Figure 6.

To dissection and refine the understanding of the model further, lesion-wise analysis, performance evaluation based on the size or area of the lesion. This could be explored further by visualizing feature maps to see which features were more active, looking at the attention weights and understanding the GRU state changed through time. Comparing our delineations with the radiologist annotations and validating on other datasets can be further beneficial to understand its operational feasibility setting from clinical point of view.

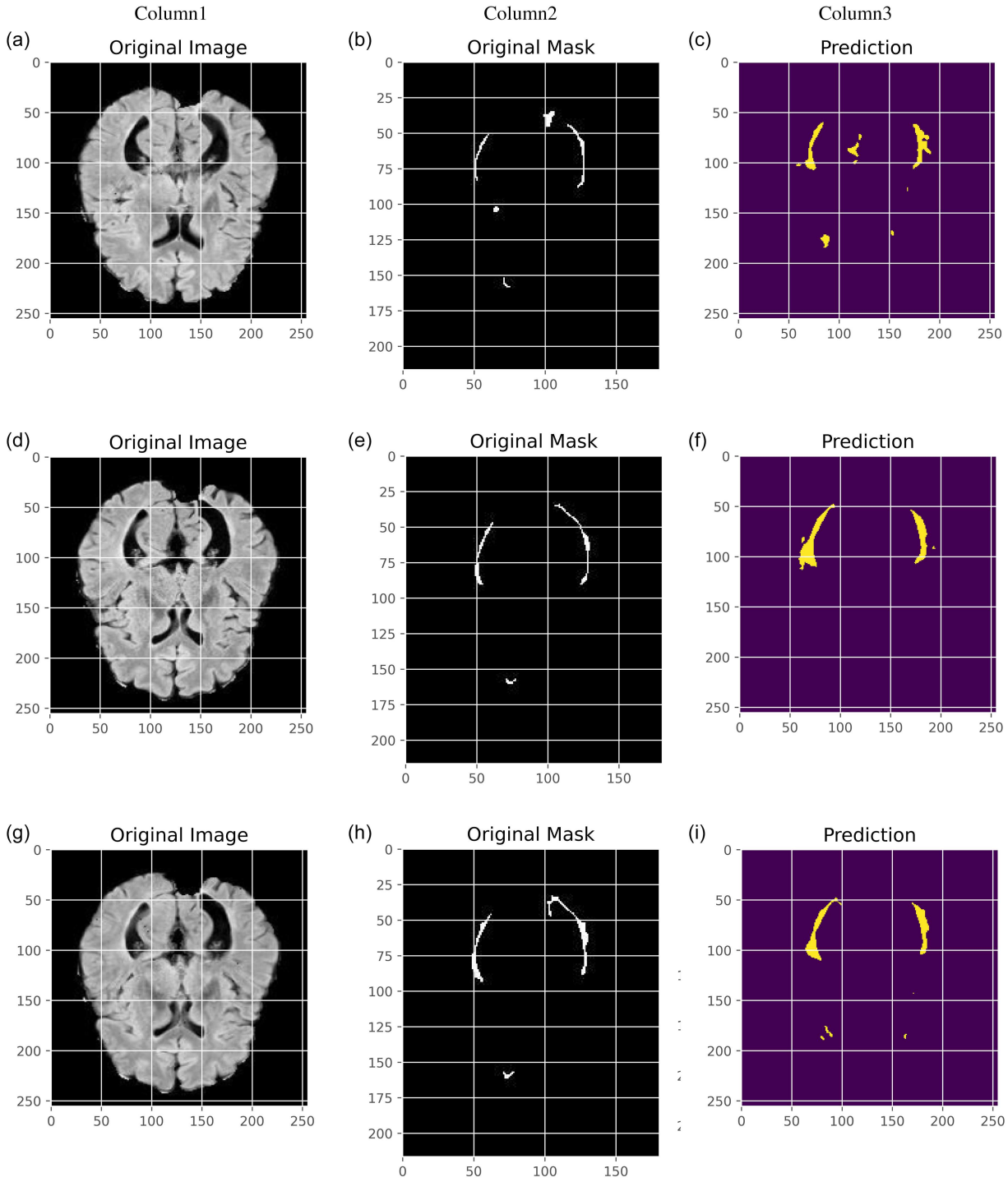
Figure 7 depicts the MS lesion segmentation of FLAIR images using proposed model. That implies, column 1 (a, d, g) shows the original slices of the FLAIR sequence, column 2 (b, e, h) displays the ground truth segmentations provided by rater 1, and column 3 (c, f, i) presents the lesion segmentations produced by the proposed approach. Table 2 gives a unique insight of different models and compare the perform in different categories for basic parameters and the result for this proposed model is quite impressive particularly when run under fold 3 cross-validation.

