TimeDistributed-CNN-LSTM: A Hybrid Approach Combining CNN and LSTM to Classify Brain Tumor on 3D MRI Scans Performing Ablation Study

SIDRATUL MONTAHA[1], SAMI AZAM[2], A. K. M. RAKIBUL HAQUE RAFID[1], MD. ZAHID HASAN[1], (Member, IEEE), ASIF KARIM[2], (Member, IEEE), AND ASHRAFUL ISLAM[3]

[1]Department of Computer Science and Engineering, Daffodil International University, Dhaka 1207, Bangladesh
[2]College of Engineering, IT and Environment, Charles Darwin University, Casuarina, NT 0909, Australia
[3]School of Computing and Informatics, University of Louisiana at Lafayette, Louisiana, LA 70504, USA

Corresponding author: Sami Azam (sami.azam@cdu.edu.au)

ABSTRACT Identification of brain tumors at an early stage is crucial in cancer diagnosis, as a timely diagnosis can increase the chances of survival. Considering the challenges of tumor biopsies, three dimensional (3D) Magnetic Resonance Imaging (MRI) are extensively used in analyzing brain tumors using deep learning. In this study, three BraTS datasets are employed to classify brain tumor into two classes where each of the datasets contains four 3D MRI sequences for a single patient. This research is composed of two approaches. In the first part, we propose a hybrid model named TimeDistributed-CNN-LSTM (TD- CNN-LSTM) combining 3D Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) where each layer is wrapped with a TimeDistributed function. The objective is to consider all the four MRI sequences of each patient as a single input data because every sequence contains necessary information of tumor. Therefore, the model is developed with optimal configuration performing ablation study for layer architecture and hyper-parameters. In the second part, a 3D CNN model is trained respectively with each of the MRI sequences to compare the performance. Moreover, the datasets are preprocessed to ensure highest performance. Results demonstrate that the TD-CNN-LSTM network outperforms 3D CNN achieving the highest test accuracy of 98.90%. Later, to evaluate the performance consistency, the TD-CNN-LSTM model is evaluated with K-fold cross validation. The approach of putting together all the MRI sequences at a time with good generalization capability can be used in future medical research which can aid radiologists in tumor diagnostics effectively.

INDEX TERMS Deep learning, brain tumor classification, 3D MRI, hybrid CNN LSTM, 3D CNN, ablation study.

I. INTRODUCTION

Brain tumor refers to the abnormal growth of lesions inside the brain tissue or spinal cords which is one of the leading causes of cancer mortality worldwide. Gliomas are a type of tumors originating in the glial cells of the brain. These tumors can be classified into two main categories benign and malignant, depending on molecular features, where benign denotes noncancerous cells and malignant denotes cancerous cells.

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

According to the World Health Organization (WHO) [1], malignant brain tumors of grades I and II are considered as low-grade glioma (LGG) whereas grade III and IV tumors are considered as high-grade glioma (HGG). Life expectancy of HGG patients is low, approximately 1 to 2 years, whereas patients having LGGs survive on average 5 to 10 years. LGGs are typically considered slow-spreading infiltrative tumors [2]. Early detection, accurate grading and correct identification of the type of tumor are imperative as a timely diagnosis can aid significantly in decreasing the mortality rate [3].

The diagnosis of tumors is usually done using two common methods named open brain biopsies and brain imaging. In stereotactic biopsies, after drilling a tiny hole into a human skull, a small slice of tissue is taken out for observation under a microscope. However, this technique is invasive and risky [4]. Risks of the biopsy test include excessive bleeding and damage to the brain caused by the biopsy needle. This may result in severe migraine headaches, stroke, coma, infection, seizures and might even lead to death [5]. Moreover, biopsies are often inefficient, time-consuming, and life-threatening due to the exceedingly invasive characteristics of the technique [6]. As a solution, noninvasive imaging procedures such as single-photon emission computed tomography (SPECT), positron emission tomography (PET), computed tomography (CT), MRI, infrared spectroscopic imaging, or sometimes a combination of all these, have been utilized for the diagnosis and classification of brain gliomas [7]. MRI can distinguish soft tissue and identify small changes in tissue density and physiological changes correlated to tumors. Furthermore, type, dimension, location of tumor and malignancy grade can be determined with conventional and advanced MRI scans which are now commonly employed to differentiate LGGs from HGGs [8]. These scans also known as MRI sequences, apprehend various possessions of tumors depending on different time and intensity settings where each sequence is regarded essential in the detection of different tumor sub-regions [9]. However, the categorization of gliomas manually by doctors is often challenging and error-prone, due to the hidden and complex characteristics of low-grade and high-grade features. The drawbacks of these diagnostic routines are that they are time-consuming and not always accurate [10]. As the number of cancer cases is growing over the period, it is quite impossible for the doctors to diagnose every patient especially at the primary stage. Moreover, in some rural areas around the world, timely diagnosis becomes hindered due to the lack of expert radiologists. In this regard, a non-invasive computer-aided fully automatic diagnostic system is necessary to aid clinical experts in diagnosis and treatment planning and reduce the mortality rate by providing more reliable, faster and more accurate tumor detection [11]. An automatic medical image examination method can lessen diagnosis time, error, increase the stability of results and thus can reduce the strain on radiologists. CNN based models have had major accomplishments in medical image research and analysis, however there remain challenges in accurately classifying grades using 3D pictures [12]. Moreover, developing a fully automated method with the highest accuracy to categorize brain tumor is still a challenging task, due to the size, texture, localization and intensity similarity of malignant regions to surrounding tissues [13].

In this research, we have introduced two deep learning approaches to classify brain tumor into HGG and LGG to compare their performance, where in first approach a hybrid model combining CNN with LSTM and wrapped by TimeDistributed function is proposed and in second approach a 3D CNN model is employed. Three BraTS 3D MRI datasets are used in this research which includes four 3D MRI sequences named T1-weighted MRI (T1), T1-weighted MRI with contrast enhancement (T1ce), T2-weighted MRI (T2), and Fluid Attenuated Inversion Recovery (FLAIR) for each patient. To date, a number of methods have been developed for the detection, segmentation and classification of gliomas using 3D brain MRI. However, in most studies, segmentation or classification is carried out using a single MRI sequence. To the best of our knowledge, no previous researches have classified HGG and LGG employing four 3D sequences of each patient as a single input. However, radiologists examine all the sequences to determine most accurate tumor grading as all of the sequences contain relevant features [14]. Following radiologist's working principle, training a deep learning model with all of these sequences might have impact on improving interpretation performance. Therefore, the aim of this paper is to propose a deep learning model with highest accuracy which is able to interpret all the 3D sequences of a patient as a single input data. From this concern, in the first approach, the model is developed in such a way that, all the four types of 3D MRI images of a patient can be passed as single input data.

