IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 22 September 2024, accepted 11 October 2024, date of publication 18 October 2024, date of current version 4 November 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3483562

RESEARCH ARTICLE

A Novel ConvLSTM-Based U-net for Improved Brain Tumor Segmentation

OSAMA MAJEED HILAL ALMIAHI[1], ALAA TAIMA ALBU-SALIH[1], AND DHAFER ALHAJIM[2]
[1]College of Computer Science and Information Technology, University of Al-Qadisiyah, Al Diwaniyah 58002, Iraq
[2]Computer Center, University of Al-Qadisiyah, Al Diwaniyah 58002, Iraq

Corresponding author: Dhafer Alhajim (dhafer.alhajim@qu.edu.iq)

This work was supported in part by the University of Al-Qadisiyah.

ABSTRACT Using 2D scans or simple 3D convolutions are two limitations of previous works on segmentation of brain tumors by deep learning, which lead to ignoring the temporal distribution of the scans. This study proposes a novel extension to the well-known U-net model for brain tumor segmentation, utilizing 3D Magnetic Resonance Imaging (MRI) volumes as inputs. The method, called ConvLSTM-based U-net + up skip connections, incorporates the ConvLSTM blocks to capture spatio-temporal dependencies in the 3D MRI volumes, and up skip connections to capture low-level feature maps extracted from the encoding path, enhancing the information flow through the network to the standard U-net architecture. A novel intensity normalization technique is used to improve the comparability of scans. This technique normalizes image intensity by subtracting the grey-value of the most frequent bin from the image. The novel method is tested on the Multimodal Brain Tumor Segmentation (BRATS) 2015 dataset, showing that the use of ConvLSTM blocks improved segmentation quality by 1.6% on the test subset. The addition of skip connections further improved performance by 3.3% and 1.7% relative to the U-net and ConvLSTM-based U-net models, respectively. Moreover, the inclusion of up skip connections could enhance the performance by 5.7%, 3.99% and 2.2% relative to the simple U-net, ConvLSTM-based U-net, and ConvLSTM-based U-net with skip connections, respectively. Finally, the novel preprocessing technique had a positive effect on the proposed network, resulting in a 3.3% increase in the segmentation outcomes.

INDEX TERMS Brain tumor, deep learning, ConvLSTM, up skip connection, U-net.

I. INTRODUCTION

Brain tumors are among the deadliest forms of cancer, with glioma being the most prevalent type [1]. From the pathological perspective, gliomas are grouped into high-grade and low-grade gliomas (HGGs and LGGs) [2], [3]. Since the survival period for glioblastoma patients is typically less than fourteen months, early diagnosis of glioma is vital to have a better treatment, surgery, and follow-up evaluations [4]. MRI is the most reliable tool to diagnose brain tumors, which is widely used by radiologists [5]. Nonetheless, due to the complexity of brain tumors, manual diagnosis from MRI images requires significant expertise and time [6]. As a result, there is a growing need for automatic and robust segmentation techniques.

The associate editor coordinating the review of this manuscript and approving it for publication was Larbi Boubchir.

The expanding capabilities of computer hardware and advancements in data science algorithms have made automated computer methods a valuable resource for interpreting medical images. Deep learning, as a branch of machine learning, has shown great promise in solving various complex problems in various fields [7], [8], [9], [10]. Deep learning methods offer automatic feature extraction, compatibility with unstructured data, cost-effectiveness, scalability for large-scale applications, and stability, making them one of the most promising choices for automatic segmentation of medical images. These methods have shown successful applications in medical image analysis [9], [11], [12], [13], paving the way to increase the speed and accuracy of clinical trials for pateients. Deep learning-based techniques, like Fully Convolutional Neural Networks (FCN) [14] and Convolutional Neural Networks (CNN) [15], are two of the most used methods for the automatic extraction of salient features from

inputs to solve underlying problems. Recently, the application of FCNs and CNNs in brain tumor segmentation has gained considerable momentum [16], [17], [18], [19], [20], [21], [22], [23], [24], [25]. Dvořák and Menze [26] approached the multiclass brain tumor segmentation by treating at as binary subtasks and employed three CNNs to tackle the subtasks. Pereira et al. [22] utilized deep CNNs with small kernels to perform automatic brain tumor segmentation using two-dimensional input patches. Havaei et al. [21] introduced a 2D two-pathway deep CNN to capture features at multiple scales, combined with a cascade structure where each CNN processed the probability outputs of the previous CNNs. The limitation of their method was that it used 2D patches, which ignored 3D dependencies. DeepMedic model was proposed by Kamnitsas et al. [20], which took benefit from 3D convolutions to extract three-dimensional contextual information from patches.

Despite the satisfactory performance of the aforementioned methods, most rely on pointwise convolutions, which means they adopt a voxel-wise classification approach. Pointwise segmentation suffers from slow inference [27]. To alleviate this limitation, FCNs [14] were introduced to label the entire samples automatically. Ronneberger et al. [16] devised an innovative extension to FCNs, named U-net, designed for medical images segmentation. An advantage of the U-net is its compatibility with a small number of training samples, and its ability to simultaneously extract global location and context information. Beers et al. [28] introduced a sequential U-net for 3D glioma segmentation. Similarly, Dong et al. [29] proposed an extension to the U-net, which was a two-dimensional fully automatic network. Chen et al. [25] proposed the VoxResNet, a voxel-wise residual network, to label voxels in 3D input patches. Walsh et al. [30] devised a lightweight U-net for real-time brain tumor segmentation, utilizing three perspective planes of MRI images, instead of the original 3D volumes. Lei et al. [31] enhanced the accuracy of brain and liver tumor segmentation by superposing multi-scale Atrous convolutions. Zhou et al. [32] invented a method to handle missing modalities in MRI images by learning the implicit relationship between the multiple modalities. Zhang et al. [33] adopted conventional 3D convolutions and the U-net architecture for multi-modality feature learning in brain tumor segmentation. Their model included two branches for learning features from two different modalities and another branch for fusing the learned features from the first two branches. Ranjbarzadeh et al. [34] devised a Cascade CNN to extract both local and global features through two distinct routes. They also proposed a novel Distance-Wise Attention mechanism to locate the center of the tumor in the brain. Archana et al. [35] adopted the U-net architecture for brain tumor segmentation and employed Bagging Ensemble K-Nearest Neighbors to detect tumor malignancy. Their model conistsed of two parts: one for segementation using a standard U-net, and another for classification. Nodirov et al. [36] utilized a 3D U-net with an attention module to increase both the speed and accuracy of brain tumor segmentation based on MRI images. However, their results showed missing label parts and lower segmentation quality because of using MobileNetV2 as the backbone of the model, which has a complex structure. More recently, Sowrirajan et al. [37] combined Long Short-Term Memory (LSTM) with the U-net architecture to segment brain tumors based on different MRI pulses. In their model, 3D MRI images were processed by 3D convolution operations in the encidng path, while feature maps from the transposed convolutions were passed to simple LSTM units to capture the temporal dependencies between frames and slices. Xu et al. [38] also proposed an extension to the U-net architecture by incorporating ConvLSTM layers. The proposed model consisted of two parts: the first part had a U-net structure with only 3D convolution layers that received different MRI slices with multiple modalities and produced the segmentation masks for each slice. Then, the masks were passed to ConvLSTM layers to capture the dependencies between different slices. However, their model did not use ConvLSTM layers to directly processes MRI images. Ranjbarzadeh et al. [39] proposed an optimized CNN, trained using the Improved Chimp Optimization Algorithm (IChOA). The process involved normalizing MRI sequences, selecting key features with a Support Vector Machine (SVM), and feeding these into the CNN for segmentation. Wang et al. [40] incorporated 3D dilated convolutions and an attention guidance block to the standard U-net model for automatic brain tumor segmentation from MRI images. This aimed to extract rich feature information and eliminate irrelevant region interference in global information. Other studies on brain tumor segmentation can be found in [41], [42], [43], and [44]. According to the cited liteeature, numerous studies on brain tumor segmentation from MRI sequences have leveraged the U-net architecture with various modifications. However, most of them have utilized conventional 3D convolutions to process 3D MRI volumes, and only a few studies have incorporated ConvLSTM and LSTM layers into their networks. Notably, the models with LSTM or ConvLSTM layers have applied these layers in the decoding path to capture the temporal dependencies between the feature maps or in the segmentations maks across different timesteps (slices) processed by conventional 3D convolutions. As a result, a model that directly processes 3D MRI volumes using ConvLSTM layers to capture temporal depenceneis between slices is still lacking. Additionally, most models in the literature have utilized the mean-calculated normalization preprocessing technique, which fails to address the intensity variations across MRI images obtained under consistent scanner, patient, and tissue conditions.