The feature maps produced by the bottleneck ASPP module, which consists of convolutional layers with different dilation rates in the proposed U-Net model, are analyzed for multicollinearity through the Variance Inflation Factor (VIF) [29]. VIF analysis is conducted to measure the dependency across feature maps, with those exhibiting a VIF value over the threshold (VIF>6) being redundant. These features are eliminated, guaranteeing the retention of only the most independent and informative attributes. The enhanced feature set is thereafter transmitted to the decoder path. Utilizing VIF analysis on the ASPP outputs and preserving only the most independent characteristics resulted in a significant enhancement in segmentation performance. The Intersection over Union (IoU) rose from 0.81 (baseline without VIF analysis) to 0.87, while the Jaccard Index enhanced from 0.74 to 0.79.

4.5. Observation and discussion

The proposed GRU-ASPP U-Net model reveals several notable merits even though the certain numerical indices look lower than

Figure 7
Results of the proposed model for lesion segmentation in FLAIR images



those of the desired approaches. One major advantage is that it performs equally well across all evaluation measures and visibly so in the fold-3 results where the TPR was roughly 0.83 coupled with an FPR of 0.17 for impressive lesion detection accuracy with low false positives and the needed segmentation precision of 0.89. More importantly, unlike many of the existing models which only present partial results, all six categories of performance measurements have been conducted on the proposed model to ensure thorough evaluation. This comprehensive assessment framework shows that the proposed model works well on different segmentation dimensions as well as their quality indicators.

Originally, our model's architecture makes a profound improvement in the aspect of temporal and spatial feature integration in comparison to the temporal-spatial network model, such as the U-Net with VGG16 and Enhanced U-Net with ResNet50 where the dominant aspect is the spatial feature. The incorporation of the desired components of the GRU model allows to overcome the problem of learning with increasing time/age data, which is required for monitoring the lesions in multiple sclerosis, the effectiveness of which is supported by a stable TPR/FPR ratio. From a clinical perspective, while U-Net++DSM shows a higher Dice score of 0.89, our model's balanced metrics (TPR: 0.83, FPR: 0.17) to achieve

Table 2
Comparative analysis of different U-Net variants with proposed GRU-ASPP U-Net model for MS lesion segmentation performance

	Evaluation metrics					
Models	**Dice score**	**IoU**	**TPR**	**FPR**	**DSC**	**Seg. accuracy**
U-Net++DSM [9]	0.89	0.74	-	-	0.89	0.98
U-Net with VGG-16 [6]	0.88	0.76	-	-	-	0.994
Enhanced U-Net with ResNet50 [7]	0.875	0.866	-	-	0.87	0.92
2D-UNet with random forest [25]	0.67	-	0.72	0.28	0.67	0.85
U-Net with AG, ECA, ASPP [26]	0.86	0.78	0.84	0.16	0.86	0.91
Transformer UNet++ with MobileNetv3 [12]	0.84	0.75	0.82	0.18	0.84	0.89
Modified Attention U-Net [14]	0.823	0.76	0.85	0.15	0.823	0.88
GRU-ASPP U-Net [Proposed] (mean score)	**0.78**	**0.69**	**0.78**	**0.22**	**0.75**	**0.87**
GRU-ASPP U-Net [Proposed] (fold-3 cross-validation)	**0.87**	**0.79**	**0.83**	**0.17**	**0.81**	**0.91**

significantly higher clinical relevance over lesion monitoring. The proposed model has the FPR of 0.17 which is lower than that of other models such as 2D-UNet with random forest which has the FPR of 0.28, which is more important in clinical decision-making is true positive/negative ratio.

