Deep Convolutional Neural Networks for Multi-Modality Brain Image Classification

Rajasree R[1,*], Dr. Rajakumar S[2] and Dr. Josephin ShermilaP[3]
[1]Assistant Professor, Department of ECE, PSNInstitute of Technology and Science, Tirunelveli, Tamil Nadu, India
[2]Professor, Department of ECE, Panimalar Engineering College, Chennai, 600123, Tamil Nadu, India
[3]Assistant Professor, Department of ECE, R.M.K. College of Engineering and Technology, Chennai, 601206, Tamil Nadu, India
[*]Corresponding Author email: rajasreer935@gmail.com

Abstract

The brain is one of the important and complex organs that controls all the metabolic activities of human body. Tumor is formed in the brain due to unwanted growth of tissue. Multimodal Brain Tumor detection is a significant process to avoid mortality and severe illness. Multimodal brain tumor detection is a significant process to diagnose the presence of the tumor. In this work, the semantic type segmentation method is proposed to detect the multimodal brain tumor.The proposed multimodal semantic image segmentation for the classification and detection of brain tumors uses deep learning Multimodal Convolutional Neural Network (MMCNN). This network proposes a Bi-directional Long Short-Term Memory (BiLSTM) based segmentation method to test the integration of segmentation, feature extraction and classification. The effectiveness of the suggested methods to detect the brain tumor is tested with different scale sizes such as (12∗12, 24∗24, and 48∗48), (64∗64, 128∗128, 240∗240). In all the three scale sizes, the proposed method exhibits an increase in the accuracy and dice coefficient parameter. The multimodal tumor Magnetic Resonance Imaging (MRI) segmentation performance is improved by combining all the models pixel information retrieved from T1, T2, T1c and FLAIR different tumor modality images. Evaluation is done with the help of BRATS15 dataset. The proposed method attains the overall classification accuracy of 95.13% and the sensitivity of 0.9068. This is a better level of prediction that can lead to efficient semantic image segmentation over other algorithms that are utilized in handling higher data volumes.A five-fold cross-validation scheme is usedin this work for validated the MRI BRATS2015 dataset.

Keywords: Multimodal; CNN;MMCNN;Bi-LSTM; semantic segmentation;T1; T2; T1c; FLAIR; MRI; U-NET;MultimodalCNN; BRATS2015; CNN.
DOI Number: 10.14704/nq.2022.20.5.NQ22659 **NeuroQuantology 2022; 20(5):3610-3627**

1. Introduction

The brain tumour images need to be processed for its clinical diagnosis. The different stages in brain tumour image processing are pre-processing, feature extraction, segmentation and characterization. The problems which occur in clinical diagnosis of brain tumour are faulty identification of tumour images and less accuracy. To avoid these kind of discrepancies, a new compelling revelation system needs to be developed that can be appropriate in a wide scope of usages.

Developing new compelling revelation systems is not an easy task to increase the accuracy and to overcome faulty identifications. Among the several brain tumours, Glioblastoma, a complicated model based tumour type is considered in this research work. It consists of four different tumour models. A multimodal based semantic type classification is proposed in this work to increase the accuracy. The proposed approach is evaluated by a deep learning based algorithm.The beginning stage of gliomas is known as Low-grade gliomas and severe stage is known as high-grade glioma. Findings state that patients with low grade glioma will live longwhen compared to high-grade glioma [1]. So, brain Tumor diagnosing in an appropriate time is very essential. To attain the structure of the Tumorimage, several techniques are being used like Computed Tomography (CT), Computed Axial Tomography scan (CAT), Positron Emission Tomography (PET). Among these brain Tumor imaging techniques, the most popular method for brain Tumor image analysis and diagnosis is MRI. MRI is widely used since the images contain only necessary information for the characterization of infectious, brain disorders, and chronic diseasesof brain image. These divergence and spatial resolutions are very useful for brain image analysis. Also the characterization of Tumor class based structures based on Tumor location and shapes are very essential in clinics [2].

The clinical diagnosis of brain tumors by MRI necessitates manual annotation and segmentation of a cumbersome quantity of multimodal MRI images. Since this manual segmentation consumes more time and effort, there has been a continuous effort to process effective and accusative segmentation through automatic segmentation methods. Traditionally the experts followed only manual segmentation methods, but, at times it seems to be inaccurate due to its time consumption and human errors. So computer-aided diagnosis system is introduced to avoid human based mistakes such as diagnostic error, missing diagnosis, time-consuming and prone to errors [3]. However, the structure of theTumorfrequently varies based on its shape, size, volume and location among patients. Since automatic segmentation is also highly complex according to the factors like texture, intensity, shape and size, in treatment planning, the images are separated based on their texture, intensity, shape and size before image processing so as to avoid such complexities [4].

In Tumortreatment planning, the images are important process that involves partitioning the obtained image into various segments. Computer-aided diagnosis systemconsists of two methods, namely, semi automatic and automatic segmentation which are very much useful in clinics to enhance the treatment planning.In semi-automatic and automatic medical image classification, failures may occurusually due to the capability to sharpen the edges, handle the noise, and eliminate incorrect responses is very intricate [5]. So to avoid such classification discrepancies the Tumor images are separated based on the Tumor present area. Based on Tumor appearance area,Tumor image segmentation is classifiedinto the edge based, contourbased model,and region based models. In contour based modelthe boundary is detected using only the local measurements.Edgebasedtechniquesare used for finding the boundaries of objects within theimages[6].

Supervised learning techniques with valid truth on the data set are mostly applied, and they also rely entirely on the choice of features and their extraction. On the other hand, generative methods give probabilistic models with prior knowledge like location and spatial extent obtained from the healthy tissues. This information helps to explore the tumor compartments by comparing the healthy regions. This method has a complication because probabilistic models generated out of permeable knowledge are difficult for conversion.Frequently used supervised techniques areK-Nearest Neighbors (KNN),Support Vector Machine (SVM), and Artificial Neural Networks (ANN).Compared with unsupervised learning, supervised learning can handle different Tumor segmentation models.The supervised

learningmodals can be classified asTumor modality based, layer based and pixel based, and these model based segmentations are carried out by CNN [7].Tumor modality based glioma image segmentation needs more accurate Tumor present regionprediction, because based only on this processing outcome, the experts diagnose the type of the Tumor.

Glioma typeTumor cases are categorized into Low Grade Glioma (LGG) and High Grade Glioma(HGG) based on the severity level of the Tumor. There are four Tumor image modalities in BRATS'15 MRI dataset they are T1, T2, T1c, and FLAIR. Eachimage modality has class labels such as necrosis, edema, non-enhancing, and enhancing Tumor. So, when handling multi modal cases it is considered very complicated, time-consuming and very difficult to predict the accurate region of interest [8]. So many researchers have undergone to overcome these difficulties.Recently, 88% ofdice score value has been attained for both HGG and LGG Tumor cases an enormously randomized treesclassification method only in Fluid Attenuated Inversion Recovery (FLAIR) image sequence . Wu et al wereproposed Conditional Random Field's (CRF) framework fordifferent Tumor patients HGG and LGG cases segmentation, which have underperformed in LGG images [9].

For small dataset both the techniques produced expected results but for large dataset optimistic results were not that accurate. To attain a perfect segmentation of large dataset based multimodal Tumor image prediction is very needful. So, a simplified and self-learning method needs to be worked out.In recent years, deep learningbased networks have achieved state-of-the-art performance in medical image segmentation. Among the existing networks, CNN has been successfully applied on medical image segmentation [10].