In this study, an innovative extension to the U-net is used for the segmentation of 3D brain MRI volumes. The vanilla U-net architecture is modified as follows: ConvLSTM [45] blocks with up skip connections are incorporated into the standard U-net architecture, replacing the conventional 3D convolutions. It is expected to have a better information

flow between the encoding and decoding branches of the network by addition of up skip connections since they ensure utilizing low-level features to enhance the network's outputs. ConvLSTM layers are included due to the temporal nature of 3D MRI images. These layers are responsible for capturing the dependencies along the third dimension. The main difference between this study and previous works involving ConvLSTM or LSTM is that this study uses a single-part model where ConvLSTM layers directly receive MRI volumes and process them through both the encoding and decoding paths, capturing the temporal dependencies between MRI slices required for accurate segmentation of brain tumors. Another distinction of this study is the use of a preprocessing technique different from conventional mean-variance scaling. This preprocessing technique is based on subtracting images' mode from each image to normalize their intensity. The proposed deep network is tested on the public dataset from the Multimodal Brain Tumor Image Segmentation Challenge (BRATS) 2015. Since previous studies did not employ ConvLSTM layers to directly process MRI volumes, and since accurately delineating the boundaries of tumors in MRI sequences depends on the temporal dependencies between slices, this motivated us to modify the standard U-net structure by incorporating ConvLSTM layers into both the encoding and decoding paths. The novelty of this work can be summarized as:

- Incorporating ConvLSTM layers into the encoding and decoding paths of the U-net model to capture the temporal dependencies in different MRI slices.
- Enahncing the information flow through the network by adding up skip connections alongside conventional skip connections.
- Enhancing the comparibility of MRI images using a robust intensity normalization technique.

The structure of the paper is as follows: Section I introduces the field of atumatic brain tumor segmentation using deep learning, highlighting its importance and summarizing recent literature in brain tumor segmentation and its limitations. Section II provides a detailed explanation of the tools and methods, including the utilized dataset, preprocessing technique, the proposed ConvLSTM U-net model with up skip connections, and the evaluation metrics. Section III presents the results obtained by the proposed model in comparision to other state-of-the-art models. Finally, Section IV, draws conclusions.

II. MATERIAL AND METHODS

A. DATASET

In this paper, the Multimodal Brain Tumor Image Segmentation Challenge (BRATS) 2015 dataset [46], [47] was used to test the proposed segmentation network. The BRATS 2015 dataset includes 220 HGGs and 54 LGGs in the training subset. The test subset consists of 110 gliomas with unknown grades. There were four separate MRI pulses, including T1 (T1-weighted), T1c (T1 contrast-enhanced), T2 (T2-weighted), and Flair, for every patient. Each modality consists of 155 slices, resulting in 620 slices per patient. All images were skull-striped. Figure 1 illustrates an instance of the 100th slice across all four modalities from the training subset. There are totally five labels, including necrosis (with label 1), edema (with label 2), non-enhancing tumor (with label 3), enhancing tumor (with label 4), and anything else (with label 0).

B. PREPROCESSING

Since MRI images exhibit varied intensity ranges and are affected differently by bias fields, this study proposes a robust intensity normalization technique to enhance image comparability as well as to modify bias field variation of the scans using the N4ITK approach [48]. Since MRI images obtained by the same equipment, from the same patient, and with the same conditions can have varied intensities, conventional mean-standard-deviation methods are not suitable for preprocessing MRI images. This is because applying the mean-standard-deviation method can result in similar scans appearing different because of these intensity variation artifacts. The innovative normalization technique proposed in this study relies on subtracting the grey-value of the most frequent bin from the image to normalize its intensity. Moreover, the standard deviation is scaled to 1. Since the white-matter constitutes 50% of the brain [49], the most frequent bin is equivalent to the white-matter. Consequently, the comparability of the MRI scans can be improved by matching the intensities of the white matter across MRI images and normalizing the overall distribution of intensities. The conventional intensity-mean-based standard deviation may have no fixed tissue concept. Thereby, robust intensity deviation is employed in this study. Robust intensity deviation is calculated according to the most frequent bin's grey-value, representing the distinction of intensity to the white matter. Considering an MRI volume with voxels $\{v_1, \ldots v_N\}$, in which the intensity of each voxel is $I_t, t = 1, 2, \ldots, T$, robust deviation is given as:

$$\tilde{\sigma} = \sqrt{\sum_{t=1}^{T} (\hat{I} - I_t)^2 / T},$$

where $\hat{I}$ represents the most frequent bin's grey-value. The overall process of intensity normalization is as follows:

1) Scale the intensity values between 0 – 255,
2) Compute the histogram of intensities using 256 bins,
3) Subtract the most frequent bin's grey-value, $\hat{I}$, from the image and multiply it by the inverted robust deviation,
4) Multiply every voxel's intensity by $\tilde{\sigma}$ and add a constant (I_0). Then, clip the intensity values between 0 and 255. Here, I_0 and $\tilde{\sigma}$ are considered equal to the grey-value of the most frequent bin and the robust deviation of the first image in the training subset, which is preprocessed by N4ITK algorithm and Step 1. For T1c, T2, and Flair slices, $I_0 = 99, 55, 75$ and $\tilde{\sigma} = 31, 37, 30$.

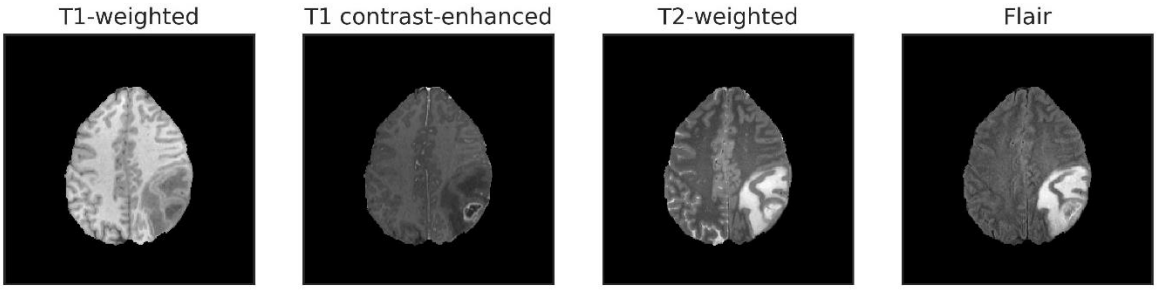

FIGURE 1. An example from the training subset showing the 100th slice of four pulses.

C. METHODS

In this study, an end-to-end deep network is proposed for improved brain tumor segmentation, named ConvLSTM-based U-net. The proposed network is an extension of the standard U-net architecture, in which ConvLSTM blocks and up skip connections are incorporated into the U-net (see Figure 2). The proposed architecture performs the brain tumor segmentation in a slice-by-slice manner consuming four MRI sequences (T1, T1c, T2, and Flair).

1) ConvLSTM

ConvLSTM is a processing unit which is responsible for extracting spatio-temporal dependencies. ConvLSTM has had successful applications in various problems, such as human action recognition [45], [50], medical image segmentation [51], [52], etc. Incorporating the convolution operation into Long Short-Term Memory (LSTM) units creates the ConvLSTM units. In other words, the simple element-wise product between the inputs and weights is changed to convolution operations, enabling it to process 2D and 3D data (e.g., MRI images). For more information on standard LSTM, please refer to [15], [53], and [54]. The integration of the convolution and LSTM operations can extract spatio-temporal dependencies. The structure of a ConvLSTM cell is similar to the structure of a standard LSTM cell (see Figure 3). The key component of a ConvLSTM cell is called the cell state, $\mathcal{C}_t$, which is responsible for conveying the extracted information from one timestep to another. Moreover, three gates, including the forget gate, input gate, and output gate control the addition and subtraction of information from one cell to another. The forget gate, f_t, decides what information should be removed from the cell state. If the forget gate outputs 0, the cell state discards the collected information, whereas, if the output grows to 1, all information is retained. The input gate, i_t, decides what new information to add to the cell state. After updating the cell state, it is combined with the output gate, o_t, to compute the hidden state, $\mathcal{H}_t$, which is passed to the next ConvLSTM cell. The hidden state contains the extracted features for each timestep. Figure 3 shows the detailed structure of a ConvLSTM cell. The aforementioned operations are given in Equations 1 through 6.