The model achieves computational efficiency while maintaining competitive performance (Dice: 0.81, IoU: 0.74 in fold-3). It outperforms models that use VGG-16 or ResNet50 through more simplified architecture than those used in folds 1, 2 and mostly 3 with 81% accuracy and IoU of 0.74 in fold 3. Hence the integration of GRU, ASPP, SE as blocks gives a perfect balance between the model capacity and the performance and is ideal for clinical practice where resources in terms of computing power maybe very limited. In addition, our model possesses high reliability in folds of cross-validation, which indicates significant improvement from mean scores to the validation of fold 3 according to all the measurements. This improvement pattern shows that the model did learn from the data and was able to effectively and significantly generalize the learning when going from 0.75 Dice score to an 0.81 score.

However, there are certain limitations to consider. The model's efficacy may be skewed by the dataset's heterogeneity, constraining its applicability to scans from varying MRI protocols or demographic groups. GRU's dependence on sequential data may prove ineffective for irregular follow-up intervals or specific MRI sequences. Furthermore, the architecture is specifically designed for MS lesions and may encounter difficulties in generalizing to other lesion types or imaging modalities.

Thus, our GRU-ASPP U-Net provides better overall assessment, better balanced score, better real-world relevance, better temporal details dealing, better computational speed, and generalize ability through cross-validation. All these benefits taken together, make our model equally suited for practical clinical use where generalizability across the multiple modalities is more useful than optimizing for a single score. Our proposed U-Net accurately detects minor variations in lesion size and quantity over time, facilitating early diagnosis and the tracking of disease development. In areas with restricted availability of skilled professionals, automated segmentation can deliver uniform and high-quality lesion evaluations, enhancing access to sophisticated diagnostic tools. Through the incorporation of longitudinal data, our model offers insights into trends in illness progression, equipping doctors with data-driven tools for prediction and therapy planning.

5. Conclusion

In this research study, presented a novel GRU-ASPP U-Net architecture specifically designed for predicting MS lesion growth by segmenting MS lesion from longitudinal MRI sequences. Our approach introduces three key innovations: the integration of GRU blocks at skip connections for temporal dependency capture, an ASPP module for multi-scale feature extraction, and SE blocks for dynamic channel recalibration. The effectiveness of these architectural enhancements is demonstrated through comprehensive experimentation on the ISBI 2015 dataset. The proposed model achieved its best performance in fold-3 cross-validation with a Dice score of 0.81 and IoU of 0.74, significantly outperforming several state-of-the-art approaches. The mean performance across all folds demonstrated robust and consistent results with a Dice score of 0.75 and IoU of 0.67. These results validate the effectiveness of our architectural innovations, particularly the synergistic combination of GRU blocks at skip connections, ASPP module for multi-scale feature extraction, and SE blocks for dynamic channel recalibration. We have compared our method with other similar approaches, and the results of these experiments reveal that the designed architecture offers better segmentation performance and quality. Furthermore, in the future, Vision Transformer blocks combined with Conv GRU can be integrated, and a dynamic attention mechanism can be adapted to enhance feature learning for better lesion characterization. Additionally, the sustainability of lightweight versions of the architecture should be verified for use in resource-constrained environments.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution Statement

Ramya Palaniappan: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing, Visualization. **Siva Rathinavelayutham:** Validation, Data curation, Supervision, Project administration.

References

[1] Lin, F., Wu, Q., Liu, J., Wang, D., & Kong, X. (2021). Path aggregation U-Net model for brain tumor segmentation. *Multimedia Tools and Applications, 80*(15), 22951–22964. https://doi.org/10.1007/s11042-020-08795-9

[2] Yousef, R., Khan, S., Gupta, G., Siddiqui, T., Albahlal, B. M., Alajlan, S. A., & Haq, M. A. (2023). U-Net-based models towards optimal MR brain image segmentation. *Diagnostics, 13*(9), 1624. https://doi.org/10.3390/diagnostics13091624

[3] Wisaeng, K. (2023). U-Net++DSM: Improved U-Net++ for brain tumor segmentation with deep supervision mechanism. *IEEE Access, 11,* 132268–132285. https://doi.org/10.1109/ACCESS.2023.3331025