From the available literatures, CNN based models are the frequent method used for multimodal MRI segmentation as well as classification for small data. According to the research CNN based techniquesproduced comparatively best results for multimodal class based segmentation. Therefore, for increasing the performance of large image dataset thedeep learning based techniques areintroduced for complex pixel based semantic type segmentation and classifications [11]. In brain Tumor analysis, multimodality MRImages is highly trained for deep learning CNNapproaches meanwhile it has complementary information of the image [12].

In recent years, deep learning-based networks have achieved state-of-the-art performance in medical image segmentation. Among the existing networks, U-Net has been successfully applied on medical image segmentation. In this work, U-Net with Densely connected convolutions (BCDU-Net) were introduced for medical image segmentation. The combinations of U-Net, bi-directional ConvLSTM and the mechanism of dense convolutions are utilized with fully connected layers. Single pathway network does not contain more information like two pathway networks.In two pathways MCNN for pixel classification and recognition,two modalitiesare used which adapts both local and global features [13].MCNN withRes-Net style classifier is introduced for semantic based segmentationto reduce the number of layers usage and increase the segmentation accuracywith two different image modalities. The numbers of layer increasedis proportional to processing time.So, to reducethe layers usage for multimodal image, CNN is adapted with UNET model. With different image modelsin CNN with UNET animage-netclassifier is usedfor large dataset to reduce the error rates [14].

In class label based semantic image classification,MCNN is a commonly used method. So deep CNN network is introduced for layer based multimodal cases, but the output does not obtain an optimistic result for the layer based Tumor diagnosis.Considering the multimodal diagnosis cases the trained output can get only based on the categories are learning and epochs. Accordingly, to produceoptimistic results the CNN is combined with Bi-LSTMbecause of itslong-term spatiotemporal features [15]. CNNwith LSTMis able to learn gestures of varying

length and complexity.In this method CNNconsists of two consecutive convolutional and pooling layers for feature map retrieval, from the retrieved feature map the features are extracted using LSTM softmax layer. In orderto improve the prediction accuracy of brain Tumor segmentation, a deep bidirectional long short term memory(Deep Bi-LSTM) method has been proposed byWanget.al. However the existing methods produced better results which have the following drawbacks such as running time,memory usage, low accuracy andnon-optimistic results. So,multi-modalbased semantic segmentation method has been proposed to overcome the drawbacks of existing methods listed in literatures.

2. Methods and Material

Modern biomedical imaging approaches with large dataset are appalling to examine. The results of segmentation depend on the exactitude and convergence time. Recently, there is a substantial need to investigate and execute new transformative calculations to tackle the issues related to glioblastoma type tumour image detection and classification. The detection and classification of glioblastoma type tumour is a complex and time consuming process. In this chapter, the extraction of the tumour from the brain image is obtained by a U-NET Convolutional Neural Network (U-NET CNN) algorithm [16]. A novel five layer U-NET type CNN structure is designed to actualize and separate the pixel features of the MRI. In this proposal deep learning based multiscale based tumour images are processed to attain the segmentation accuracy. Small-scale, largescale and multiscale MRI images are tested with the proposed architecture on standard benchmark MRI datasets collected from BRATS2015 [17]. In the present work, more than 500 brain MR images with different sizes are used for evaluation. The performance of the proposed method is validated by the outcome of the ground truth done by the neurologist.

2.1 Imaging Dataset

To assess the competence of the proposed approach, BRATS'15 is applied as an image dataset for the testing and training operations BRATS'15 dataset consists of 220 HGG and 54 LGG MRI scans of several patients. For the experimentations, 34 HGG and 120 LGG are expended as a training set, and 20 HGG and 70 LGG are utilized as the testing set from the BRATS'15. Multimodal MRI information is used for each patient in the BRATS'15 dataset in addition to that four MRI filtering systems (Pereira et al. 2016). These systems have been rectified for the wellness of each patient with T1-weighted (T1), T1-weighted imaging with gadolinium-ameliorating differentiation (T1c), T2-weighted (T2), and FLAIR. T1, T2, and FLAIR pictures are co-enrolled into the T1c information for each patient, necessitating recruiting spatial determinants, re-sampled and mediated into 1×1×1 mm3 through the size of 240∗240∗155. BRATS data sets are skull stripped and co-registered. The first step implicates preprocessing of 2D MRI slices extracted from the 3D volume for every patient where the intensity normalization and in-homogeneity alteration are done. Afterwards, two primary sets of features are extracted from every preprocessed image. In manual segmentation, the four intra-tumor classes are distinguished by distinct labels like necrosis (1), edema (2), non-enhancing (3), and enhancing tumor (4). In this work, T1, T2, T1c, and FLAIR images are accustomed to segment the whole tumor regions.

2.2 Overall Framework

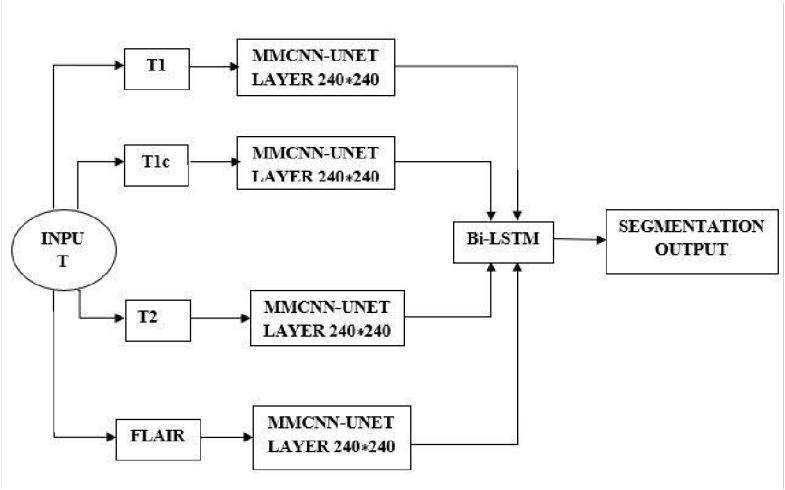

Figure 1: Overall Block Diagram

The overall block diagram of the proposed method is shown in Figure1. Multimodal Tumor identification and segmentation is intricate because the presence of intra-Tumoral structure is in irregular structure. In order to reduce the intricacies, a MMCNN with Bi-LSTM structure based semantic segmentation method is proposed for both LGG and HGG Tumor cases. In this proposal the BRATS'15 MRI dataset is given as input, it contains four Tumor image models for each patients such as T1, T1c, T2 and FLAIR. The performances of the class label-wise architectures is increased by using Multimodal input CNN-UNET layer with each modalities image patch size is 240∗240. After the UNET training the segmented 240∗240 size feature maps of four models are combined in a Bi-LSTM network for tumor class prediction. Bi-LSTM model is trained for identify the class labels of the intra tumoral structure. The 240∗240 scale value feature maps from different modalities are given as the input to Bi-LSTM layer for classification. Finally, the labeling of intra tumoral class is done in the Bi-LSTMs full connection softmax layer.