$$i_t = \sigma\left(W_{xi} \circledast X_t + W_{hi} \circledast \mathcal{H}_{t-1} + b_i\right) \quad (1)$$

$$f_t = \sigma\left(W_{xf} \circledast X_t + W_{hf} \circledast \mathcal{H}_{t-1} + b_f\right) \quad (2)$$

$$\tilde{\mathcal{C}}_t = tanh\left(W_{xC} \circledast X_t + W_{hC} \circledast \mathcal{H}_{t-1} + b_C\right) \quad (3)$$

$$\mathcal{C}_t = f_t \odot \mathcal{C}_{t-1} + i_t \odot \tilde{\mathcal{C}}_t \quad (4)$$

$$o_t = \sigma\left(W_{xo} \circledast X_t + W_{ho} \circledast \mathcal{H}_{t-1} + b_o\right) \quad (5)$$

$$\mathcal{H}_t = o_t \odot tanh\left(C_t\right) \quad (6)$$

where $\mathcal{C}_{t-1}$, $\mathcal{C}_t$, $\mathcal{H}_{t-1}$, and H_t, are the cell states and hidden states at timesteps $t-1$ and t, respectively, x_t is the input at timestep t, W_{**} is the convolution kernel between two states, $\odot$ denotes the element-wise product, and $\circledast$ is the convolution operation. A ConvLSTM layers is composed of several ConvLSTM units, while a ConvLSTM block is constructed from two layers containing multiple ConvLSTM cells (see Figure 4).

2) UP SKIP CONNECTIONS

As shown in Figure 2, up skip connections are added between the encoding and decoding paths to enhance the network's connectivity by facilitating the learning of multi-level features. There are two advantages to adding up skip connections. First, up skip connections connect the downsampling and upsampling parts, allowing the decoding layers to retain low-level information extracted by the shallower layers of the encoding path. Consequently, the network recovers the spatial information lost during the encoding process. Second, the gradient flow during backpropagation is improved by reducing the semantic gap between the low-level features from shallower layers and the high-level features from deeper layers, leading to more effective training of the shallow layers. As Figure 2 illustrates, if there are N layers in the network, the standard skip connections integrate the output of layer i and layer $N-i+1$, whereas the proposed up skip connections integrate layers i, $N-i+1$, and $N-i+2$. Figure 5 illustrates that the up skip connections in the upsampling layers follow a weighted strategy. To ensure the feature maps are of the same size during the weighted addition process, an identical kernel size (3×3) is assigned both in the up skip connection layers and transposed convolution (upsampling) layers, with strides of 2×2. In addition, a 1×1 convolution kernel with a stride of 1×1 is used to preserve the balance in feature size. Figure 5 illustrates the weighted addition which

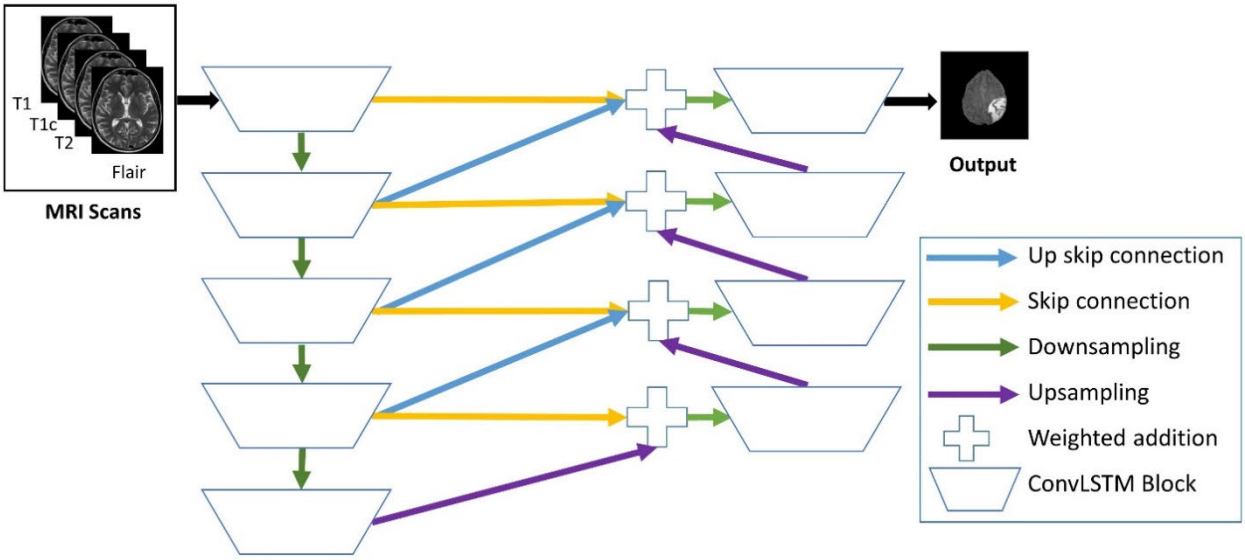

FIGURE 2. Structure of the proposed ConvLSTM-based U-net.

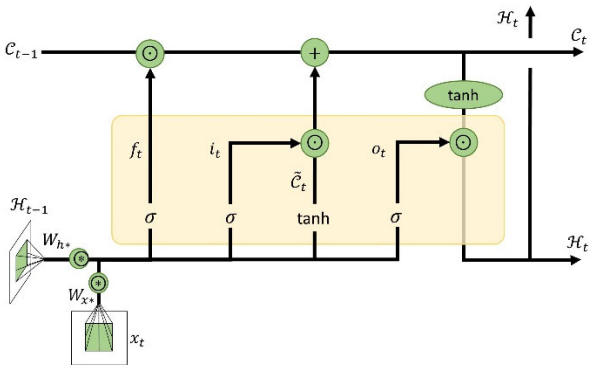

FIGURE 3. Structure and operations of a ConvLSTM cell.

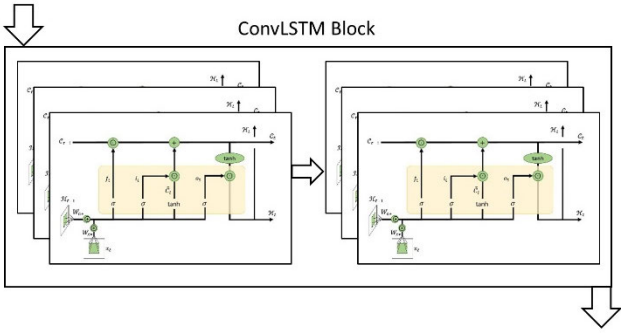

FIGURE 4. Schematic of a ConvLSTM block.

is expressed as:

$$x_0 = \alpha_1 . x_{skip} + \alpha_2 . x_{upskip} + \alpha_3 . x_{upsampling} \quad (7)$$

where x_{skip}, x_{upskip}, and $x_{upsampling}$ denote the conventional skip connections, up skip connections, and the upsampling operations, respectively. In the end, α_1, α_2, and α_3 are the coefficients of x_{skip}, x_{upskip}, and $x_{upsampling}$, respectively, which are learned during the network training process.

We opted for a weighted sum instead of the simple concatenation because the weighted sum allows for a more intelligent combination of the features by emphasizing certain components more than others based on their relevance to the task. In other words, with the weighted sum strategy, the model learns the importance of feature maps coming from different sources (skip connection, up skip connection, upsampling). On the other hand, concatenation would increase the dimensionality, which could introduce redundancy or unnecessary complexity. The weighted sum, however, provides a more compact and interpretable representation while preserving essential information.

TABLE 1. Libraries used to implement the proposed model in this study.

Library	Version	Link
Tensorflow	2.12.0	https://tensorflow.org
Keras	2.12.0	https://keras.io/
NumPy	1.24	https://numpy.org/
Matplotlib	3.6.2	https://matplotlib.org
Scikit-learn	1.2	https://scikit-learn.org/

D. EVALUATION METRICS

In this paper, four evaluation metrics are utilized to assess the performance of the proposed brain tumor segmentation model. These metrics encompass sensitivity, positive predictive value (PPV), Dice score, and Jaccard similarity [49].

The Dice score determines the similarity between the ground truth (T) and predicted (P) masks. The Dice score

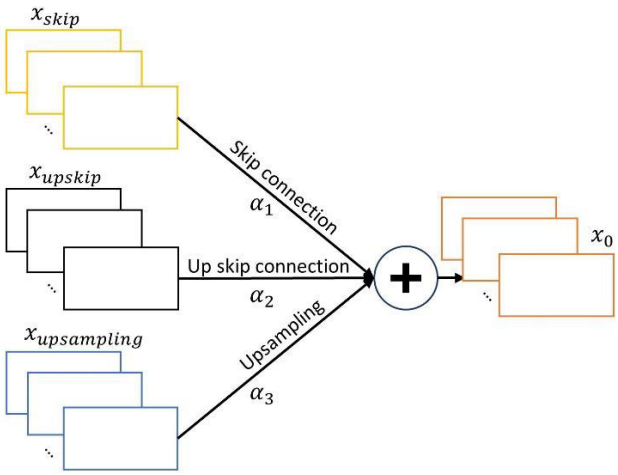

FIGURE 5. Detailed structure of the weighted addition.