[4] Yang, T., Song, J., Li, L., & Tang, Q. (2020). Improving brain tumor segmentation on MRI based on the deep U-net and residual units. *Journal of X-Ray Science and Technology, 28*(1), 95–110. https://doi.org/10.3233/xst-190552

[5] Ghazal, M. T. (2023). A robust U-Net-based approach for accurate brain tumor segmentation using multimodal MRI data. *NTU Journal of Engineering and Technology, 2*(3), 27–33. https://doi.org/10.56286/ntujet.v2i3.692

[6] Ghosh, S., Chaki, A., & Santosh, K. C. (2021). Improved U-Net architecture with VGG-16 for brain tumor segmentation. *Physical and Engineering Sciences in Medicine, 44*(3), 703–712. https://doi.org/10.1007/s13246-021-01019-w

[7] Zhu, J., Zhang, R., & Zhang, H. (2024). An MRI brain tumor segmentation method based on improved U-Net. *Mathematical Biosciences and Engineering, 21*(1), 778–791. https://doi.org/10.3934/mbe.2024033

[8] Magadza, T., & Viriri, S. (2023). Efficient nnU-Net for brain tumor segmentation. *IEEE Access, 11,* 126386–126397. https://doi.org/10.1109/ACCESS.2023.3329517

[9] Ali, M., Gilani, S. O., Waris, A., Zafar, K., & Jamil, M. (2020). Brain tumour image segmentation using deep networks. *IEEE Access, 8,* 153589–153598. https://doi.org/10.1109/ACCESS.2020.3018160

[10] Micallef, N., Seychell, D., & Bajada, C. J. (2021). Exploring the U-Net++ model for automatic brain tumor segmentation. *IEEE Access, 9,* 125523–125539. https://doi.org/10.1109/ACCESS.2021.3111131

[11] Basaran, B. D., Matthews, P. M., & Bai, W. (2022). New lesion segmentation for multiple sclerosis brain images with imaging and lesion-aware augmentation. *Frontiers in Neuroscience, 16,* 1007453. https://doi.org/10.3389/fnins.2022.1007453

[12] Langat, G., Zou, B., Kui, X., & Njagi, K. (2024). An efficient multiple sclerosis segmentation framework using hybrid dilated convolution–based adaptive mobilenet mechanism. *Journal of Electrical Systems, 20,* 1960–1995. https://doi.org/10.52783/jes.1790

[13] Sadeghibakhi, M., Pourreza, H., & Mahyar, H. (2022). Multiple sclerosis lesions segmentation using attention-based CNNs in FLAIR images. *IEEE Journal of Translational Engineering in Health and Medicine, 10,* 1800411. https://doi.org/10.1109/JTEHM.2022.3172025

[14] Aishwarya, R., Ganesan, S., & Babu, T. R. (2023). MT-MRANBrainTumorNet: A novel brain tumour and severity level classification framework using meta-heuristic-assisted transformer based multiscale residual attention network. *Journal of Propulsion Technology, 44*(3), 3296–3321. https://doi.org/10.52783/tjjpt.v44.i3.1775

[15] Saeedi, S., Rezayi, S., Keshavarz, H., & Niakan Kalhori, S. R. (2023). MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques. *BMC Medical Informatics and Decision Making, 23*(1) 16. https://doi.org/10.1186/s12911-023-02114-6

[16] Das, S., Dey, N. S., & Mounika, M. (2023). Automated brain tumor segmentation in MRI: An enhanced mask generation approach. In *2023 7th International Conference on I-SMAC,* 891–897. https://doi.org/10.1109/I-SMAC58438.2023.10290265

[17] Aghalari, M., Aghagolzadeh, A., & Ezoji, M. (2021). Brain tumor image segmentation via asymmetric/symmetric UNet based on two-pathway-residual blocks. *Biomedical Signal Processing and Control, 69,* 102841. https://doi.org/10.1016/j.bspc.2021.102841

[18] Gessert, N., Krüger, J., Opfer, R., Ostwaldt, A.-C., Manogaran, P., Kitzler, H. H., ..., & Schlaefer, A. (2020). Multiple sclerosis lesion activity segmentation with attention-guided two-path CNNs. *Computerized Medical Imaging and Graphics, 84,* 101772. https://doi.org/10.1016/j.compmedimag.2020.101772