2.3 MMCNN-UNET layer

2.3.1CNN-UNET Layer

Pixel-wise semantic segmentation necessitates the labeling of individual pixels. In this semantic segmentation method, tumor labeling is grounded on the classes and is performed by segregating all the pixels into individual categories. Figure 2 illustrates the proposed five layer novel CNN U-NET layer architecture. The CNN U-NET system comprises of three layers namely the input layer, convolutional layers, and full connection layer. Contracting path or down sampling path, bottleneck path, and expanding path or upsampling path are three U-NET architecture paths. Every up and down sampling path consists of 4 blocks, and the down sampling block is built with a 3∗3 convolution layer with an activation function and a 2∗2 max-pooling layer. The number of feature maps are doubled over in every pooling operation in the U-NET layers. The input MRI is a 240∗240-pixel image, starting with 32 feature maps for the first block in the pooling operation, 64 for the second, and sequentially repeated. This contracting path pooling operation intends to obtain the input image context information for segmentation. The skip connections are utilized for retrieving coarse contextual details and transferring

*e*ISSN 1303-5150

them to the upsampling path. The network connectedness between the contracting paths and the expanding paths is constituted in U-NET architecture with the bottleneck path. Two convolutional layers build the bottleneck for connecting with the expanding path.

The first block is a deconvoluted layer with stride two operations in the expanding path, and the second layer establishes concatenation between the corresponding cropped feature maps obtained from the contracting path to the previous layer. Likewise, the third and fourth blocks are operated with a 2∗2 convolution layer and an activation function. The expanding path function is to locate the exact location by combining contextual data from the contracting path. Similar input sizes can be used for the next layer because of dense layers. Kernels are utilized as the learning parameters on convolution layers, and the kernel sizes are independent of the input image size 240∗240. This method is similar to the multimodal scale input image streams planned with CNN U-NET size 240∗240 for classification. Rather than classifying the whole input image, the contracting and expanding paths are accustomedto determine a pixel-wise segmentation map of the input MRI. This computerarchitecture is contrived for pixel-wise semantic segmentation, where the fully convolutional layer renders the feature maps of the entire image.

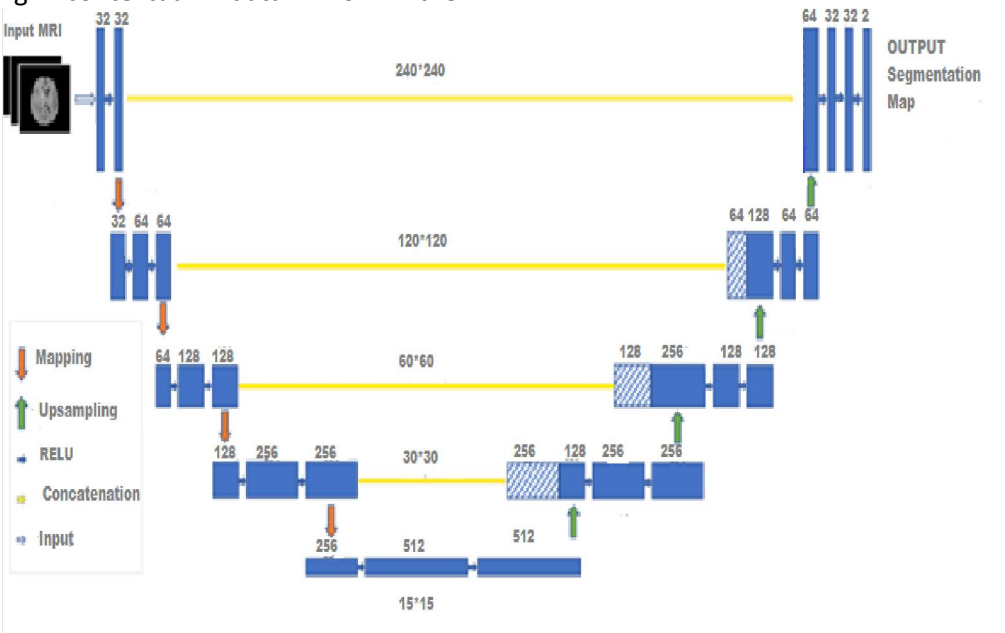

Figure 2: CNN-UNET Layer Architecture

Since the architecture is in U shape, it is termed as U-NET, and the architecture is symmetric and consists of two major parts. The left part is called the contracting path, constituted by the general convolutional process; the right part is an expansive path constituted by transposed 2d convolutional layers. In each process, the input layer constitutes two convolutional layers; the input MR brain image changes as 32, as the convolution process would increase the depth of the image. The red arrow pointing down is the max pooling process that halves the image size (the size reduced from 1142∗1142→ 1138∗1138 is due to padding issues). In layer 2, the number of channels changes from 32 → 64, as the convolution process would increase the image depth.

The red arrow pointing down is the max pooling process that halves the image size (the size reduced from 572∗572 → 568∗568 is due to padding issues). In layer 3, the number of channels changes from 64 →128. The red arrow pointing down is the max pooling process that halves the image size (the size reduced from 284∗284 → 280∗280). In layer 4, the number of channels changes from 128 → 256. The red arrow pointing down is the

max pooling process that halves the image's size (the size reduced from 140∗140 → 136∗136) due to padding issues. Finally, the fifth is a bottleneck where the number of channels changes from 256 →512. The red arrow pointing down is the max pooling process that halves the image size (the size reduced from 68∗68→ 64∗64 is due to padding issues). The image at this moment has been resized to 64∗64∗512. The expansive path of the image is upsized due to its original size.

Transposed convolution is an upsampling technique that expands the size of images. It does some padding on the original image, followed by a convolution operation. After the transposed convolution, the image is upsized from 64∗64∗512 → 128∗128∗256, and then, this image is concatenated with the corresponding image from the contracting path and together makes an image of size 128∗128∗512. The reason behind this is to combine the information from the previous layers to get a more precise prediction. The fourth and fifth convolution layers are added together. In the output layer, the last layer is a convolution layer with 1 filter of size 2∗2 where there is no dense layer in the entire network. The rest left is the same for neural network training. The input and output layers are concatenated with 240∗240, line second to fifth layers are concatenated with the upsampling layer by 120∗120, 60∗60, 30∗30 and 15∗15 accordingly. The experimental results are calculated using a percentage match between segmentation result and ground truth. Figure 3 shows the four intra-Tumoral modalities used in this method such as T1, T1c, T2 and FLAIR of a patient MRI slices.

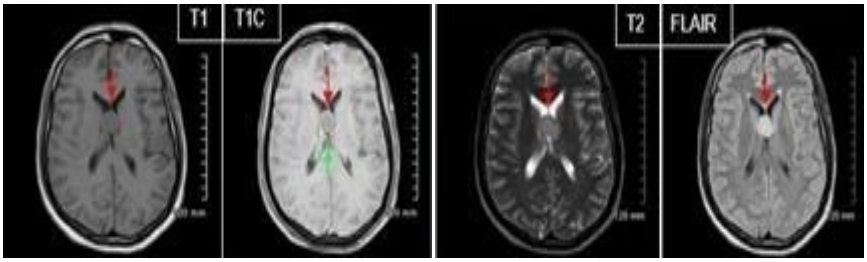

Figure 3: T1, T1c, T2 and FLAIR of patient MRI slices

Five hundred epochs are distributively trained for retrieving a complete tumor segmentation mask. Equated with the original U-NET model, the dice score calculation is performed by halving the feature map usage in this proposed methodology. This approach is implemented since it involves less time for training similar pads in the convolutional layers. Hence, the cropping function across the concatenation of feature maps are avoided.