The proposed model is named TimeDistributed-CNN-LSTM (TD-CNN-LSTM), where each layer of CNN architecture is wrapped with a TimeDistributed function. In this process, LSTM is introduced to learn higher level parameters while feature extraction is accomplished using CNN. LSTM can deal with the issue of vanishing gradients efficiently, unlocking certain memory positions in a spatial context. The input layer is configured using a TimeDistributed wrapper to pass the four 3D images of a single patient as one input. However, interpreting all the MRI sequences together with optimal performance especially in 3D is quite challenging. In order to get the optimal model architecture depending on the highest accuracy, an ablation study is performed by changing different hyper-parameters and layer architecture. In the second approach, a 3D CNN model is trained using each of the four 3D MRI sequence respectively. The purpose of introducing the second approach is mainly to show the effectiveness of our first approach by analysing all the MRI sequences of a patient at once. For both of the models, BraTS 2018 and BraTS 2019 are combined to increase the number of images and used as train dataset where BraTS 2020 dataset is employed as test dataset. To evaluate the performance of both models, evaluation metrics named accuracy, precision, recall, F1-score, specificity, mean absolute error (MAE), root mean squared error (RMSE) and the Area Under Curve (AUC) have been analyzed. The results obtained with the TD-CNN-LSTM and 3D CNN models are compared to determine which technique yields the best performance having consistency. Results suggest that our proposed TD-CNN-LSTM model outperforms 3D CNN in terms of accuracy, which evidences that instead of analysing single sequences, employing all the sequences can significantly aid in improving performance. Moreover, ablation study has been proved to

be an effective approach in developing a deep learning model with the highest performance. Therefore, in future medical research on 3D MRI imaging, performance of the model can be improved effectively by employing all the imaging of a patient and carrying out the ablation study.

The rest of the paper is laid out in eight sections. The lLiterature review is conducted in section II which also includes the limitations of previous studies. Section III discusses the aim and scope of this study. Datasets that are used in this research are discussed in section IV and visualization of 3D MRI images is shown in section V. Preprocessing of these 3D images is conducted in section VI. Both of the approaches (hybrid CNN LSTM and 3D CNN) proposed in this study are briefly explained in section VII. The computed results of the entire study are showcased in section VIII along with a comparison with the proposed hybrid approach and existing literature. Lastly, section IX presents the conclusion of this study.

II. LITERATURE REVIEW

In recent years, there have been major advances in the field of Brain Tumor Classification. Deep convolutional neural networks (DCNN) are often used because of their high accuracy but this comes at the cost of a long computational time for each epoch. Jude *et al.*, [15] employed DCNN by modifying the training algorithm to classify brain tumors into four classes. A total of 220 images comprising of T1, T2 and T2 flair sequences, were used in this work. Their proposed model DCNN achieved an average accuracy of 96.4% accuracy in classifying the tumors. Of these images, 80 images were in the training set where four classes have the equal number of train data (20 images). This is a highly balanced dataset and there is a significant shortage in training images. In a multi-class classification problem, usually for such a small number of data, the model cannot be trained properly and thus may result in increasing false interpretation rate. However, 2D MRI with size of 256∗256 was used in this study where employing 3D MRI and increasing the number of images might aid to improve their performance. Mzoughi et al [16] proposed a 3D CNN network to classify brain gliomas into LGG and HGG resulting in an accuracy of 96.49% in the validation dataset. T1-graded 3D MRI sequences from 'BraTS' 3D MRI dataset, containing 209 HGG and 75 LGG MRI scans were utilized in this work. However, except for accuracy, no other performance evaluation matrics are explored in this study and employing other MRI sequences might be an effective approach. DeepSeg model has been proposed by the author of this study [13], to detect and segment brain tumors using MRI images. In this research, the BraTS 2019 dataset was employed containing 259 HGG and 76 LGG MRI scans with four sequences where only the Flair MRI sequence was used for segmentation of size 224 × 224. Mohamed *et al.*, [17] created a tumor grading classifier with a pre trained VGG16 model and a CNN based U-net architecture. The Cancer Imaging Archive (TCIA) containing imaging data for LGG MRI with sequences T1 pre-contrast,

FLAIR, and T1 post-contrast of total 110 patients were employed for both segmentation and classification tasks. After segmenting the images, using a VGG16 based CNN classifier, they were able to classify LGG into grade II and grade III with an accuracy of 89%, a sensitivity of 87% and a specificity of 92%. However, in tumor segmentation only 2D FLAIR sequence was used. In classification, they used 2D MRI data instead of using 3D MRI which provides more details of the brain tumor. Training and validating a CNN model with different sequences may result in inconsistent performance as these contain different information and features of brain tumors. A similar study regarding this issue was conducted by Banzato *et al* [18]. They used a transfer learning model GoogLeNet to classify meningiomas and gliomas across all sequences respectively. A total of 80 cases, 56 meningiomas and 24 gliomas were used in their study. Each case offers brain MRI image of 512×512 dimensions which was scaled down to 224×224 pixels. The authors trained the model using post-contrast T1 images to develop their proposed classifier named trCNN. The highest test accuracy of 94% was recorded with post contrast T1 sequence whereas for Pre contrast T1 and T2 sequences, accuracy of 91% and 90% were achieved respectively. However, this study employed 2D data and the model was trained with each of the MRI sequences individually. A comparison in classification performance between 2D CNN and 3D CNN models in brain tumor classification has been done by the authors of this study [19], They used TCIA dataset comprising of 108 MRI images and the BraTS 2018 dataset containing 210 HGG and 75 LGG images where each dataset contains T1, T1-Gd, T2, and FLAIR sequences. A 3D brain tumor segmentation model, based on U-net architecture, was proposed. The authors recorded an accuracy of 97.1% with 3DConvNet (3D CNN) which was higher compared than the 96.3% accuracy gained with the 2DCNN model. Thiruvenkadam *et al.*, [20] developed six CNN classifiers in order to find the optimal model to classify HGG and LGG lesions in brain MRI images. All their models were constructed by changing the combination of hyper-parameters. For training these models BraTS 2013 dataset was used where 2D slices (240×240) are extracted from 3D brain MRI images. Each model had been tested with eight different volumes of Whole Brain Atlas (WBA) dataset and the average results from all eight volumes were calculated. They achieved the highest accuracy of 88.91% from the model five layers with stopping criteria and batch normalization (FLSCBN). Though, 2D CNNs are widely used in the field of deep learning, they are not flawless in computer vision tasks with 3D volumetric data. To address this issue, some researchers experimented with LSTM models to classify brain tumor in 3D MRI images. Such an approach was proposed by Iram *et al.*, [21] who introduced a combination of CNN and LSTM for the classification of 3D MRI images. They used three transfer learning models named AlexNet, ResNet, and VGGNet for feature extraction and classified with a LSTM model utilizing the features respectively. The BraTS 2015 dataset containing four MRI

sequences having an image dimension of 240 × 240 × 155 was chosen where only the FLAIR sequence was used for their experimentation. However, researchers balanced the dataset by keeping 60 images in HGG and 60 images in LGG class. Their model performed best with the highest accuracy of 84% using VGG16 as feature extractor. A similar approach was proposed by Rukmani *et al.*, [12] who used VGG16 and AlexNet model for feature extraction and LSTM to classify tumor based on the features. For dataset, BraTS 2015 was used containing 250 HGG and 50 LGG cases of four MRI sequences. However, only FLAIR sequence was utilized for the experiment. With their LSTM model, the highest classification accuracy of 85% was recorded with features extracted by the VGG16 network, whereas features extracted by AlexNet resulted in an accuracy of 70%. Other authors [11] developed a CNN architecture of 22 layers for classifying brain tumors into three classes using 3064 2D slices of T1 weighted contrast enhancement MRI images instead of 3D sequence. Training the model using an image size of 256 × 256 and employing 10-fold cross validation the highest accuracy of 96.56% was recorded.

Most of the studies described above contain a few common limitations of employing only a single sequence, employing 2D slices instead of 3D sequences, and not experimenting with the proposed model using different configurations. The performance of their models could be improved with optimal classification accuracy if these issues could be addressed.