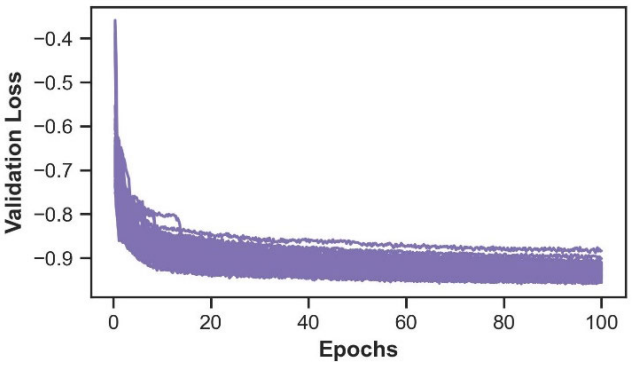

FIGURE 6. Validation loss versus epochs during the grid search.

ranges from 0 to 1, with values closer to 1 indicating a high similarity. Assuming that the predicted segmentation map is $P \in (0, 1)$ and the corresponding ground truth is $T \in (0, 1)$, which are binary maps, the Dice score can be formulated as:

$$Dice\,(P, T) = \frac{2 \times |P_1 \cap T_1|}{(|P_1| + |T_1|)} \tag{8}$$

where $|P_1 \cap T_1|$ is the intersection of both sets, and $|P_1|$ and $|T_1|$ denote the number of voxels where $P_1 = 1$ and $T_1 = 1$, respectively.

The Jaccard similarity indicates the proportion of the correct predictions and is a widely used evaluation metric in medical image segmentation [55]:

$$Jaccard\,(P, T) = \frac{|P_1 \cap T_1|}{|P_1 \cup T_1|} \tag{9}$$

The PPV and Sensitivity measure the accuracy of the network in correctly segmenting positive voxels. The PPV is calculated as the portion of correctly classified positives, while Sensitivity is the ratio of true positives to all ground truth positives. Equation 10 and 11 provide the mathematical formulation of PPV and Sensitivity, respectively.

$$PPV\,(P, T) = \frac{|P_1 \cap T_1|}{|T_1|} \tag{10}$$

$$Sensitivity\,(P, T) = \frac{|P_1 \cap T_1|}{|T_1|} \tag{11}$$

All codes are implemented in Python using open-source machine learning, deep learning, and data science libraries as listed in Table 1.

III. RESULTS AND DISCUSSION

The performance of a deep neural network is highly affected by the selection of its hyperparameters, including the number of hidden layers, the number of kernels, kernel sizes, activation functions, optimizer, etc. In this study, a grid search with 5-fold cross-validation was used to find the optimal hyperparameters for the proposed network. Table 2 presents the optimal hyperparameters of the network and their investigation ranges. The negative Dice score was used as the loss function. In addition, early stopping was employed to avoid overfitting, with training halted if the validation loss remained unchanged for 20 consecutive epochs. Hyperparameters that did not have an investigation range were not changed during the grid search.

The cross-validation losses against epochs during the grid search are shown in Figure 6. To select the optimal hyperparameters, those resulting in the lowest cross-validation loss were selected. The ultimate cross-validation loss was calculated as the average validation loss over the 5 folds.

Before the grid search, all images underwent intensity normalizing. Figure 7 illustrates the impact of intensity normalization on three random samples of T2 scans. It is clear that the proposed normalization technique has improved the comparability of the MRI scans. The effect of intensity normalization is also examined through the intensity histogram of 20 randomly chosen T2 scans, as shown in Figure 8, revealing that high intensities are normalized to lower values.

Three cases were considered for training the proposed network: without any skip connections, with skip connections, and with skip and up skip connections. The results from these three cases were compared with each other and with the base (standard U-net) model.

The convergence curves for the three training scenarios are shown in Figure 10. As depicted, the ConvLSTM-based U-net + up skip connections achieved the lowest train and validation losses. Although the training loss of the ConvLSTM-based U-net is lower than that of the ConvLSTM-based U-net + skip connections, it exhibits a higher validation loss than the ConvLSTM-based U-net + skip connections. It is worth pointing out that the U-net was not trained in this study; its results were obtained from the literature. Figure 10 demonstrates the visual segmentation outputs of the proposed network and the standard U-net on a Flair sample from the test subset. In this figure, the yellow region indicates necrosis and non-enhancing tumor, the

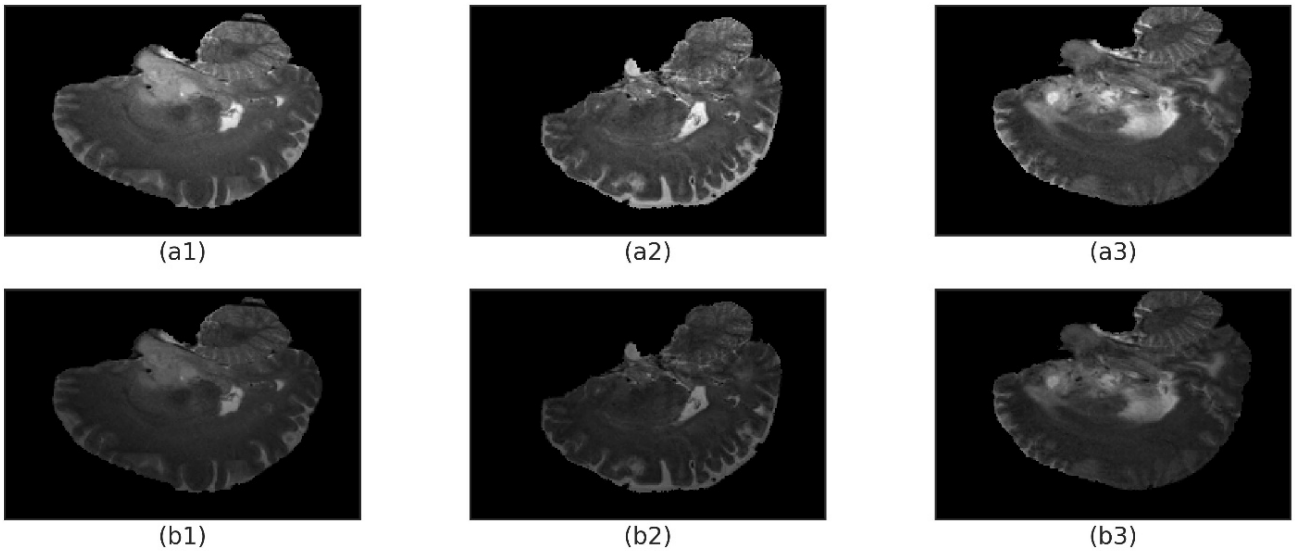

FIGURE 7. Randomly selected T2 scans before (a1-a3) and after (b1-b3) intensity normalization.

red region indicates enhancing tumor, and the blue region indicates edema. As shown, the ConvLSTM-based U-net + up skip connections has resulted in the best segmentation mask. Moreover, the ConvLSTM-based U-net + skip connection yielded a better segmentation mask compared to the ConvLSTM-based U-net without skip connections, highlighting the importance of skip connections. As stated earlier, the reason behind the better performance of the network with skip and up skip connections is enahnced information flow and information integration during the feed forward and backpropagation processes. This enhancement facilitated the recovery of lost information in the encoding path and resulted in more effective training of the network. Overall, the proposed methods outstandingly outperformed the standard U-net. Table 3 reports the average and standard deviation of the evaluation metrics for the utilized networks, state-of-the-art models, and the standard U-net on 110 test cases.

Based on Table 3, the proposed segmentation neural network has outperformed other networks across all performance metrics. In this regard, the standard U-net model yielded the lowest evaluation metrics. The inclusion of LSTM layers significantly improved the performance of the network; for example, the LSTM Multi-modal U-net reached a Dice score of 0.873, which is 0.009 higher than the dice score of the standard U-net. Other evaluations metrics, including sensitivity, PPV, and Jaccard similarity, also demonstrated improvement with the addition of LSTM and ConvLSTM layers. Among the literature, the LSTM Multi-modal U-net showed the best performance, since it utilized ConvLSTM layers to capture the dependencies between the segmentation masks obtained from different MRI slices. This indicates that ConvLSTM layers are appropriate tools for processing 3D MRI volumes, which are time-lapses of 2D images. However, the performance of these methods from the literature is generally similar, since none employed ConvLSTM layers in the encoding path of the network to process MRI volumes. Moreover, the use of skip connections and up skip connections could improve upon the results of the ConvLSTM-based U-net, supporting the hypothesis that the these connections help recover the lost information in the encoding path of the network, leading to enhanced information flow in the feed forward pass and improved gradient flow during backpropagation, ultimately resulting in better learning of shallower layers.