Up-convolutions 2∗2 are executed throughout the upsampling process applying the Keras upsampling 2D function. Moreover, in the transposed model, the Keras convolution and 2D transpose function are dispensed to complete the up-convolution 2∗2. The upsampling mechanism is designed by a 2∗2 convolution network where a ReLU succeeds every convolution network with a 2∗2 max-pooling process. During this process, the feature maps are heightened and doubled the number of feature channels. In the upsampling path, the feature map of 2∗2 convolutional layers are accustomed to dilute the feature channels. The crop feature is required to avert the loss of border pixels in the convolution process. Therefore, the crop feature map is settled up with two 3∗3 convolution layers. They are passed through a ReLU layer. The entire convolution layers are grouped with 2∗2 max-pooling operations in the final layer to select the intended tile size. The desired numbers of classes are held back with 1∗1 convolutions in the final layer to map each of the 32-component feature vectors.

2.3.2. MMCNN

In this method deep learning based multimodal convolutional neural network is

designed for intra-Tumoral case images of 240∗240 sizes. Here, the different size images are used for testing the input image sizes but the highest segmentation accuracy output has attained by only the 240∗240 size images. Because of the small scale values like 64∗64,120∗120 are only contains small image informations than 240∗240. So that only in this work was proposed with an image size of 240∗240. In this method, all the four Tumor modalities are used because of the different modals features have more image feature information than single model. All the modalities are combined together with their local and global features throughout the classification. Five-layer 32∗32 block size architecture of MMCNN is designed with one input layer and five convolutional layers such as C1, C2, C3, C4, and C5. ForTumor modal imageclassificationpurpose along with the convolutional layer one max pooling layer is added. The convolutional layers are the building blocks of CNN's because based on the convolutional layers only the features are obtained. Each convolutional layer is added with afilter kernel size of 3∗3.

In the first convolutional layer C1,the image feature map of input MRI is obtained based on the related pixels present in the image. After the 5th convolutional layers(C5), the pooling function will deliver the needful feature map i.e., highest value feature map at every time interval. It is alsovery useful for withdrawing the size of the corresponding feature map. Individually, weights of the predicted four different modalities feature maps are joined together to produce a new pathway network structure. Proposed MMCNN is used to extract the method of contradictoryresponsibilities in anexact hierarchy of feature. In addition to that, the outputs of four models are combined using the full connection layer for the final classification. The four modals features are arranged by full connection layer and utilize cooperatively for classification.

2.5 MMCNN with Bi-LSTM

2.5.1 LSTM

This approach is a multi-modal deep learning model for enhancing image segmentation with large-scale images possessing multiple modalities. It is based on combining MCNN based pixel recognition techniques with LSTM. In the LSTM layer, the long-term spatiotemporal features of images have been obtained, and U-NET patch sequences have been identified. Features from the U-NET have been combined and used with LSTM for finding the events for patch sequences. These LSTM aided patch sequences have been used in creating new patch sequences for classification. In this network analysis, the entire image with the convolutional features has been learned amongst individual frames and temporal dynamics. Graphics Processing Unit (GPU) based technique has been utilized for the processing intended to cut down the CPU training time.Traditional neural network models had combined layers in two adjacent layers, and nodes between them. This network structure made unexpected effects while using a larger data sequence [18-21]. To overcome this, the nodes between the hidden layers have been connected in RNN models, and these inputs belonging to hidden layers include the output of these layers. Gradient disappearance during the training process is a demerit which makes RNN complicated within long sequences. This eventually results in transmitted long sequences that are very complex to handle, and this has been managed with the LSTM method [22, 23].

2.5.2 Bi-LSTM

Traditional neural network models had combined layers in two adjacent layers, and nodes between them. This network structure made unexpected effects while using a larger data sequence. To overcome this, the nodes between the hidden layers have been connected in RNN models, and these inputs belonging to hidden layers include the output of these layers. Gradient disappearance during the training process is a demerit which makes RNN complicated within long sequences. This eventually results in transmitted long sequences that are very complex to handle, and this has been managed with the LSTM method.

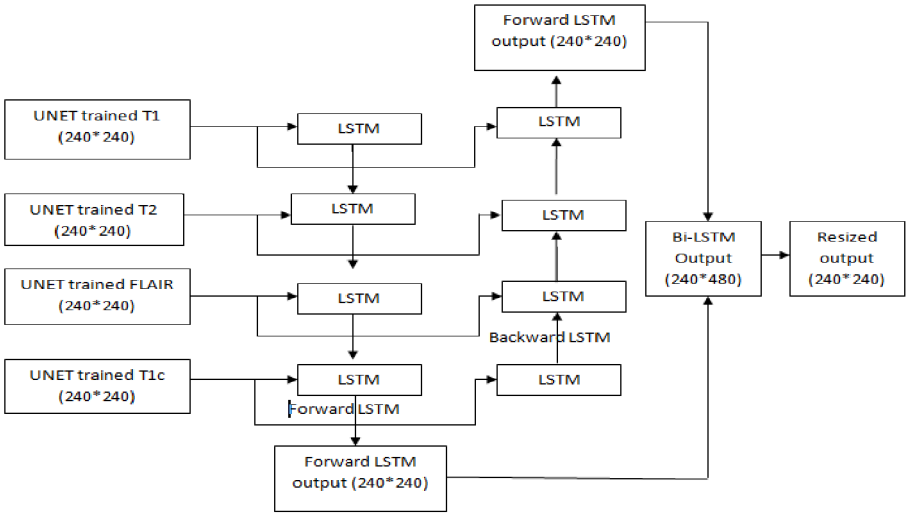

Figure4: BILSTM structure

Figure 4 depicts the BiLSTM layer structure; memory blocks in the LSTM network substitute the hidden layer summation units of standard RNN. The LSTM architecture has a set of recurrently connected subnets called memory blocks. Each block has three multiplicative units, which are input, output, and forget gates with one or more self-connected memory cells. These multiplicative units impart constant analogs of reading, writing, and reset operations for the cells. These gates are nonlinear summation units that collect activation functions from the outside and inside the blocks and control the cell's activation function using multiplication functions.

To overcome the gradient problems in memory cell storage and access of information, LSTM avoids multiplicative long period operation. Bidirectional LSTM (BiLSTM) is an upgraded version of LSTM and is with two LSTM layer superimpositions. In BiLSTM, the layer learns long-term bidirectional dependencies between time steps of time series or data sequence. These dependencies between time and data can be used for the network to learn the complete time series at each time step. BiLSTM trains two instead of a single LSTM over the input sequence in problems where all time steps of the input sequence were available. In BiLSTM, the input sequences have been connected in the forward direction, and the reversed copy has been connected in the reverse direction. It provided additional context information to the network and results in a faster and full learning network in the output. So the BiLSTM network has been trained with the input sequence in the forward and backward direction. Both the LSTM layers have been connected to the same output layer, i.e., the output at every momentary time knows the complete information of the entire sequence.

2.5.2 MCNN with Bi-LSTM

In thismethod 240∗240 size input image is used instead of small-scale images, because a large size image contains more image information capacity than small scale sizes. After the features extraction inU-NET modelclassification is done by the full connection classification layer. Each and every pixel'sis subjected to classificationthroughoutthe convolution process.In this work,feature recognitionis done by considering the local and global regions features. MMCNNwithBi-LSTM based network isusedfor evaluate the image continuousby learning the convolution features among each and every frames. And temporal dynamics between frames based onground truth or temporal feature design. In this work, the classifications are made by the Bi-LSTM network's softmax layer. So in this classification approach Bi-LSTMis in a position to detect same length matched events for

patch sequences retrieval.Also, the patches are made byusing the longer context information between thesequence'sframes.