III. RESEARCH AIM AND SCOPE

The aim of this research is to mimic radiologist's analysis pattern into a deep learning model regarding to tumor identification using 3D MRI scans. As Artificial Intelligence is developed following the concept of human brain, we hypothesize that, if a deep learning model is developed based on the analysis scheme of radiologists, an accurate, reliable and effective performance might be achieved that can outperform other existing approaches. The main contributions of this study can be summarized as follows:

1. In this study, three 3D MRI datasets named BraTS 2018, BraTS 2019 and BraTS 2020 are used to increase the volume of images and conduct the experiments.
2. In pre-processing step, all the 3D images are normalized to a standard scale of 0-1 using the min-max normalization technique and resized to 128 x 128 x 32 to increase overall performance and decrease the computational complexity.
3. Instead of using single MRI sequences, all the sequences and the Region of Interest (ROI) of a particular patient are aimed to pass to the CNN model as one input.
4. A hybrid TD-CNN-LSTM model is proposed where the layers of CNN are wrapped with a TimeDistributed function, in order to pass all the 3D images of a single subject as one input.
5. A 3D CNN model is introduced and trained with single sequences, in order to compare the performance with

the TD-CNN-LSTM network in terms of accuracy and performance consistency.
6. Both of the models are trained, validated with a combined dataset of BraTS 2018 and BraTS 2019 and tested with BraTS 2020 dataset.
7. An ablation study of ten cases is performed to determine the best configuration of the proposed TD-CNN-LSTM model in terms of the highest possible accuracy and lowest possible time complexity by altering any components or layers from the architecture.
8. Both of the models are evaluated using different performance metrics: Accuracy, Precision, Recall, Specificity, F1-score, MAE, RMSE and AUC score.
9. To check the performance consistency of TD-CNN-LSTM model rigorously, the model is trained multiple times using K-fold cross validation with different k values.

Our entire interpretability pipeline is represented in Fig. 1.

IV. DATASET

For the classification task, three BraTS datasets named BraTS 2018 [22], BraTS 2019 [23] and BraTS 2020 [24] are explored in this research where the datasets are composed of two classes, HGG and LGG. Each of the dataset contains four 3D MRI sequences and one corresponding ROI for a single patient or subject in NIfTI files. Additionally, each 3D volume contains 150 2D slices/images of MRIs from different sections of the brain. The single slices are of dimension 240 x 240 and in single channel grayscale format.

In BraTS 2018 dataset, a total of 282 cases are found where 208 cases are of class HGG and 74 cases are of class LGG. BraTS2019 dataset comprises 331 cases where 257 cases are found in class HGG and 74 cases are found in class LGG. The BraTS 2020 dataset was initially divided into training and a validation set while collecting from the resource. The validation set comprises 120 cases of both classes HGG and LGG but the grade is not publicized in the provided comma separated value (CSV) file [25]. Therefore, only the training dataset is used for this research. In this regard, a total of 365 useable cases are collected from BraTS 2020 dataset, where 291 cases are of class HGG and 74 cases are of class LGG. In all three datasets, there is a significant imbalance between HGG and LGG classes and a scarcity of LGG class can be seen. To address this issue, BraTS 2018 and BraTS 2019 datasets are combined to train and evaluate both of the models. In the combined dataset, a total of 361 subjects are present where 210 cases are of class HGG gathered from BraTS 2018 dataset and 151 cases are of class LGG gathered from both BraTS 2018 and BraTS 2019 datasets thus minimizing the scarcity of LGG class. Lastly, BraTS 2020 dataset is employed as a test set with the purpose of evaluating how both of the models perform with an external test dataset (BraTS 2020).

The three datasets are summarised in Table 1.

Fig. 2 illustrates the forming of training and testing dataset used in this research as well as axial slice of four
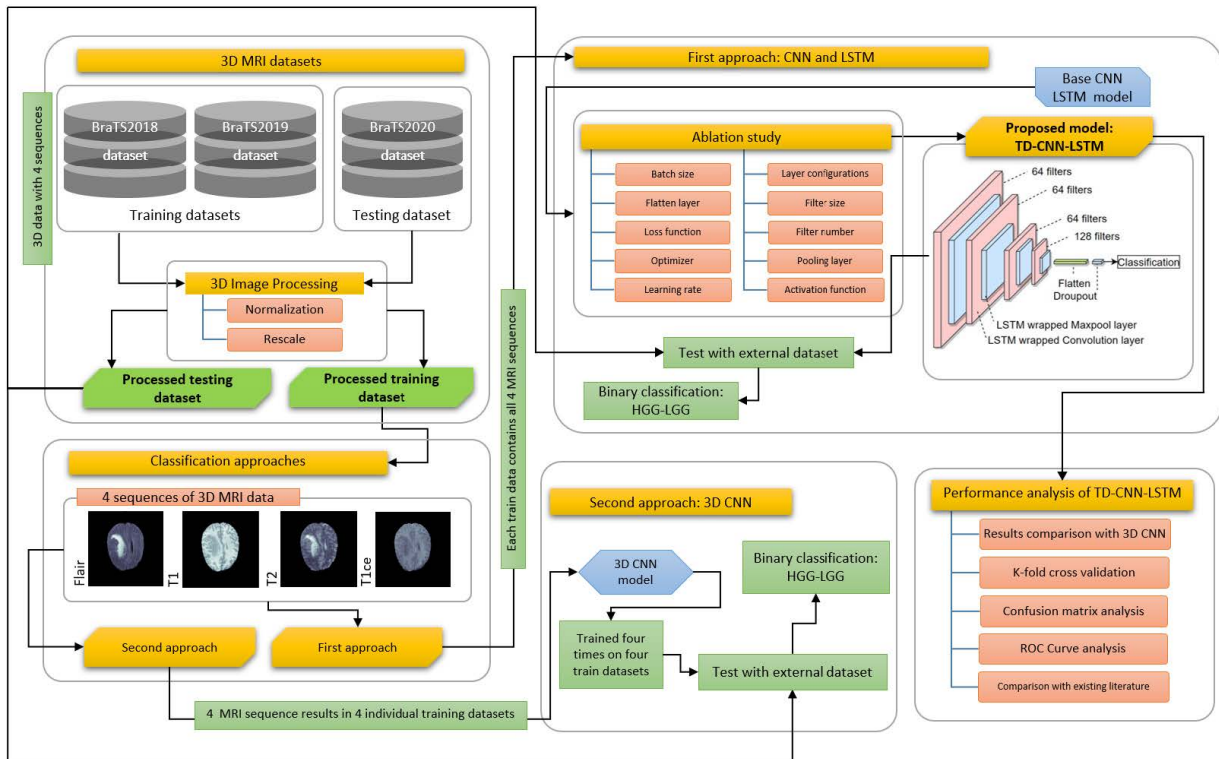
FIGURE 1. Workflow of the entire interpretability pipeline.

TABLE 1. Dataset description.