As Table 3 shows, the addition of ConvLSTM blocks has increased the mean Dice score by 1.6%, sensitivity by 1%, PPV by 2.8%, and Jaccard similarity by 1.1% compared to the standard U-net. In the meanwhile, the improvement of these evaluation metrics when integrating skip connections and ConvLSTM blocks is 3.3%, 2.7%, 6.6%, and 2.3% compared to the standard U-net, respectively. In contrast, the improvements after adding only the ConvLSTM blocks were 1.7%, 1.7%, 3.6%, and 1.1%, respectively. Finally, the inclusion of up skip connections, while keeping the skip connections and ConvLSTM blocks, improved the evaluation metrics by 5.7%, 5%, 10.1%, and 4.5% compared to the standard U-net; 3.99%, 4.1%, 3.6%, and 3.3% compared to the ConvLSTM-based U-net; and 2.2%, 2.3%, 3.3%, and 2.2% compared to the ConvLSTM-based U-net + skip connections, respectively.

To facilitate a straightforward comparison, the improvements are summarized in Table 4. As can be deduced from Table 4, the PPV showed the highest increase in each stage, which is a vital metric, since it is a measure showing that the existence of brain tumor and its type is being detected precisely. In addition to the overall improvement in the

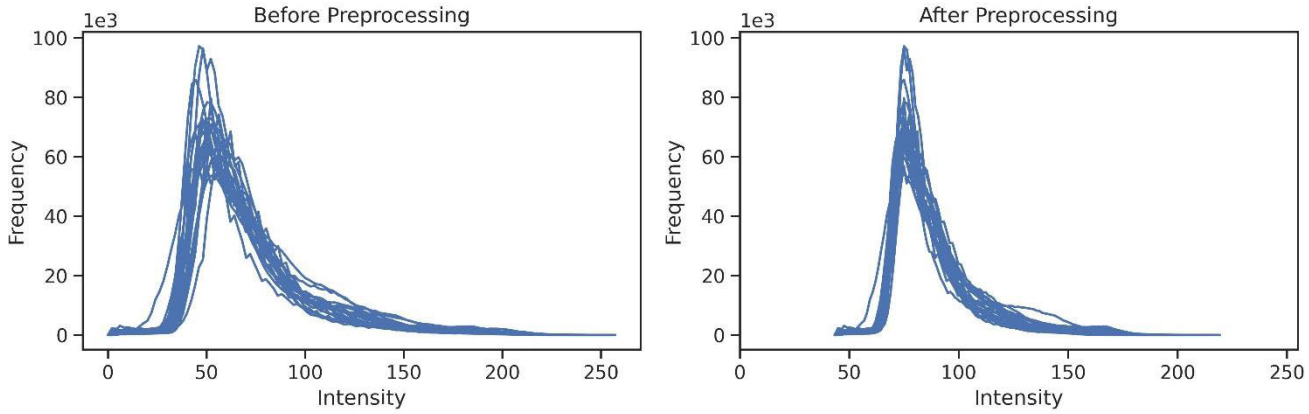

FIGURE 8. Histogram of the image intensity of 20 samples of T2 slices before and after preprocessing by the proposed intensity normalization method.

TABLE 2. Optimal hyperparameters for the proposed network obtained by 5-fold cross-validation.

Hyperparameter	Optimal value	Investigation ranges
No. of ConvLSTM blocks	9	3, 5, 7, 11
No. of layers in each block	2	1,2,3
No. of ConvLSTM kernels	16,32,64,128,256,128,64,32,16,1	8 – 1024
Activation functions	ReLU	ReLU, ELU, sigmoid, tanh
Optimizer	Adam	Adam, RMSProp, SGD
Learning rate	0.01	0.01, 0.005, 0.001
Batch size	64	16,32,64,128
Epochs	100	-
Early stopping	20	-
Loss function	$-\frac{2\times\lvert P_1\cap T_1\rvert}{(\lvert P_1\rvert+\lvert T_1\rvert)}$	-

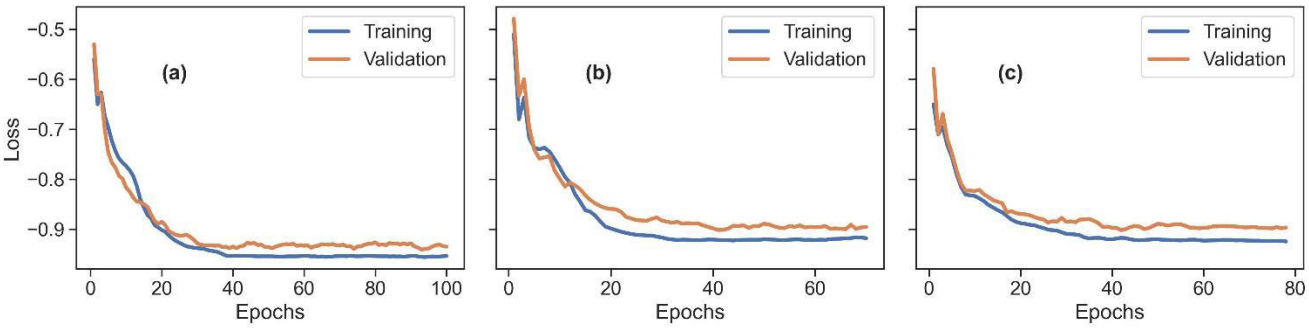

FIGURE 9. Learning curves for the proposed methods. a) ConvLSTM-based U-net + Skip and Up skip connections, b) ConvLSTM-based U-net + skip connections, c) ConvLSTM-based U-net.

mean evaluation metrics, the reduction in the standard deviation of these metrics further indicates the models' improved performance because of capturing the temporal nature of the MRI volumes. This reduction in the standard deviation reaches its peak in the proposed segmentation network in this study. This indicates that the addition of ConvLSTM layers to the encoding path and up skip connections make the predictive models more robust and confident about their predictions. A low standard deviation means less scattering of the evaluation metrics, implying that most of the values are concentrated around the mean.

In the following, the influence of incorporating the robust intensity normalization method is investigated in Table 5. This table reports the evaluation metrics of the ConvLSTM-based U-net + up skip connections and the standard U-net, both with and without the proposed preprocessing technique. As Table 5 shows, the segmentation performance of both the proposed ConvLSTM-based U-net and the standard

TABLE 3. Performance of the different segmentation methods (Test Subset).

Method		Dice score	Sensitivity	PPV	Jaccard
U-net	Mean	0.864	0.851	0.832	0.884
	STD	0.015	0.014	0.016	0.014
UNet-LSTM [37]	Mean	0.868	0.853	0.836	0.886
	STD	0.014	0.13	0.14	0.12
Cross-modality [33]	Mean	0.870	0.852	0.841	0.886
	STD	0.015	0.012	0.014	0.012
LSTM Multi-modal U-net [38]	Mean	0.873	0.856	0.850	0.889
	STD	0.012	0.012	0.014	0.014
ConvLSTM U-net	Mean	0.878	0.859	0.856	0.894
	STD	0.011	0.012	0.014	0.013
ConvLSTM U-net + skip connections	Mean	0.893	0.874	0.887	0.904
	STD	0.012	0.010	0.013	0.012
ConvLSTM U-net + up skip connections	**Mean**	**0.913**	**0.894**	**0.916**	**0.924**
	STD	**0.010**	**0.008**	**0.009**	**0.011**

TABLE 4. Improvement of the evaluation metrics of brain tumor segmentation by the addition of ConvLSTM blocks, skip connections, and up skip connections relative to each other and the standard U-net.