For multimodal class-basedTumor cases, the input matching scores are measured from the retrieved features for semantic classification. If multiple matches are present the weighted averaging is used to combine the similar scores for Tumorclasses' identification. After the class labelformation,the fusion is to be done. In this work the score level fusion is made by classlabel matcher. If the label matcher provides a similarity score, it comes under same feature vector labels. Similarly, four labels are assigned by label names1, 2, 3, 4 i.e., necrosis (1), edema (2), non-enhancing (3) and enhancing Tumor (4). Then the similar value feature vector labels are combined by fusion. Fusion is obtained by normalizing the corresponding scores which are in the same range.Thenthe similar normalized scores values among the fourTumorclassessuch as necrosis, edema, non-enhancing and enhancing Tumor are combined. Experiments have shown that the MMCNN-Bi-LSTM network is trained efficiently. To evaluate the performance by applying the network to original MR image sequences to find events forclasses.

3 Experiment Setup

MMCNN methodology is proposed for large-scale multi modal images based on deep learning method. In this method, the MMCNN based pixel recognition technique is combined with Bi-LSTM. In the Bi-LSTM layer, the long-term spatiotemporal features of images are used for obtaining class based segmentation map.For finding the events for class sequences, the retrieved features from both U-NET and Bi-LSTM are used. These class sequences are very useful to create new class sequences for classification. U-NET with Bi-LSTM network analyses the image continuous and discover the convolutional features between every frame and sequential dynamics among the frames. To reduce the training time of the method instead of CPU system the proposed model is processed using Graphics Processing Unit (GPU) system.

The BRATS'15 dataset is used for the system evaluation which contained images of several Tumor types, which is divided into 24,800 training samples and 9,600 test samples. The tests are conducted using each patient's three adjacent image slices from each Tumor modal imagessuch as T1, T1c, T2 and FLAIR. For testing the manually segmented MRITumor masks are used. These manually segmented masks are the ground truth and used for assessing the systems efficiency. In the UNET training process these ground truth masks were used. The results are calculated bydeterminetheamountof match between ground truth and predicted segmentation results.

To achieve best results CNNs U-NET layer was used. In the U-NET layer the convolutional layers are used for convolving an image with kernels to obtain the feature maps. In each convolution layers the kernels weights are used to connect the previous layer through a feature map element. In back propagation, in order to enhance the input characteristics the kernel weights are adapted during the training phase. In testing and training phase the context informations are necessary. So, for context information retrieval the neighbourhood information is taken into accountusing kernels. In each neural unit a non-linear activation function is applied, which is responsible for non-linear data transformation. Several convolutional layers are stacking up the extracted features become more accurate, because the feature are more trained based on the increased number of convolution layers and the training depth. To speed up training process in this method Rectifier linear units (RELU) are used.

In the feature maps pooling, operator combines the nearby spatial features. This combination of probable features makes the output feature representation more compact and invariant to insignificant details.It also decreases the next stage computational load. Finally, all the features are joined using max-pooling operator.Dropout layer and data augmentation are used to reduce the over fitting also thedata augmentation can be used for increase the size of training sets.

Categorical Cross-entropy is used for minimizing the loss functions in our proposal.

Tumorimages ofsingle-scale multi-modal classification problems with 4 intra tumoralclasses are proposed.To assess the competence of the proposed approach, BRATS'15 is applied as an image dataset for the testing and training operations BRATS'15 dataset consists of 220 HGG and 54 LGG MRI scans of several patients (Havaei et al., 2015). For the experimentations, 34 HGG and 120 LGG are expended as a training set, and 20 HGG and 70 LGG are utilized as the testing set from the BRATS'15. Multimodal MRI information is used for each patient in the BRATS'15 dataset in addition to that four MRI filtering systems (Pereira et al., 2016). These systems have been rectified for the wellness of each patient with T1-weighted (T1), T1-weighted imaging with gadolinium-ameliorating differentiation (T1c), T2-weighted (T2), and FLAIR (Vishnuvarthanan et al., 2018). T1, T2, and FLAIR pictures are co-enrolled into the T1c information for each patient, necessitating recruiting spatial determinants, re-sampled and mediated into 1×1×1 mm3 through the size of 240∗240∗155.

In this methodology, brain tumor segmentation is executed by a single-scale multi-modal classification problem with four classes: necrosis, edema, non-enhancing, and enhancing tumor. However, in brain tumor images, the classes are imbalanced, so samples from the underrepresented classes and random samples from others are expended. Moreover, the number of samples of necrosis and enhancing tumors is smaller in the LGG training set. To surmount this, the intensities of HGG and LGG are normalized and calculated. To train the U-NET for HGG and LGG, this methodology extracts around 450,000 and 335,000 patches, where 45% of these patches represent normal tissue in HGG and 55% in LGG. The learning rate is linearly reduced after each epoch during the training stage.

In evaluation of segmentation the following four metrics are considered: Dice Coefficient (DC), Positive Predictive Value (PPV) sensitivity, and accuracy. The DC assesses the overlap between the manual and the automatic segmentation. It is defined as,

$$DiceCoeffvient = \frac{2TP}{2TP + FP + FN} \quad (1)$$

TP (True Positive) is the number of cases correctly identified as unhealthy, and FP (False Positive) is the number of cases incorrectly identified as unhealthy. TN (True Negative) is the number of cases correctly identified as healthy, and FN (False Negative) is the number of cases incorrectly identified as healthy. PPV is a measure of the amount of FP and TP, defined as,

$$PositivePredictiveVdue = \frac{TP}{TP + FP} \quad (2)$$

Sensitivity is useful to evaluate the number of TP and FN detections, and it is defined as

$$Sensitivity = \frac{TN}{TN + FP} \quad (3)$$

Finally, accuracy is calculated for finding the segmentation result of the method compared with manual segmentation results. The expression for Accuracy is given in equation 4.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Sensitivity and accuracy values are presented in percentage throughout this thesis.

4 Experiment Results and Discussion

This section analyzes themetrics, input scale size selection and the performance of the plane over extractedclasses of the proposed method. In addition to that, comparisons between proposed methods to other methods which are using the same dataset are made. BRATS'15 dataset is taken for validation of the proposed technique. The dataset contain MRI pictures of 9,600 test sets and 24,800 training sets. A fivefold cross-validation assessment scheme is utilized. The tests are directed by utilizing three adjacentimages from every model of T1, T1c, T2 and FLAIR of each patient. The quantitative outcomes are determined utilizing a quantity rate of coordinate between ground truth and predicted results.

*e*ISSN 1303-5150

To get the results of every individual Tumor region separately on the proposed approach by the choiceover the performance improvement. Theincrease in performance is evaluated by the mean gain ofDSC, PPV and Sensitivity three metrics. Then the comparisons are made between the CNN (UNET) with small scale input pixel values to the proposed methods 240∗240 scale value. Which are obtained bycalculateall metrics usingMMCNN methodology for the data sets.Then the metricsare calculated usingMMCNN-BILSTMmethod.

Finally every metrics for the two methodsare calculated bythe subtractions and getting the mean gain. The metric of every experiment is reported in Table1.