Dataset name	BraTS 2018	BraTS 2019	BraTS 2020
Type of image	3D brain MRI	3D brain MRI	3D brain MRI
Image format	NIfTI	NIfTI	NIfTI
Size of the images	240 x 240 x 150	240 x 240 x 150	240 x 240 x 150
No. of classes	Two	Two	Two
Name of classes	HGG, LGG	HGG, LGG	HGG, LGG
No. of useable subjects	282	331	365
No. of 3D images in each subject	5	5	5
Name of the four sequences and corresponding ROI of each subject	FLAIR, T1, T1ce, T2 and Seg	FLAIR, T1, T1ce, T2 and Seg	FLAIR, T1, T1ce, T2 and Seg
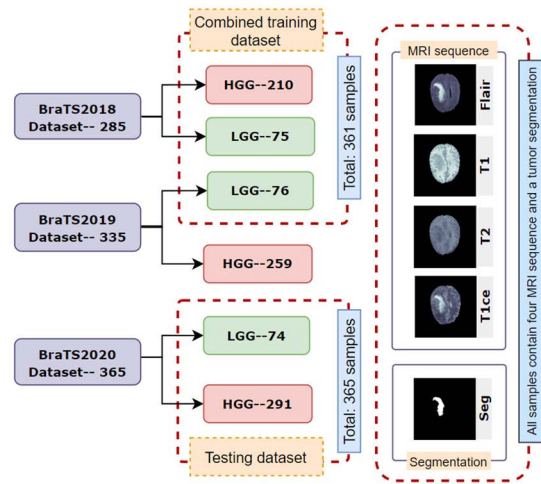
FIGURE 2. Dataset description with four imaging modalities and formation of training and testing datasets.

standard MRI sequences for a single patient of each of the datasets.

V. VISUALIZATION OF 3D IMAGE

A 2D image consists of pixels in single or multi-channel whereas 3D MR images comprise of 3D cubes or voxels. For reading, visualizing and writing 3Dneuroimaging data, NiBabel which is a Python package is commonly-used. Loading or reading a NIfTI file using the load() function of NiBabel ensures the encoding of all the information in the file where

each detail is known as 'attribute'. When visualizing a 3D image with NiBabel, it initializes a list in which it iterates over all the 155 slices of the 3D volume and whenever a volume is read, each slice is appended sequentially in the list. The number of voxels in a 3D image can be calculated using the following formula:

$$V_t = S_t \times H_s \times W_S \tag{1}$$

where, V_t is the total number of voxels in an image, S_t is the total number of 2D slices in an image, H_s is the height of each slice and W_s is the width of each slice.

As explained before, the number of slices in each image is 155, for our study the height and width of each slice is 240 (Fig. 3). The total number of voxels in each image can therefore be calculated as 155 X 240 X 240 = 8928000 voxels per image. To get a better understanding of the 3D sequences and their characteristics, a few images are visualized by converting them into 2D slices.
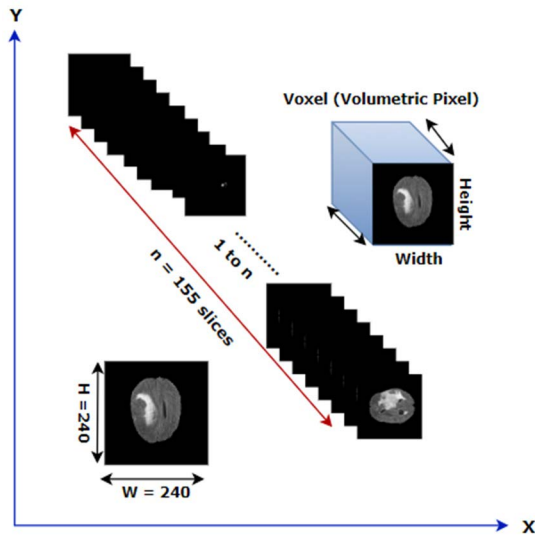
FIGURE 3. Visualization of 3D brain MRI.

A. PHYSICS OF MRI IMAGING SEQUENCES

MRI provides intense details of the brain, the spinal cord and the vascular anatomy, across three dimensions, axial, sagittal and coronal. Another advantage of MRI is that it can be used to observe blood flow and vascular malformations in brain tissues. However, all of the sequences contain certain and diverse features of brain tumor having different appearances which are analysed to determine the presence and grade of cancer. T1ce images are eminent in visualising blood brain barrier (BBB) interruption, while T2 and FLAIR images are recognized for distinguishing tumor boundaries and peritumoral edema [26]. The 3D FLAIR sequence ensures a good signal with fewer voxels, a high-spatial resolution with high signal to noise ratio (SNR), and good Cerebrospinal Fluid (CSF) suppression without of CSF flow artifacts [27]. The cystic constituents and edema are better portrayed with the T2- weighted than the T1- weighted MRI sequence. In comparison with Axial T2-weighted and Axial T2-weighted FLAIR images, Axial FLAIR image with nulling of the signal from the cerebrospinal fluid shows the metastatic lesions more visibly. FLAIR MRI sequence can aid in tumor detection efficiently by providing precise information on tumor infiltration [13]. However, T1 visualises a LGG tumor better. Due to hyper cellularity, the tumor region

appears as a hypo intense signal in T1- weighted images and as a hyper intense signal in T2- weighted images [28]. Moreover, T1 is recommended for the segmentation of tumor from non-affected brain cells. T1ce makes the tumor borders visible. With the T2 sequence, the edema (fluid) surrounding the tumor is more visible. FLAIR is suitable to distinguish the edema area from cerebrospinal fluid [29]. T2 and FLAIR MRI sequences are suitable for evaluating extracellular fluid in brain parenchyma [30].

1) FLAIR

In the imaging procedure of the FLAIR sequence, the motion of water molecules is suppressed [31]. The FLAIR sequence is produced using very long Time to Echo (TE) and Repetition Time (TR) times that ensure that abnormalities, for instance edematous tissues, remain brighter while the normal CSF appears darker (Fig. 4). Therefore, FLAIR is considered as a highly effective MRI sequence to distinguish the edema region from the CSF. In addition, white matter appears dark grey, the cortex light grey and fat appears light in a FLAIR MRI sequence.
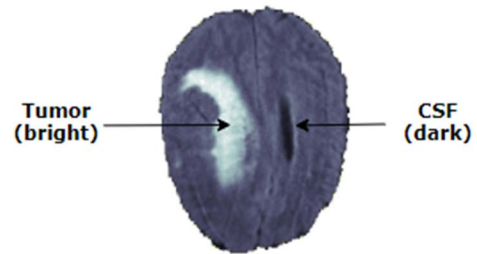

FIGURE 4. Axial view of FLAIR sequence.

2) T1

T1-weighted MRI sequences are generated with short TE and TR times which make the CSF darker (Fig. 5). T1 is a widely used sequence for the analysis of brain tumor patterns, as it allows an easy annotation of the healthy cells. In addition, white matter appears light, and the cortex grey and fat are bright in a T1-weighted MRI sequence.
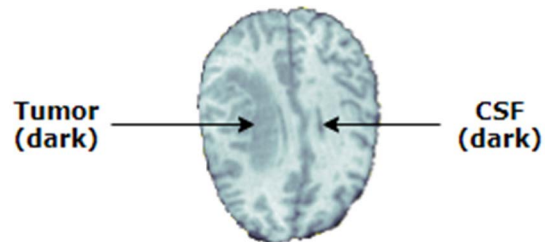

FIGURE 5. Axial view of T1 sequence.