Method	Metric	Improvement relative to the U-net (%)	Improvement relative to the ConvLSTM U-net (%)	Improvement relative to the ConvLSTM U-net + skip connections (%)
ConvLSTM-based U-net	Dice score	1.6	-	-
	Sensitivity	1	-	-
	PPV	2.8	-	-
	Jaccard	1.1	-	-
ConvLSTM U-net + skip connections	Dice score	3.3	1.7	-
	Sensitivity	2.7	1.7	-
	PPV	6.6	3.6	-
	Jaccard	2.3	2.1	-
ConvLSTM U-net + up skip connections	Dice score	5.7	3.99	2.2
	Sensitivity	5	4.1	2.3
	PPV	10.1	3.6	3.3
	Jaccard	4.5	3.3	2.2

U-net deteriorates when the intensity normalization method is ignored. This is because the worsened comparability of different MRI images when they are not preprocessed by the proposed preprocessing technique. All runs were conducted using an Nvidia GeForce GTX 1660 Ti with 1536 CUDA cores and 6 GB of GPU memory.

In the final analysis, the evaluation metrics are reported for segmenting each region: necrosis and non-enhancing tumor (NNT), enhancing tumor (ET), and edema. To do so, each time one region in the segmentation mask is considered as class 1, while all other parts of the image are treated as class 0. Table 6 shows these metrics for each region and method. According to the results, all methods have a better performance on edema compared to their overall performance (Table 3). However, they exhibit inferior performance on necrosis and non-enhancing tumor (NNT) and enhancing tumor (ET) compared to their overall performance. This is primarily because most pixels in the images correspond to edema, while fewer pixels are associated with NNT and ET. However, just like the overall performance, the performance of the methods when evaluated separately on regions show significant improvement by transitioning from the vanilla U-net to the ConvLSTM Unet + up skip connections.

The results in Tables 3 and 6 are also comprehendable from Figure 10. As illustrated in this figure, the most similar segmentation mask to the ground truth mask is the prediction by the proposed ConvLSTM U-net with up skip

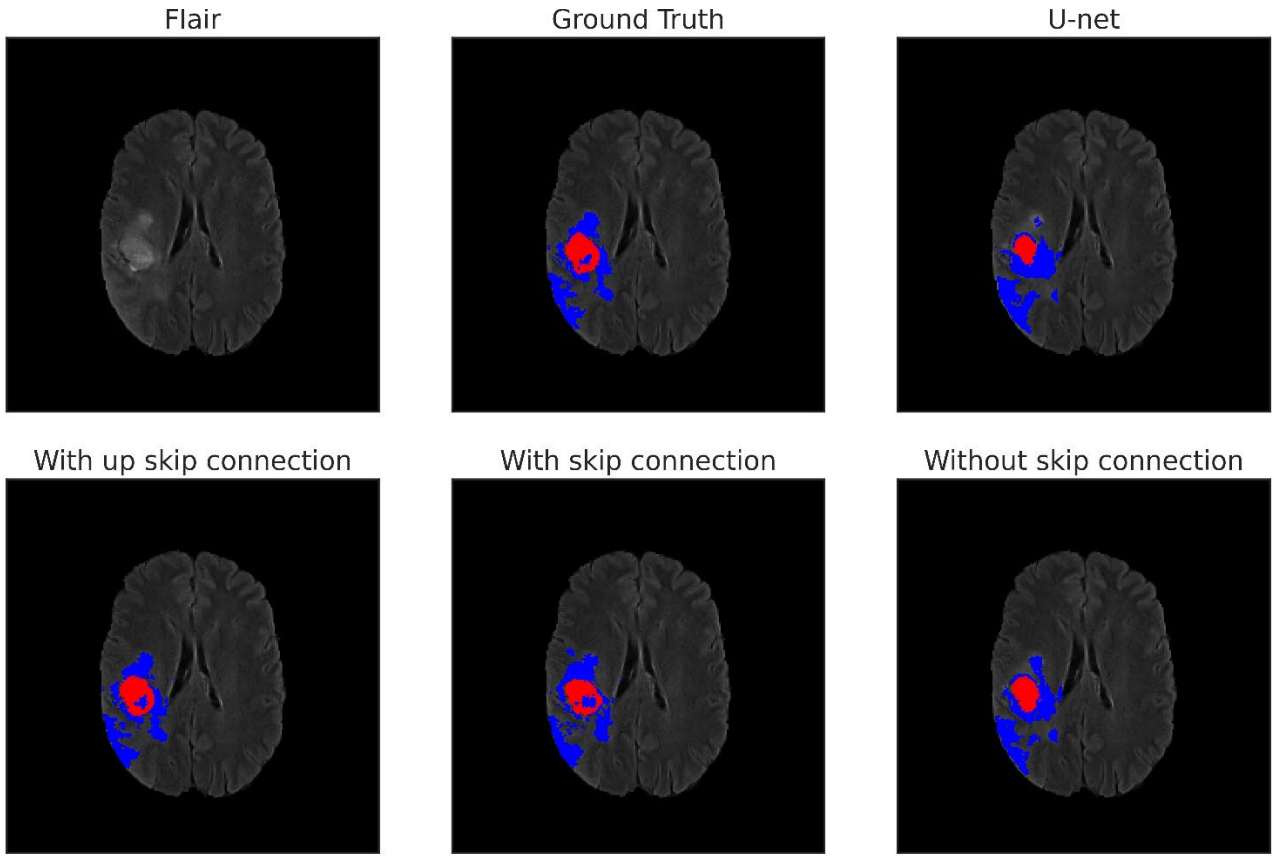

FIGURE 10. Segmentation results on a flair scan from the test dataset obtained from four segmentation methods.

TABLE 5. Comparison of the performance of both the proposed network and the standard U-net Model with and without the intensity normalization technique.

Method	Dice score	Sensitivity	PPV	Jaccard
U-net with robust intensity normalization	0.864	0.851	0.832	0.884
U-net without robust intensity normalization	0.842	0.816	0.776	0.854
Proposed model with robust intensity normalization	**0.913**	**0.894**	**0.916**	**0.924**
Proposed model without robust intensity normalization	0.884	0.860	0.855	0.898

connections. Moreover, the improvements are evident not ony in the overall segmentation but also is in each region (NNT, ET, and Edema). As can be seen, the delicate structure of each region is better captured by the proposed method, while other methods tend to produce lumped and smoothed boundaries for the region of interest. For example, the standard U-net has failed to correctly identify some red regions, capturing only a small part of it, and the upper part of the blue region (above the red region) is missing. It has also identified some healthy regions as edema, increasing the risk of false positives. When a ConvLSTM layer is added without using skip connections, the segmentation mask shows some improvement. The upper section of the blue region appears, and the red region grows. However, it is still not accurate enough, as both red and blue regions contain false positives and false negatives. This is because of information loss caused by contraction in the encoding path. By incorporating skip connections and up skip connections, this information loss is mitigated, leading to a much better segmentation by preserving information in the contraction path. Consequently, both ConvLSTM layers and up skip connections are necessary for reliable brain tumor segmentation in MRI volumes.

It should be noted that the proposed network in this study is limited to 3D MRI images, and applying it to 2D images is not recommended. Since 2D images do not have a temporal distribution in the third dimension, it is not logical to apply ConvLSTM2D layers to 2D inputs. It is suggested to explore the effect of the addition of attention modules, such as self-attention, transformers, etc. on the segmentation results. Additionally, integrating inception modules could enhance the ability of the network to simultaneously capture both delicate and large patterns in MRI scans.

TABLE 6. Evaluation metrics of the segmentation methods on each tumor region.

Method	Metric	NNT	ET	Edema
U-net	Dice score	0.844	0.856	0.874
	Sensitivity	0.835	0.849	0.856
	PPV	0.823	0.827	0.838
	Jaccard	0.865	0.882	0.890
ConvLSTM U-net	Dice score	0.867	0.875	0.883
	Sensitivity	0.849	0.858	0.862
	PPV	0.851	0.855	0.858
	Jaccard	0.885	0.893	0.897
ConvLSTM U-net + skip connections	Dice score	0.884	0.892	0.896
	Sensitivity	0.865	0.871	0.878
	PPV	0.880	0.887	0.889
	Jaccard	0.895	0.903	0.907
ConvLSTM U-net + up skip connections	Dice score	**0.908**	**0.910**	**0.916**
	Sensitivity	**0.889**	**0.893**	**0.896**
	PPV	**0.909**	**0.916**	**0.918**
	Jaccard	**0.918**	**0.923**	**0.926**

IV. CONCLUSION

An innovative deep learning-based brain tumor segmentation model was proposed in this study, which was extended the standard U-net architecture. The improvements included replacing traditional convolutional layers with ConvLSTM blocks and integrating up skip connections into the network. ConvLSTM blocks were employed due to the nature of 3D MRI volumes, which are time-lapses of 2D scans. Thereby, to capture the temporal distribution in the third dimension, ConvLSTM blocks were used to simultaneously extract spatio-temporal features. In addition, a novel, robust intensity normalization method was utilized to enhance the comparability of MRI volumes. The results on the BRATS 2015 dataset proved that the addition of ConvLSTM blocks could enhance the segmentation performance of the standard U-net across all evaluation metrics. Also, incorporating up skip connections further improved the network's performance measured, with the Dice score increasing by 5.7%, 3.99%, and 2.2% compared to the standard U-net, ConvLSTM-based U-net, and ConvLSTM-based U-net with skip connections, respectively. Therefore, it was proven that the integration of skip and up skip connections into the network recovers information lost during the encoding stage, enhances information flow during the feed-forward pass, and improves gradient flow in the decoding stage, making the learning of shallower layers more effective. Finally, the evaluation of the innovative preprocessing technique showed that it enahnced the performance of the proposed segmentation network by 3.3%, proving its effectiveness in increasing the comparability of MRI volumes.