Table 1: The Tumor region based output metrics

Scale size	Tumor region	Dice Coefficient	Positive Predictive Value	Sensitivity
CNN-Scale1 (64∗64)	CompleteTumor region	0.84	0.832	0.863
	Core region	0.703	0.81	0.76
	Enhancing region	0.57	0.79	0.68
CNN-Scale2 (128∗128)	CompleteTumor region	0.842	0.851	0.854
	Core region	0.73	0.82	0.81
	Enhancing region	0.565	0.75	0.65
CNN-Scale3 (240∗240)	CompleteTumor region	0.8753	0.8632	0.8949
	Core region	0.803	0.83	0.82
	Enhancing region	0.61	0.77	0.69
MMCNN-BILSTM (240∗240)	CompleteTumor region	0.9427	0.9491	0.9531
	Core region	0.84	0.86	0.85
	Enhancing region	0.72	0.81	0.78

From the Table1 it is concluded that our experiments performs well in completeTumor region, core region and enhancing region over other CNN's different small scale values over.Small dataset kind of cases imaging pictures the tissue separation is very small. It is the reason for most of the case in machine learning with large data setfailed to explore data augmentation. Havaei et al. deliberation of its application, however found to be ineffective in their system.Brain Tumors are implanted by intra-tumoral structures with completely different volumes leading to unbalanced range samples of every category. So, the representation of different classes impairs the performance of the CNN. Therefore, to investigatedata augmentation for balancing the quantity of samples from each classes, samples extracted from necrosis and enhancing regions in HGG is used as trainingsamples ofLGG.Sampling the data from HGG to LGG is done for enhancing the training of the CNN.

In this work ReLU activation function is used for the gradients, which may increase the convergence during the optimization and lead to fast training.So, it is investigated that the effectivenessof this activation functions in brain Tumor segmentation.To maintain the similarsmall kernels sizes using cascaded layers of small 3x3 kernels is the advantage of this method.Whereas,the number of kernels' size reduction and more non-linear transformations allowing on the data's are the best accuracy improving criteria's. BrainTumor segmentation of this method is done by cascaded each convolutional layers before every max-pooling operationin the proposed UNET architecture.This operation is performed bysingle layer with larger kernels with the correspondingupsampling field. So, in this case we changed the group of layers such as 1, 2, 3,4 and 5 by single convolutional layer with 3x3 kernels each in down sampling, while

in the up sampling we changed the groups of layers 1,2,3,4 and 5 by one layer with 2x2 kernels each. The Dice Coefficient and Positive Predictive Values of different CNN scale values and proposed method over 500 epochs are described in Table 1.

Table 2: Dice Coefficient and Positive Predictive Values of results of the proposed method

Name	Dice Coefficient	Positive Predictive Value	Sensitivity
CNN-Scale1 (64∗64)	0.84	0.832	0.863
CNN-Scale2 (128∗128)	0.842	0.851	0.854
CNN-Scale3 (240∗240)	0.8753	0.8632	0.8949
MMCNN-BiLSTM (240∗240)	0.8963	0.9047	0.9068

Table 2 shows that the MMCNN-BiLSTM achieves better Dice coefficient and PPV thanthe CNN scale methods. The proposed methodobtains Dice Coefficient is0.9136which is 7.83%, 6.85%, 3.09% better than the other methods respectively. The Positive Predictive Value of the proposed method is0.9261which is 9.81%, 8.11%, 5.71% better than the CNN scalemethod respectively in tumor segmentation.However, in the proposed experiment,500 epochs obtained improvement result over with the other scales because of the 240∗240 scale contains more information than individual scales.Table 2exhibits the sensitivity of the proposed model that compare with ground truth with500 iterations. FromTable 2, it can be inferred that the proposed method clearly predicts more sensitivityvalues of MMCNN over other three smallscales CNN.The proposed method obtains 0.9314 sensitivity which is 6.27%, 7.24%, 2.84% better than the CNN scales in Tumor segmentation using LGG and HGG.However, the proposed experiment 500 epochs obtained improvement result over with the other scales because of the 240∗240 scale contains more information than small scale values.Figure 5 exhibits the segmentation accuracy results of the proposed method and compare with ground truth of the 500 iterations.

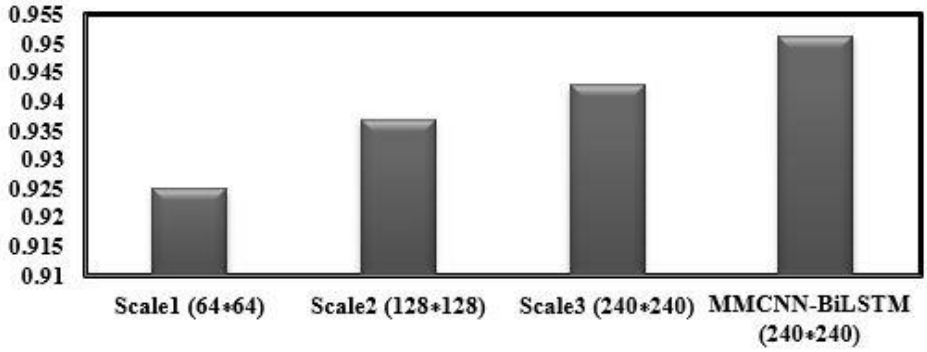

Figure 5: Segmentation accuracy results of MMCNN with different size of pixels

Figure5represents the responsiveness impacts of top three information picture fix sizes on cerebrum Tumor picture division. In this diagram X hub portrays the best three information picture fix sizes and Y pivot means precision. This strategy improves Tumor division utilizing the T1, T2, T1c, FLAIR pictures for LGG and HGG cases. It got 0.9513 precision which is 2.99%, 1.74%, 1.13% better than the CNN scale1 (64∗64), CNN scale2 (128∗128) and CNN scale3 (240∗240) techniques individually north of 500 ages. From this outcome, the proposed strategy accomplishes preferred execution over the other CNN scale models over Accuracy. The aftereffect of this analysis in HGG, LGG cases is great in view of the elements advanced by the U-NET are figured in neighborhood and worldwide districts. Notwithstanding, in the proposed approach, 100 ages acquired

improvement result over with different scales in light of the consolidated 64∗64,128∗128 and 240∗240 scales contains more data than individual scales.