3) T2

T2-weighted sequences are generated with long TE and TR times, making the CSF brighter (Fig. 6). In a T2 sequence, the
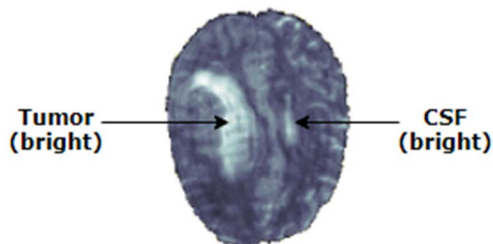
FIGURE 6. Axial view of T2 sequence.

edema region is brighter than in other MRI sequences. Like FLAIR, white matter appears dark grey, the cortex light grey and fat appears light in a T2-weighted MRI sequence.

4) T1CE

Finally, in T1ce sequences, the brain tumor borders appear brighter because of the accumulation of contrast agent. This is caused by the interruption of the blood-brain blockade in a proliferative brain tumor area (Fig. 7). Analyzing this sequence, the necrotic core and the active cell regions can be differentiated precisely. Some studies suggest that T1ce is more sensitive than with the other sequences [32]. Moreover, T1ce shows details of regional angiogenesis and the integrity of the blood–brain barrier in the tumor region.
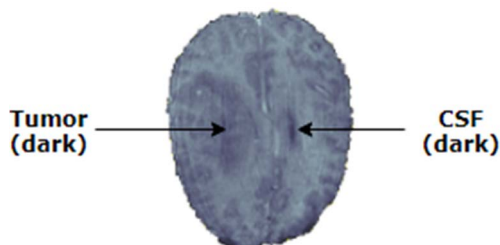

FIGURE 7. Axial view of T1ce sequence.

Table 2 summarizes the brain MRI sequences based on their characteristics and pictorial appearance.

TABLE 2. Comparison among four MRI sequences.

SEQUENCE	TR	TE	TUMOR	CSF
FLAIR	VERY LONG	VERY LONG	BRIGHT	DARK
T1	SHORT	SHORT	DARK	DARK
T2	LONG	LONG	BRIGHT	BRIGHT
T1CE	LONG	LONG	DARK	DARK

3D MRI images can be observed in three dimensions namely, Sagittal, Axial, and Coronal, which enables medical specialists to examine the shape of the tumors (Fig. 8).

VI. PREPROCESSING PHASE

Classification of brain tumors using 3D MRI scans is always challenging and computationally intensive due to the complex architecture of 3D images. Pre-processing operations are therefore required to enhance the model's performance [4].
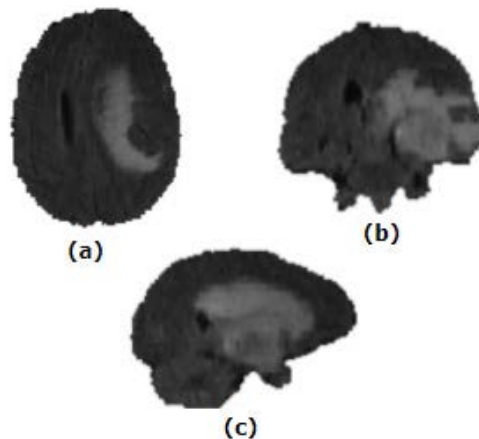

FIGURE 8. (a) Sagittal, (b) Axial, and (c) Coronal plans of FLAIR sequence.

In this research, two pre-processing steps intensity normalization and rescaling are performed to all of the four sequences of the dataset.

A. NORMALIZATION AND RESCALING

As intensities differ in MRI images, due to imaging procedures having different TE and TR, these 3D images need to be normalized. As scanning of patients is likely to be performed in different environments; intensity normalization plays an important role in brain tumor classification.

Data normalization ensures that each input parameter of a model has a similar data distribution transforming the floating-point feature values from their regular range into a new arbitrary standard range that is usually 0 to 1. The distribution of such pixels resembles a Gaussian curve. Data normalization is done when the approximate higher and lower pixel boundaries of an image are identified and data is approximately homogeneously distributed within that particular range. Min-max normalization [33], a widely-used technique, is adopted in our study to normalize the pixels. Algorithm 1 describes the min-max normalization process.

$$z_i = \frac{x_i - min(x)}{max(x) - min(x)} \qquad (2)$$

where, x refers to the pixel values ($x1,...,xn$) and z_i is the ith resultant normalized data.

The procedure for computing the normalization used in our paper can be explained using the pseudo-code of Algorithm 1:

Resizing 3D images accelerate the training process since most existing systems cannot handle the massive volume of 3D image data. After normalization, the dataset is rescaled to 128 × 128×32 voxels [34] because of the GPU memory limitations. Here, to ensure computational efficiency, only the middle 32 slices are utilized instead of all 155 slices of the brain and the original size of the 3D MRI image (240 × 240 pixels) was down-scaled to 128 × 128 pixels [35].

Algorithm 1 Pseudo-Code of min-max Normalization

   START
1. **DERIVE** mx = max(dataset) //highest pixel value
2. **DERIVE** mn = min(dataset) //lowest pixel value
3. L = len(dataset) //length of the dataset
4. **FOR** image in range of L:
5.   **CALCULATE** s = image – mn //subtraction of lowest pixel value from each image
6.   **CALCULATE** r = mx – mn //deriving range of dataset by deducing minimum pixel value from maximum pixel value
7.   **DERIVE** z = s / r // z denotes the normalized value
8. **ENDFOR**
   END

VII. PROPOSED APPROACHES

As stated, in this research, two fully automatic deep learning approaches are carried out for the classification of glioma brain tumors into HGG and LGG.

In the first approach, an integrated TD-CNN-LSTM network is employed which receives all the four sequences as a single input. In Fig. 9, an example of the sequences of a single HGG input, denoted as HGG [0], consisting of four 3D images is shown where all four 3D MRI images for a single subject is passed to the TD-CNN-LSTM network at once as an input. In the second approach, we propose a 3D CNN architecture to classify brain tumors, deriving deep feature interpretation from a single MRI sequence, where the model is trained four times for four MRI sequences respectively.
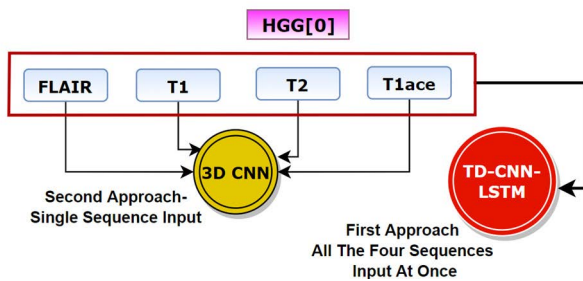
FIGURE 9. Visualization of two proposed approaches.

A. DATASET SPLIT

As described, we have employed three datasets for the interpretation of the study where BraTS 2018 and BraTS 2019 datasets are merged to train both of the models and BraTS 2020 dataset is left as the test dataset. In this regard, the preprocessed combined dataset is split into two sets: training dataset and the validation dataset with a ratio of 80:20. After splitting, the training set of subjects holds a total $289 \times 4 = 1156$ 3D images and the validation set of 72 subjects consisting of a total $71 \times 4 = 288$ 3D images and the external

test dataset contains 365 subjects composed of a total $365 \times 4 = 1460$ 3D images.

For second approach, the data split ratio is kept similar to the first approach with the difference that there are now four pairs of training validation and test datasets (one pair for each 3D MRI sequences).