REFERENCES

[1] A. Işın, C. Direkoğlu, and M. Şah, "Review of MRI-based brain tumor image segmentation using deep learning methods," *Proc. Comput. Sci.*, vol. 102, pp. 317–324, Jan. 2016, doi: 10.1016/j.procs.2016.09.407.

[2] M. Soltaninejad, G. Yang, T. Lambrou, N. Allinson, T. L. Jones, T. R. Barrick, F. A. Howe, and X. Ye, "Supervised learning based multimodal MRI brain tumour segmentation using texture features from supervoxels," *Comput. Methods Programs Biomed.*, vol. 157, pp. 69–84, Apr. 2018, doi: 10.1016/j.cmpb.2018.01.003.

[3] M. Soltaninejad, G. Yang, T. Lambrou, N. Allinson, T. L. Jones, T. R. Barrick, F. A. Howe, and X. Ye, "Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 12, no. 2, pp. 183–203, Feb. 2017, doi: 10.1007/s11548-016-1483-3.

[4] E. G. Van Meir, C. G. Hadjipanayis, A. D. Norden, H. K. Shu, P. Y. Wen, and J. J. Olson, "Exciting new advances in neuro-oncology: The avenue to a cure for malignant glioma," *CA, A Cancer J. for Clinicians*, vol. 60, no. 3, pp. 166–193, May 2010, doi: 10.3322/caac.20069.

[5] S. Bauer, R. Wiest, L.-P. Nolte, and M. Reyes, "A survey of MRI-based medical image analysis for brain tumor studies," *Phys. Med. Biol.*, vol. 58, no. 13, pp. R97–R129, Jul. 2013, doi: 10.1088/0031-9155/58/13/r97.

[6] A. Yonekura, H. Kawanaka, V. B. S. Prasath, B. J. Aronow, and H. Takase, "Automatic disease stage classification of glioblastoma multiforme histopathological images using deep convolutional neural network," *Biomed. Eng. Lett.*, vol. 8, no. 3, pp. 321–327, Aug. 2018, doi: 10.1007/s13534-018-0077-0.

[7] R. Yousefzadeh and M. Ahmadi, "Well trajectory optimization under geological uncertainties assisted by a new deep learning technique," *SPE J.*, vol. 29, no. 9, pp. 4709–4723, Sep. 2024.

[8] R. Yousefzadeh, A. Kazemi, M. Ahmadi, and J. Gholinezhad, "Dimensionality reduction methods used in history matching," in *Introduction to Geological Uncertainty Management in Reservoir Characterization and Optimization*. Cham, Switzerland: Springer, 2023, pp. 75–91, doi: 10.1007/978-3-031-28079-5_4.

[9] M. Y. Ansari, I. A. C. Mangalote, P. K. Meher, O. Aboumarzouk, A. Al-Ansari, O. Halabi, and S. P. Dakua, "Advancements in deep learning for B-mode ultrasound segmentation: A comprehensive review," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 8, no. 3, pp. 2126–2149, Jun. 2024, doi: 10.1109/TETCI.2024.3377676.

[10] N. Rasool and J. I. Bhat, *Brain Tumour Detection Using Machine and Deep Learning: A Systematic Review*. Cham, Switzerland: Springer, 2024, doi: 10.1007/s11042-024-19333-2.

[11] M. Y. Ansari, S. Mohanty, S. J. Mathew, S. Mishra, S. S. Singh, J. Abinahed, A. Al-Ansari, and S. P. Dakua, "Towards developing a lightweight neural network for liver CT segmentation," in *Medical Imaging and Computer-Aided Diagnosis*. Singapore: Springer, 2023, pp. 27–35.

[12] M. Y. Ansari, I. A. C. Mangalote, D. Masri, and S. P. Dakua, "Neural network-based fast liver ultrasound image segmentation," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jun. 2023, pp. 1–8, doi: 10.1109/ijcnn54540.2023.10191085.

[13] M. Y. Ansari, M. Qaraqe, R. Righetti, E. Serpedin, and K. Qaraqe, "Unveiling the future of breast cancer assessment: A critical review on generative adversarial networks in elastography ultrasound," *Frontiers Oncol.*, vol. 13, pp. 1–9, Dec. 2023, doi: 10.3389/fonc.2023.1282536.

[14] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440, doi: 10.1109/CVPR.2015.7298965.

[15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[16] O. Ronneberger, P. Fischer, and T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation* (Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9351. 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[17] M. Saha and C. Chakraborty, "Her2Net: A deep framework for semantic segmentation and classification of cell membranes and nuclei in breast cancer evaluation," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2189–2200, May 2018, doi: 10.1109/TIP.2018.2795742.

[18] C. Han, Y. Duan, X. Tao, and J. Lu, "Dense convolutional networks for semantic segmentation," *IEEE Access*, vol. 7, pp. 43369–43382, 2019, doi: 10.1109/ACCESS.2019.2908685.

[19] K. Kamnitsas, W. Bai, E. Ferrante, S. McDonagh, M. Sinclair, N. Pawlowski, M. Rajchl, M. Lee, B. Kainz, D. Rueckert, and B. Glocker, *Ensembles of Multiple Models and Architectures for Robust Brain Tumour Segmentation* (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 10670. 2018, pp. 450–462, doi: 10.1007/978-3-319-75238-9_38.

[20] K. Kamnitsas, C. Ledig, V. F. J. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, Feb. 2017, doi: 10.1016/j.media.2016.10.004.

[21] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017, doi: 10.1016/j.media.2016.05.004.

[22] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1240–1251, May 2016, doi: 10.1109/TMI.2016.2538465.

[23] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, and Y. Fan, "A deep learning model integrating FCNNs and CRFs for brain tumor segmentation," *Med. Image Anal.*, vol. 43, pp. 98–111, Jan. 2018, doi: 10.1016/j.media.2017.10.002.

[24] A. Myronenko, "3D MRI brain tumor segmentation using autoencoder regularization," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. Brainlesion Workshop*, 2019, pp. 311–320.

[25] H. Chen, Q. Dou, L. Yu, J. Qin, and P.-A. Heng, "VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images," *NeuroImage*, vol. 170, pp. 446–455, Apr. 2018, doi: 10.1016/j.neuroimage.2017.04.041.

[26] P. Dvořák and B. Menze, "Local structure prediction with convolutional neural networks for multimodal brain tumor segmentation," in *Medical Computer Vision: Algorithms for Big Data*. Cham, Switzerland: Springer, 2016, pp. 59–71.

[27] Y. Xue, T. Xu, H. Zhang, L. R. Long, and X. Huang, "SegAN: Adversarial network with multi-scale L_1 loss for medical image segmentation," *Neuroinformatics*, vol. 16, nos. 3–4, pp. 383–392, Oct. 2018, doi: 10.1007/s12021-018-9377-x.

[28] A. Beers, K. Chang, J. Brown, E. Sartor, C. Mammen, E. Gerstner, B. Rosen, and J. Kalpathy-Cramer, "Sequential 3D U-Nets for biologically-informed brain tumor segmentation," 2017, *arXiv:1709.02967*.

[29] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using U-net based fully convolutional networks," in *Medical Image Understanding and Analysis* (Communications in Computer and Information Science), vol. 723. Cham, Switzerland: Springer, 2017, pp. 506–517, doi: 10.1007/978-3-319-60964-5_44.

[30] J. Walsh, A. Othmani, M. Jain, and S. Dev, "Using U-Net network for efficient brain tumor segmentation in MRI images," *Healthcare Anal.*, vol. 2, Nov. 2022, Art. no. 100098, doi: 10.1016/j.health.2022.100098.

[31] T. Lei, R. Wang, Y. Zhang, Y. Wan, C. Liu, and A. K. Nandi, "DefED-net: Deformable encoder-decoder network for liver and liver tumor segmentation," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 6, no. 1, pp. 68–78, Jan. 2022, doi: 10.1109/TRPMS.2021.3059780.