[19] Krüger, J., Opfer, R., Gessert, N., Ostwaldt, A.-C., Manogaran, P., Kitzler, H. H., ..., & Schippling, S. (2020). Fully automated longitudinal segmentation of new or enlarged multiple sclerosis lesions using 3D convolutional neural networks. *NeuroImage: Clinical, 28,* 102445. https://doi.org/10.1016/j.nicl.2020.102445

[20] Walsh, J., Othmani, A., Jain, M., & Dev, S. (2022). Using U-Net network for efficient brain tumor segmentation in MRI images. *Healthcare Analytics, 2,* 100098. https://doi.org/10.1016/j.health.2022.100098

[21] Yuan, Y., Li, Z., Tu, W., & Zhu, Y. (2023). Computed tomography image segmentation of irregular cerebral hemorrhage lesions based on improved U-Net. *Journal of Radiation Research and Applied Sciences, 16*(3), 100638. https://doi.org/10.1016/j.jrras.2023.100638

[22] Sarica, B., Seker, D. Z., & Bayram, B. (2023). A dense residual U-net for multiple sclerosis lesions segmentation from multi-sequence 3D MR images. *International Journal of Medical Informatics, 170,* 104965. https://doi.org/10.1016/j.ijmedinf.2022.104965

[23] Alruily, M., Said, W., Mostafa, A. M., Ezz, M., & Elmezain, M. (2023). Breast ultrasound images augmentation and segmentation using GAN with identity block and modified U-Net 3+. *Sensors, 23*(20), 8599. https://doi.org/10.3390/s23208599

[24] Chen, Y., Wang, L., Ding, B., Huang, Y., Wen, T., & Huang, J. (2023). Radiologically based automated segmentation of cardiac MRI using an improved U-Net neural algorithm. *Journal of Radiation Research and Applied Sciences, 16*(4), 100704. https://doi.org/10.1016/j.jrras.2023.100704

[25] Hashemi, M., Akhbari, M., & Jutten, C. (2022). Delve into Multiple Sclerosis (MS) lesion exploration: A modified attention U-Net for MS lesion segmentation in brain MRI. *Computers in Biology and Medicine, 145,* 105402. https://doi.org/10.1016/j.compbiomed.2022.105402

[26] Carass, A., Roy, S., Jog, A., Cuzzocreo, J. L., Magrath, E., Gherman, A., ..., & Pham, D. L. (2017). Longitudinal multiple sclerosis lesion segmentation: Resource and challenge. *NeuroImage, 148,* 77–102. https://doi.org/10.1016/j.neuroimage.2016.12.064

[27] Zhang, X., Liu, C., Ou, N., Zeng, X., Xiong, X., Yu, Y., ..., & Ye, C. (2021). CarveMix: A simple data augmentation method for brain lesion segmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021: 24th International Conference,* 196–205. https://doi.org/10.1007/978-3-030-87193-2_19

[28] Saboor, A., Li, J. P., Ul Haq, A., Shehzad, U., Khan, S., Aotaibi, R. M., & Alajlan, S. A. (2024). DDFC: Deep learning approach for deep feature extraction and classification of brain tumors using magnetic resonance imaging in E-healthcare system. *Scientific Reports, 14*(1), 6425. https://doi.org/10.1038/s41598-024-56983-6

[29] Kianfar, N., Mesgari, M. S., Mollalo, A., & Kaveh, M. (2022). Spatio-temporal modeling of COVID-19 prevalence and mortality using artificial neural network algorithms. *Spatial and Spatio-temporal Epidemiology*, *40*, 100471. https://doi.org/10.1016/j.sste.2021.100471

How to Cite: Palaniappan, R., & Rathinavelayutham, S. (2026). GRU-ASPP U-Net: Enhanced Multiple Sclerosis Lesion Segmentation Through Temporal-Spatial Feature Integration in Longitudinal MRI. *Journal of Computational and Cognitive Engineering*, *5*(2), 341–353. https://doi.org/10.47852/bonviewJCCE52024394