B. FIRST APPROACH (TD-CNN-LSTM MODEL)

Classification using just one MRI sequence could potentially lead to substandard performance as all of the sequences contain certain characteristics of brain tumor. Individual MRI sequences comprise independent details whereas all MRI sequences combinedly may provide coherent radiomic and relevant clusters of features [36]. As explained in section V-A, each sequence contains important features and if all the sequences are fed into a deep learning network, the necessary parameters from all these MRI sequences can be extracted effectively. This technique can hypothetically lead to a better performance. Therefore, we integrate LSTM into the CNN architecture wrapped with TimeDistributed function with the aim of extracting the 3D context of slices in a sequential manner. As stated, the optimal model configuration is attained through ablation study by training the model several times with different configurations. Hence, we initially generate a base model, perform ablation study on it and obtain the optimal configuration of our proposed TD-CNN-LSTM model in terms of highest performance.

1) PROPOSED MODEL

CNNs combined with RNNs have been yielding promising results on several complex computer vision tasks. This approach provides effective solutions by detecting the hidden outlines in visual data using back propagation [37]. The proposed model is titled Time-Distributed_CNN_LSTM (TD-CNN-LSTM) as the architecture is deployed by combining CNN and LSTM and wrapped with TimeDistributed layer [38]. This model consists of four parts: 2D convolution for feature extraction, pooling layer for feature reduction over the sequences, an LSTM layer, and a final classification layer [39]. In this hybrid network, CNN handles the spatial dependencies while LSTM deals with the temporal dependencies. Contrary to CNN, LSTM is proficient in sequence processing by adaptively apprehending long-term dependences and nonlinear dynamics of sequential data. LSTMs are powerful in analyzing a dataset of a sequential nature but configuring LSTM with CNN is complex. To overcome this complexity, a TimeDistibuted function is used to configure the input shape and proceed to convolution and pooling layers.

A TimeDistibuted layer typically adds an additional dimension to the input shape of the corresponding argument-layer. This enables the CNN to receive multiple frames as one input. TimeDistributed layer behaves as a layer wrapper, wrapping the CNN model itself when employed to an input tensor. This wrapper allows the addition of a layer to each sequential slice of input data where inputs can be in 3D.

In our experiment, this function is used on each convolution, pooling, and input layer. Here, the input dimension is denoted as input (batch_size, frame, width, height, channel) where batch_size is the pre-defined batch size of the model, the frame is the number of frames to input at once, width is the width of the image, height is the height of the image and channel is the number of slices of the image.

In our study, the batch_size is 'None' which indicates that the pre- defined batch size of the model that will be applied while training the model. As we have four 3D images for each subject, the number of frames is equal to four. The last three parameters (height, width, channel) are 128, 128 and 32 respectively. As we have 32 slices for a 3D image, 32 is denoted as the number of channels for each 3D data. TimeDistributed convolutions are convolutions directed through a TimeDistributed wrapper. This permits the application of any layer to every temporal slice (or frame) of the input individually. In this study, the temporal frames for the MRI data are derived from the 3D volumes [40]. Generally, the TimeDistributed approach is useful when, solving a computer vision task which requires a complex model.

LSTM is a variant of RNN introduced to capture global sequence dependencies of the input data. The spatio-temporal parameters processed by LSTM help the model in identifying hidden outlines in challenging frame-to-frame sequences [37]. In this process, the basic features extracted by the CNN are passed to the LSTM layer as inputs to ascertain the temporal dependencies. After that, LSTM layer receives a sequence of CNN outputs as input accruing the temporal dependencies of the frames of all four MRI sequences [41]. An LSTM memory cell comprises of three components the forget gate, the input gate, and the output gate. The LSTM operation procedure consists of the following steps [42]:

- The output value of the last instant and the input value of the present instant is passed as input into the forget gate. The output of the forget gate is calculated using (3):

$$f_t = \sigma \left(W_f \cdot [h_{t-1}, x_t] + b_f \right) \quad (3)$$

where, the range of f_t is (0, 1), W_f represents the weight of the forget gate, b_f represents the bias of the forget gate, x_t represents the input value of the present time, and finally h_{t-1} denotes the output value of the last instant.

- The output value of the last instant and the input value of the present is passed as input into the input gate. The output value and candidate cell state of the input gate are calculated using (3) and (4):

$$i_t = \sigma \left(W_i \cdot [h_{t-1}, x_t] + b_i \right) \quad (4)$$
$$\tilde{C}_t = \tanh \left(W_c \cdot [h_{t-1}, x_t] + b_c \right) \quad (5)$$

where, the value range of i_t is (0,1), W_i is the weight of the input gate, b_i is the bias of the input gate, W_c is the weight of the candidate input gate, and b_c is the bias of the candidate input gate.

- The current cell state is updated using (6):

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (6)$$

where the values range of C_t is (0, 1).

- The output and value of h_{t-1} and x_t are inputted into the output gate at time t. The output o_t of the output gate is acquired as follows using (7):

$$o_t = \sigma \left(W_o [h_{t-1}, x_t] + b_o \right) \quad (7)$$

where, the value range of o_t is (0,1), W_o is the weight of the output gate, and b_o is the bias of the output gate.

- The output value of LSTM is acquired by computing the output of the output gate and the state of the cell, applying following formula (8):

$$h_t = o_t * \tanh (C_t) \quad (8)$$

Like RNN, LSTM has time steps and "MEMORY" for each time step. In our study, time steps represent the MRI image sequence. The diagram in Fig. 10 explains the mechanism of LSTM cell at time step t.
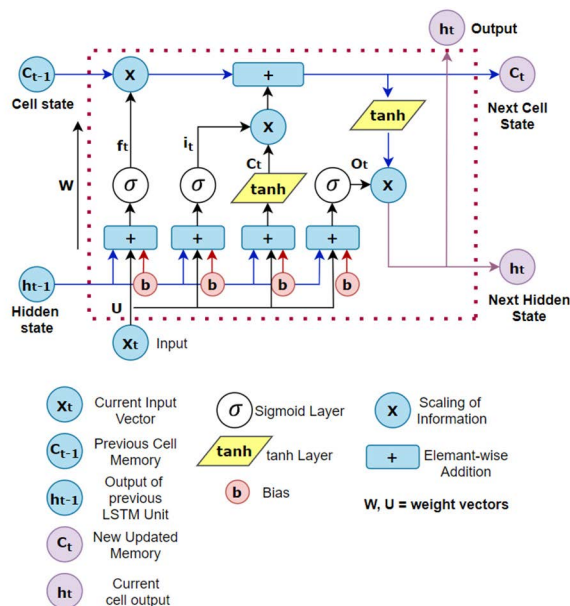

FIGURE 10. Operation procedure of LSTM.

The Forget gate (f), Input gate (I) and Output Gate(O) employ the Sigmoid activation function and candidate layer uses *Tanh* as the activation function. The data can be added, removed or updated to the cell state through sigmoid gates [43]. The method of recognizing and eliminating data is determined by the sigmoid function which takes the output of the last LSTM unit (h_{t-1}) at time t − 1 and the present input (x_t) at time t. The sigmoid function decides which parts of the previous output should be eliminated by the forget gate f_t. After inputting the weights vector.dot (U) and previous hidden state.dot (W), these gates concatenate these and apply the activation function, subsequently producing vectors f, C, I, O in a range of 0 to 1 for Sigmoid and

-1 to 1 for Tanh for every time step. C_{t-1} and C_t denote the cell states at a particular time $t-1$ and t, respectively, and b represents the bias. Memory state C of LSTM cell is where the memory or context of input is stored. This stored information can be modified in different time steps. The output values (h_t) are determined based on the output cell state (o_t). A sigmoid layer determines which portion of the cell state will be used for the output. Next, the output of the sigmoid gate o_t) is multiplied with the values produced by the tanh layer from the cell state (C_t).