[32] T. Zhou, S. Canu, P. Vera, and S. Ruan, "Latent correlation representation learning for brain tumor segmentation with missing MRI modalities," *IEEE Trans. Image Process.*, vol. 30, pp. 4263–4274, 2021, doi: 10.1109/TIP.2021.3070752.

[33] D. Zhang, G. Huang, Q. Zhang, J. Han, J. Han, and Y. Yu, "Cross-modality deep feature learning for brain tumor segmentation," *Pattern Recognit.*, vol. 110, Feb. 2021, Art. no. 107562, doi: 10.1016/j.patcog.2020.107562.

[34] R. Ranjbarzadeh, A. B. Kasgari, S. J. Ghoushchi, S. Anari, M. Naseri, and M. Bendechache, "Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images," *Sci. Rep.*, vol. 11, no. 1, May 2021, Art. no. 10930, doi: 10.1038/s41598-021-90428-8.

[35] K. V. Archana and G. Komarasamy, "A novel deep learning-based brain tumor detection using the bagging ensemble with K-nearest neighbor," *J. Intell. Syst.*, vol. 32, no. 1, Jan. 2023, Art. no. 20220206, doi: 10.1515/jisys-2022-0206.

[36] J. Nodirov, A. B. Abdusalomov, and T. K. Whangbo, "Attention 3D U-Net with multiple skip connections for segmentation of brain tumor images," *Sensors*, vol. 22, no. 17, p. 6501, Aug. 2022, doi: 10.3390/s22176501.

[37] S. R. Sowrirajan, L. Karumanan Srinivasan, A. D. Kalluri, and R. K. Subburam, "Improved brain tumor segmentation using UNet-LSTM architecture," *Social Netw. Comput. Sci.*, vol. 5, no. 5, p. 496, Apr. 2024, doi: 10.1007/s42979-024-02799-0.

[38] F. Xu, H. Ma, J. Sun, R. Wu, X. Liu, and Y. Kong, "LSTM multi-modal UNet for brain tumor segmentation," in *Proc. IEEE 4th Int. Conf. Image, Vis. Comput. (ICIVC)*, Jul. 2019, pp. 236–240, doi: 10.1109/ICIVC47709.2019.8981027.

[39] R. Ranjbarzadeh, P. Zarbakhsh, A. Caputo, E. B. Tirkolaee, and M. Bendechache, "Brain tumor segmentation based on optimized convolutional neural network and improved chimp optimization algorithm," *Comput. Biol. Med.*, vol. 168, Jan. 2024, Art. no. 107723, doi: 10.1016/j.compbiomed.2023.107723.

[40] Z. Wang, Y. Zou, H. Chen, P. X. Liu, and J. Chen, "Multi-scale features and attention guided for brain tumor segmentation," *J. Vis. Commun. Image Represent.*, vol. 100, Apr. 2024, Art. no. 104141, doi: 10.1016/j.jvcir.2024.104141.

[41] Z. Liu, L. Tong, L. Chen, Z. Jiang, F. Zhou, Q. Zhang, X. Zhang, Y. Jin, and H. Zhou, "Deep learning based brain tumor segmentation: A survey," *Complex Intell. Syst.*, vol. 9, no. 1, pp. 1001–1026, Feb. 2023, doi: 10.1007/s40747-022-00815-5.

[42] Z. Qian, L. Xie, and Y. Xu, "3D automatic segmentation of brain tumor based on deep neural network and multimodal MRI images," *Emergency Med. Int.*, vol. 2022, pp. 1–9, Aug. 2022, doi: 10.1155/2022/5356069.

[43] M. D. Giammarco, F. Martinelli, F. Mercaldo, and A. Santone, "High grade brain cancer segmentation by means of deep learning," *Proc. Comput. Sci.*, vol. 207, pp. 1633–1640, Jan. 2022, doi: 10.1016/j.procs.2022.09.220.

[44] K. Munir, F. Frezza, and A. Rizzi, "Deep learning hybrid techniques for brain tumor segmentation," *Sensors*, vol. 22, no. 21, p. 8201, Oct. 2022, doi: 10.3390/s22218201.

[45] Z. Li, Y. Liu, X. Guo, and J. Zhang, "Multi-convLSTM neural network for sensor-based human activity recognition," *J. Phys., Conf.*, vol. 1682, no. 1, Nov. 2020, Art. no. 012062, doi: 10.1088/1742-6596/1682/1/012062.

[46] M. Kistler, S. Bonaretti, M. Pfahrer, R. Niklaus, and P. Büchler, "The virtual skeleton database: An open access repository for biomedical research and collaboration," *J. Med. Internet Res.*, vol. 15, no. 11, p. e245, Nov. 2013, doi: 10.2196/jmir.2930.

[47] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, and L. Lanczi, 'The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015, doi: 10.1109/TMI.2014.2377694.

[48] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee, "N4ITK: Improved N3 bias correction," *IEEE Trans. Med. Imag.*, vol. 29, no. 6, pp. 1310–1320, Jun. 2010, doi: 10.1109/TMI.2010.2046908.

[49] K. Fritscher, P. Raudaschl, P. Zaffino, M. F. Spadea, G. C. Sharp, and R. Schubert, "Deep neural networks for fast segmentation of 3D medical images," in *Medical Image Computing and Computer-Assisted Intervention*. Cham, Switzerland: Springer, 2016, pp. 158–165.
[50] A. Sanchez-Caballero, D. Fuentes-Jimenez, and C. Losada-Gutiérrez, "Exploiting the ConvLSTM: Human action recognition using raw depth video-based recurrent neural networks," 2020, *arXiv:2006.07744*.
[51] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera, "Bi-directional ConvLSTM U-Net with Densley connected convolutions," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 406–415, doi: 10.1109/ICCVW.2019.00052.
[52] H. Song, W. Wang, S. Zhao, J. Shen, and K. M. Lam, *Pyramid Dilated Deeper ConvLSTM for Video Salient Object Detection* (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11215. 2018, pp. 744–760, doi: 10.1007/978-3-030-01252-6_44.
[53] R. Yousefzadeh and M. Ahmadi, "Fast marching method assisted permeability upscaling using a hybrid deep learning method coupled with particle swarm optimization," *Geoenergy Sci. Eng.*, vol. 230, Nov. 2023, Art. no. 212211, doi: 10.1016/j.geoen.2023.212211.
[54] R. Yousefzadeh and M. Ahmadi, "Improved history matching of channelized reservoirs using a novel deep learning-based parametrization method," *Geoenergy Sci. Eng.*, vol. 229, Oct. 2023, Art. no. 212113, doi: 10.1016/j.geoen.2023.212113.
[55] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," *BMC Med. Imag.*, vol. 15, no. 1, p. 29, Dec. 2015, doi: 10.1186/s12880-015-0068-x.

OSAMA MAJEED HILAL ALMIAHI received the bachelor's degree in computer science, the master's degree in information security, and the Ph.D. degree in information security, in 2007, 2013, and 2017, respectively. He is currently a Senior Lecturer with the College of Computer Science and Information Technology, University of Al-Qadisiyah. His research interests include image processing and analysis, deep learning, and cryptography.

ALAA TAIMA ALBU-SALIH received the B.Sc. degree in computer science from the University of Al-Qadisiya, Iraq, in 2005, the M.Sc. degree in information technology from Babylon University, Iraq, in 2013, and the Ph.D. degree in computer engineering from the Ferdowsi University of Mashhad, Mashhad, Iran, in 2018. He is currently a Senior Lecturer with the College of Computer Science and Information Technology, University of Al-Qadisiyah. He has more than 40 books in the field of machine learning, deep learning, and data science. He is the founder of the "Deep Learning in Arabic" website, which contains books, articles, projects, questions and answers, and a glossary in these fields. He also has mobile applications in these fields. His research interests include artificial intelligence, wireless sensor networks, mobility models, routing, software-defined networking, the Internet of Things, mathematical modeling, optimization, and unmanned aerial vehicles. He has several publications in these fields.

DHAFER ALHAJIM received the B.S. degree in computer technique engineering from The Islamic University of Najaf, Iraq, in 2012, and the M.Sc. degree (Hons.) in computer engineering in a subfield of computer architecture from Tabriz University, Iran, in 2018. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering, Faculty of Engineering, Shahid Chamran University of Ahvaz, Iran. He is also a Lecturer with the Computer Center, University of Al-Qadisiyah. His research interests include image processing, machine learning, deep learning, the IoT, and cloud computing.