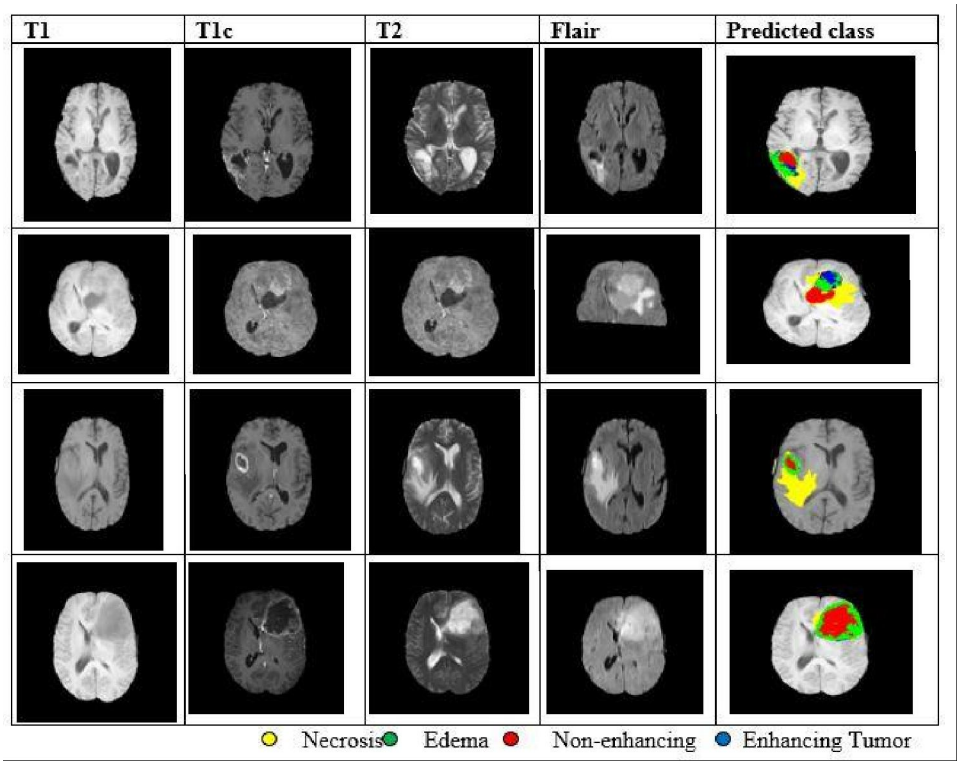

Figure 6: The randomly selected segmentation results of three different patients

Figure 6delineates haphazardly chosen division aftereffects of three unique patients utilizing MRI. It shows three patients with four unique Tumor modalities that were accurately distinguished and fragmented. The pictures of the principal second third and fourth lines are the three unique patients input picture methodology, for example, T1, T1c, T2 and FLAIR of manual explanation in the preparation set. The fifth column is the three distinct patients anticipated result of three fix scales division result. The division yields comprising of four models like corruption, edema, non-endlessly improving Tumor that are shown in the outcomes as yellow, green, red and blue variety veils individually.

4.1 Comparison between different small-scale inputs

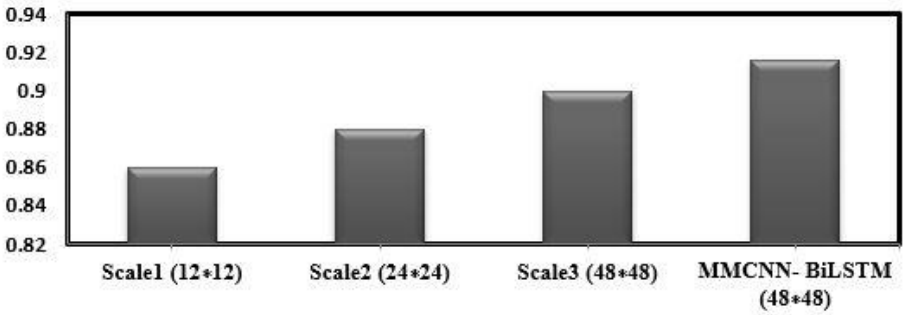

Figure 7: The segmentation accuracy of methods integrating low scale values

Figure 7 shows the precision aftereffects of the limited scale sizes like 12∗12, 24∗24, 48∗48 sizes. In this correlation the three little information fix sizes are utilized for single little pathway MMCNN-BiLSTM. What's more, each fix size goes about as single pathway of our U-NET. Examination is additionally made among them. From Figure 7, it is plainly clear that the limited scale CNNs Tumor order exactness is improved when contrasted and

increment fix sizes. All in all, MMCNN-BiLSTM with 48∗48 fix sizes is perform better compared to the next single fix sizes, for example, 24∗24 and 12∗12. This is on the grounds that 48∗48 fix size gives more worldwide scale data to CNNs highlight learning. Sadly 12∗12 and 24∗24 preparation occasion have less precision contrasted and 48∗48 pathway CNNs. This is without a doubt a result of the absence of worldwide elements in 12∗12 pathway. Clearly through consolidating both worldwide and nearby highlights of 48∗48 scale in MMCNN-BiLSTM and the exactness is improved and comes to almost 0.916 for the BRATS'15 dataset. From Figure 7 plainly clear that, the limited scale patches doesn't contain a lot of data than the high scale patches. So in the proposed strategy the 64∗64, 128∗128 and 240∗240 fix scale values are utilized for increment the division precision. Figure 8 outlines the upsides of our proposed MMCNN-BiLSTM model over limited scope values.

4.2 Comparison between combinations of top three Scale values

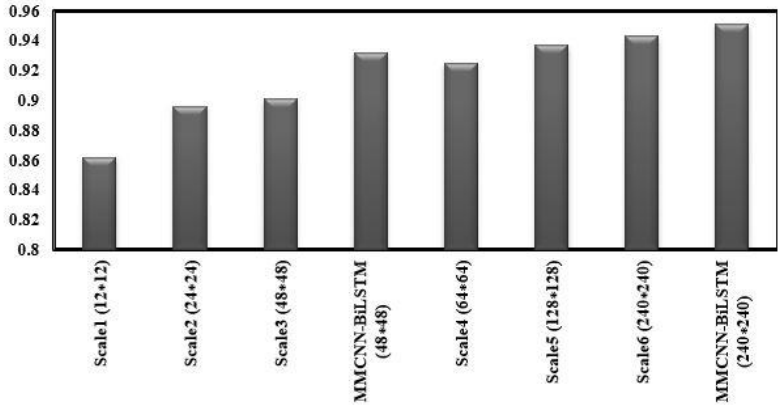

Figure 8: segmentation accuracy comparisons between low scale values with top three scales

Figure 8 outlines the exactness aftereffects of the different scale sizes. In this examination the main three appropriate fix sizes are utilized for single-pathway MMCNN-BiLSTM. The rising scale upsides of CNNs produces further developed Tumor characterization precision contrasted and single way CNNs which is displayed in Figure 8. Overall, the preparation of individual fix sizes containing 64∗64 fix sizes performs better compared to the next single fix sizes, for example, 128∗128 and 240∗240. This is on the grounds that 240∗240 fix size gives worldwide scale data to CNNs highlight learning. 128∗128 and 64∗64 preparation occasion have less precision when contrasted and 240∗240 single pathway CNNs. This is presumably a result of the absence of worldwide elements in 64∗64 way pathway.

In the proposed strategy, contrasted and the 12∗12, 24∗24, and 48∗48 scales to the best three scales created precision around 0.92 which is 0.4 exactness low qualities than 240∗240, 128∗128, and 68∗68 scales. Clearly through joining both worldwide and nearby highlights by 240∗240, 128∗128, and 68∗68 scales and the precision is improved and comes to almost 0.9513 for the dataset. Figure 9 delineates the upsides of our proposed MMCNN - BiLSTM model over condition of craftsmanship strategies.