In our study, the information for all MRI sequences is stored in memory state C. The final output is based on analyzing all features of the four MRI sequences (Fig. 11).
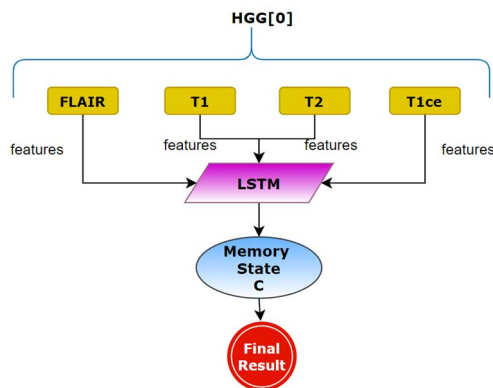

FIGURE 11. Operation of LSTM on all four 3D MRI sequences.

2) BASE MODEL
We initialize our experiment from a base CNN model containing two convolutional layers each followed by a maxpool layer where each layer is wrapped with a TimeDistributed function. To begin with, the network consists of 5×5 sized convolutional kernels, the number of kernels in each convolutional layers is set to 32, 'ReLU' is chosen as activation function, 'softmax' is utilized as the activation function for the final layer, 'categorical_crossentropy' is chosen for loss function and lastly, optimizer Adam with a learning rate of 0.001 was chosen. The epoch number for training is set to 100 epochs.

Batch size is set to 16 while training the model. The input shape of the 3D images is $4 \times 128 \times 128 \times 32$ where 4 denoted the number of sequences in a 3D data, 128×128 denotes the height $\times$ width of each 3D slices and 32 refers to depth (number of 2D slices in a 3D image) of each 3D image.

3) ABLATION STUDY
The idea of 'ablation' is based on animal experiments where nerves in particular regions of the brain were eradicated in a selective manner, with the purpose of investigating the behavioural effects of this destruction and thereby obtaining a better understanding of the function of these regions. The term ''ablation study'' is nowadays mostly used in the context

of neural networks [44] to observe the model's performance by studying the effect of altering some components [45]. Therefore, the effects of ablation are investigated on our base model with ten study cases by changing different parameters, the number of layers and the filters to evaluate the impact of these parameters in the proposed architecture [46]. In this way, with the alteration of different components in every study case, the optimal component is chosen and proceeds further to the next study cases based on the highest accuracy. Moreover, the issue of time complexity is taken into account while not compromising accuracy. Thus, after completing all the study cases, the best-performed configuration of our proposed TD-CNN-LSTM model can be achieved with highest accuracy and lowest time complexity. The results are described in section VIII-B.

4) TD-CNN-LSTM ARCHITECTURE
After performing ablation study of ten cases on the base model, the optimal model architecture is generated for which the highest performance is recorded with the lowest possible computational complexity. In the architecture illustrated in Fig. 12, the input shape is a five-dimensional data vector (None, 4, 128, 128, 32), where 4 denotes the five MRI sequences to be inputted simultaneously and 128, 128, 32 is the input dimension of every MRI sequence.

The network comprises two blocks containing a total of 15 layers: four convolutional layers and four maxpool layer four batch normalization layers, a flatten layer, and an LSTM layer followed by a dense layer. Each of the layers is wrapped with a TimeDistributed function that allows each layer to be applied to all temporal slices of the 3D MRI sequences. Each block consists of a 2×2 kernel sized convolutional layer followed by a 2×2 maxpool layer. First, the input layer is fed to Block-1 containing the first convolution layer of 64 filters. This extracts the features from the 3D input image and generates a feature map. Then a maxpool layer is employed to scale down the size of the resulting feature map (4, 128, 128, 64 to 4, 64, 64, 64) and a batch normalization layer normalizes the output of maxpool layer. In Block 2, the second convolution layer of 64 filters extracts more features from the Block-1 resultant feature map. The generated feature map is again scaled down (4, 64, 64, 64 to 4, 32, 32, 64) and again a batch normalization layer normalizes the feature maps. Similarly Block-3 and Block-4 produces more feature maps and scale down the feature maps even more and lessen computational time. Resultant feature maps after Block-4 are of dimensions 4, 8, 8, 128. Block-4 is followed by a Flatten layer that generates a 1D vector from the pooled feature map of the previous layers. The output feature vector is then entered into the LSTM layer which has 128 hidden units. As dropout factor of 0.5 to 0.8 is considered as standard values, a dropout of factor 0.5 is also included in the LSTM layer as 0.5 value yields the best possible outcome. The output layer of this architecture is equipped with a Fully Connected (FC) layer with sigmoid activation function that takes the resulting 1D tensor and gives an output 0 or 1 which respectively
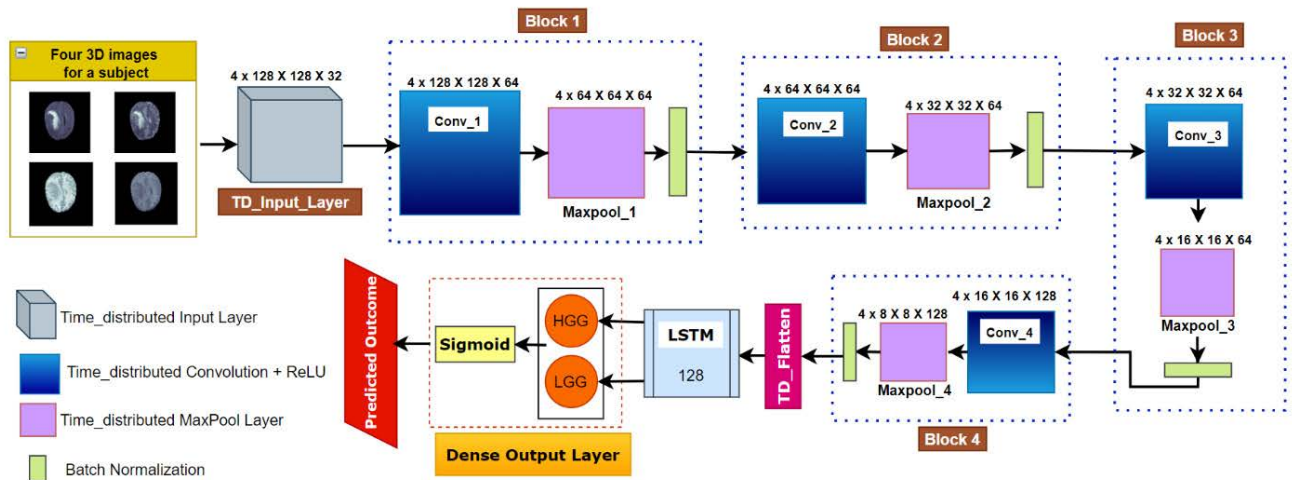
FIGURE 12. TD-CNN-LSTM model architecture.

indicates HGG or LGG. This function takes any real-valued input (x) and generates outputs with a value between zero and one. In this process, a 3D image will be considered as a series of 2D slices that can be represented as X = {x1, x2, · · ·, xn} where n denotes the number of slices. From these slices, the convolution layer extracts the required features, and the maxpool layers shrink the individual slices. Finally, the sequence is condensed to P = {p1, p2, · · ·, pm} where p is the feature vector of 2D slices.