4.3 Comparison with other state-of-the-art methods

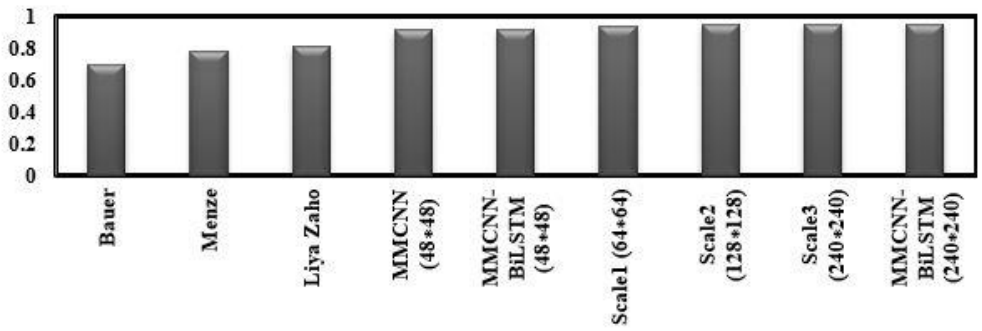

Figure 9: classification accuracy of proposed method with state-of-the-art techniques

The graph shows the effects of the four CNN based techniques for BRATS. Additionally, in light of a similar assessment the proposed technique performed well when contrasted and different strategies accessible in the writing. Figure delineates the examination with different techniques to our proposed strategy. The proposed work is contrasted and the best four performing calculations. Top edge of the blue square shape represents the mean precision of every strategy. Precision of single-pathway MMCNN of 64∗64, 128∗128, and 240∗240 scale sizes is higher than different techniques. Notwithstanding, execution of MMCNN with scale size of 240∗240 is superior to conventional limited scope CNN strategies, which shows that the size 240∗240 is an appropriate decision. The proposed strategy acquires a precision worth of 0.9513 which is 0.26 higher than the Bauer et al., (2013) model. The proposed strategy beats the Menze et al., (2014) technique by 0.17 precision i.e., Menze strategy got just an exactness of 0.78 which is 0.17 lower than the proposed strategy. As indicated by Menze et al., the most troublesome errands in mind Tumor division are the division of the center locale for LGG and the upgrading area for HGG. The precision of proposed MMCNN achieves 0.14 higher than the Liya Zaho et al., (2015) strategy ie., Zaho's technique happen just 0.81 yet in proposed strategy accomplishes 0.9513 , so MMCNN is a steady strategy. So the proposed MMCNN-LSTM beats well. Be that as it may, the exhibition of MMCNN with fix sizes 64∗64, 128∗128, and 240∗240 is better compared to the customary MCNN technique over 0.42 exactness values. From the above expressed examinations with best in class strategy the precision of proposed MMCNN - LSTM technique.

Liya Zhao proposed limited scope based CNN strategy, which performed better in the center district in DSC and outflanked all techniques in the total and center areas in awareness. In any case, a huge drop in execution in PPV in similar locales is noted, it is gathered that likely the technique by Zhao over fragmented the Tumor. Be that as it may, the strategy proposed by Menze, Bauer and MCNN techniques delivered less exactness in center district contrasted and the proposed technique. Considering the presentation in single informational collection, Menze and Bauer techniques were comparative in dividing the total Tumor. For the most part the strategy proposed by Menze was better just in the center area in the mean time the proposed technique in portraying the entire construction. Considering the cutting edge, the MMCNN based approaches have been utilized for bigger channels and shallow models, with highlights figured by the U-NET as contribution to a LSTM for organized expectation.

5. Conclusion

Clinically brainTumorrecognition and analysis are difficult task in recent days.So,Tumorsegmentation wants to be precise and efficient. To conquer thesemantic type segmentationdifficulties the MMCNN with BiLSTM has been proposed. And, also to improve the segmentation over large-scale image dataset withmulti modalities such as T1, T2, T1c, and FLAIRproposed intending for pixel feature extraction. Deep convolution networks-basedU-NET method conquers superior results for the core Tumor regionsand finest results for the complete

Tumor regions producing state-of-the-art result. Theproposed methods performance evaluation is done by a benchmark datasetBRATS'15 it contains various patients MRI of HGG and LGG cases.The proposed method obtained high accuracy value than the Bauer, Menze, Liya Zaho and traditional MCNN methods. The results of the proposed method when compared with the manual delineate ground truth produced economical and robust segmentation.In future, other CNN's architectures such as Alex net, Z netmay be designed to further exploit its self-learning property and improve segmentation accuracy by extracting richer boundary information.

Statements and Declarations

Funding
Authors did not receive any funding.

Conflicts of interests
Authors do not have any conflicts.

Availability of data and material
Available on request.

Code availability
Not applicable.

Acknowledgment
The Author with a deep sense of gratitude would thank the supervisor for his guidance and constant support rendered during this research.

References

1. Jin L et al. A survey of MRI-based brain tumor segmentation methods. Tsinghua Science and Technology, 2014. 19(6): 578-595.
2. Whittle I.R. The dilemma of low grade glioma. Journal of Neurology, Neurosurgery & Psychiatry, 2004. 75(suppl 2): ii31.
3. Kang E. et al. Deep Convolutional Framelet Denosing for Low-Dose CT via Wavelet Residual Network. IEEE Transactions on Medical Imaging, 2018. 37(6): 1358-1369.
4. Bengio Y. et al.Representation Learning: A Review and New Perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013. 35(8): 1798-1828.
5. Fan-Hui K. Image retrieval based on Gaussian Mixture Model. in *2012 International Conference on Machine Learning and Cybernetics*. 2012.
6. Savitha R. et al.Projection-Based Fast Learning Fully Complex-Valued Relaxation Neural Network. IEEE Transactions on Neural Networks and Learning Systems, 2013. 24(4): 529-541.
7. Zhang Y. et al.Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Transactions on Medical Imaging, 2001. 20(1): 45-57.
8. Shahzadi I. et al. CNN-LSTM: Cascaded Framework For Brain Tumour Classification, 2018. 633-637.
9. Akkus Z. et al. Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions. Journal of Digital Imaging, 2017. 30(4): 449-459.
10. Moon W.K. et al. Computer-Aided Tumor Detection Based on Multi-Scale Blob Detection Algorithm in Automated Breast Ultrasound Images. IEEE Transactions on Medical Imaging, 2013. 32(7): 1191-1200.
11. Gonçalves V.M. et al.A systematic review on the evaluation and characteristics of computer-aided diagnosis systems. Revista Brasileira de Engenharia Biomédica, 2014. 30: 355-383.
12. Brosch T. et al. Deep 3D Convolutional Encoder Networks With Shortcuts for Multiscale Feature Integration Applied to Multiple Sclerosis Lesion Segmentation. IEEE Transactions on Medical Imaging, 2016. 35(5): 1229-1239.
13. Zaharchuk G. et al. Deep Learning in Neuroradiology. American Journal of Neuroradiology, 2018.
14. Pereira S. et al. Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. IEEE Transactions on Medical Imaging, 2016. 35(5): 1240-1251.

15. Zhao L, K. Jia, Multiscale CNNs for Brain Tumor Segmentation and Diagnosis. Computational and Mathematical Methods in Medicine, 2016. 2016: 7.
16. Soltaninejad M. et al. Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI. International Journal of Computer Assisted Radiology and Surgery, 2017. 12(2): 183-203.
17. Dong H. et al. Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks, in Medical Image Understanding and Analysis2017. 506-517.
18. Hu, Y, XiaY.3D Deep Neural Network-Based Brain Tumor Segmentation Using Multimodality Magnetic Resonance Sequences. in Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. 2018. Cham: Springer International Publishing.
19. Wang X. et al. Beyond Frame-level CNN: Saliency-Aware 3-D CNN With LSTM for Video Action Recognition. IEEE Signal Processing Letters, 2017. 24(4): 510-514.
20. Li S. et al. Generating image descriptions with multidirectional 2D long short-term memory. IET Computer Vision, 2017. 11(1): 104-111.
21. Tsironi E. et al. An analysis of Convolutional Long Short-Term Memory Recurrent Neural Networks for gesture recognition. Neurocomputing, 2017. 268: 76-86.
22. Bauer S. et al. A survey of MRI-based medical image analysis for brain tumor studies. Phys Med Biol, 2013. 58(13): p. 0031-9155.
23. Menze B.H. et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Transactions on Medical Imaging, 2015. 34(10): p. 1993-2024.

*e*ISSN 1303-5150