The proposed model is trained with binary cross-entropy as the loss function, using the following equation [5]:

$$L_C = -\frac{1}{n} \sum_{i=1}^{n} \{y_i \log(f_i) + (1 - y_i) \log(1 - f_i)\} \quad (9)$$

where, n denotes the number of samples, y_i represents the true label of a particular sample and f_i denotes its predicted label.

C. SECOND APPROACH (3D CNN MODEL)

As explained earlier, to evaluate the effectiveness of our proposed TD-CNN-LSTM model, we have trained a 3D CNN model having 3D convolution and 3D maxpool layer and compared the performance. The model is trained four times on these four training sets of 3D images, leading to different results for the 3D MRIs: FLAIR, T1, T2 and T1ce.

The 3D CNN provides a comprehensive feature map, analysing the volumetric spatial information and integrating nonlinear 3D contextual information [16]. A 3D network detects 2D structures such as edges and corners in 3D manner [47]. The proposed CNN architecture is constructed of four layers: convolutional, sub-sampling, batch normalization (BN) and FC layers equipped with a sigmoid activation function for HGG and LGG classification. The input layer of the proposed 3D model receives a multi-channel brain MRI which can be denoted as [5]:

$$I \in \mathbb{R}^{m \times w \times h} \quad (10)$$

where, m denotes the number of channels and w and h represent the width and height of the channels respectively.

Compared to a 2D model, 3D CNN is computationally intensive and requires more memory. Each convolutional layer of our proposed network is structured with a 3×3 kernel which enables the model to perform faster as smaller size convolutional kernels yield improved efficiency due to the lesser number of trainable parameters. The function 3D convolutional layer can be stated as [48]:

$$u_{kj}^l(x, y, z) = \sum_{\delta_x} \sum_{\delta_y} \sum_{\delta_z} F_k^{l-1}(x + \delta_x, y + \delta_y, z + \delta_z) \times W_{kj}^l(\delta_x, \delta_y, \delta_z) \quad (11)$$

here, voxel positions for any given 3D image are denoted as x, y and z, respectively, the weight of j-th 3D kernel, connecting the k-th feature maps of the l-1 layer and the j-th feature maps in the layer l, is represented by $W_{kj}^l(\delta_x, \delta_y, \delta_z)$, the k-th feature maps in the layer l-1 is F_k^{l-1}, and the kernel sizes corresponding to the x, y and z are δ_x, δ_y, and δ_z respectively The kernel filter's convolutional response is $u_{kj}^l(x, y, z)$.

Rectified linear unit (ReLU) which is a commonly used activation function for deep CNN due to its computational efficiency and reduced likelihood of vanishing gradients is employed in each convolutional layer. The function can be defined as [48]:

$$F_j^l(x, y, z) = max\left(0, b_j^l + \sum_k u_{jk}^l(x, y, z)\right) \quad (12)$$

where, b_j^l represents the bias term of the j-th feature map of l-th layer's, $F_j^l(x, y, z)$ is attained by adding the response maps of the j-th 3D feature map's individual convolution filters.

In a 3D max-pooling procedure, the maximum value in the cubic region is the input to the consequent feature volume. The equation can be stated as [49]:

$$h_{xy}^l = max_{i=0,...,a, j=0...,s} h_{(x+i)(y+j)}^{l-1} \quad (13)$$

where, h^l = activation of layer l.

A BN layer is utilized to condense preliminary covariate shift. The FC layer classifies the data, performing nonlinear operations on the generated parameters to obtain the output [49].

$$z^l = W^l h^{l-1} \tag{14}$$

where, h^l = activation of layer l and W^l = learnable parameter

The BN layer is generally intended to expedite the training progression of the CNN model. It normalizes the output of each layer.

$$y_i = \gamma x_i' + \beta \tag{15}$$

$$x_i' = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \tag{16}$$

$$u_B = \frac{1}{m} \sum_{i=1}^m x_i \tag{17}$$

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2 \tag{18}$$

1) 3D CNN ARCHITECTURE

The 3D CNN architecture is comprised of five Blocks and a depth size of 16. Each block has a 3D convolutional layer with a kernel size of 2×2, followed by a 2×2 maxpool layer and a BN layer. The input layer can take 3D images of size 128,128,32,1 as input. The 3D CNN architecture is illustrated in Fig. 13.

The input (128, 128, 32, 32) goes through Block-1 where the first convolutional layer with 32 filters extracts the features from the 3D image and generates a feature map. The feature map is then scaled down with a maxpool layer (126, 126, 62, 32 to 62, 62, 30, 32) followed by a BNlayer. Block-2, Block-3 and Block-4 have convolutional layers with 64, 128 and 256 filters respectively. This is followed by a maxpool layer and a BNlayer, as in Block-1. Each of these blocks extracts features from the previous feature map and then scales down the feature map further. The output feature map from Block-4 is of size 8, 8, 3, 256 which goes to a global average pooling layer which applies average pooling to all spatial dimensions. The output from this layer goes through a dropout layer with a factor of 0.5 and a FC layer of 512 neurons. The output layer consists of a FC layer equipped with a Sigmoid function that classifies the output into 0 or 1 which represents HGG and LGG respectively. The model is trained with 'binary cross-entropy' as the loss function.

VIII. RESULT ANALYSIS

In this section the effectiveness of both binary classifier models: TD-CNN-LSTM and 3D CNN are evaluated by employing several evaluation metrics (both model classification metrics and model error metrics). As explained, proposed TD-CNN-LSTM model is achieved by employing extensive ablation studies. The 3D CNN model is trained and validated separately on the FLAIR dataset, the T1 dataset, the T1ce dataset and the T2 dataset in order to justify the effectiveness of TD-CNN-LSTM. Evaluation metrics are calculated

for both of the models to evaluate the performance of the classifiers in predicting classes HGG and LGG successfully. In this section, the results of ten cases of ablation study, results of TD-CNN-LSTM and 3D CNN model, and results of K-fold cross validations are discussed. At the end of the section, a performance comparison of previous studies with our research is presented.

A. EVALUATION METRICS

Often, while doing binary classification, a single metric (accuracy score) is used to evaluate the performance of the classifier. This approach is flawed, however, as the classifier might yield poor results upon evaluation with other evaluation metrics. To properly evaluate the effectiveness of a binary classifier, several evaluation metrics are required. Keeping this in mind, both the TD-CNN-LSTM and 3D CNN models are evaluated using different classification metrics: precision (Pre), Recall, (Rec), Specificity (Spe), F1-score (F1) and accuracy (ACC). Two model error metrics: MAE and RMSE are also calculated. AUC value is calculated by generating Receiver operating characteristic (ROC) curves for all models [44].

For the computation of all evaluation metrics, a confusion matrix for all trained models is generated for all trained models. This provides the values of true positive (TP), true negative (TN), false positive (FP), false negative (FN) that which are required for computing all the evaluation metrics used in this section.

$$ACC = \frac{TP + TN}{(TP + TN + FP + FN)} \tag{19}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{20}$$

$$Specificity = \frac{TN}{(TN + FP)} \tag{21}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{22}$$

$$F_1 = 2 * \frac{Precision * Recall}{(Precision + Recall)} \tag{23}$$

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \tag{24}$$

$$RMSE = \left[\sum_{j=1}^N \left(d_{f_i} - d_d \right)^2 / N \right]^{\frac{1}{2}} \tag{25}$$

B. RESULT OF ABLATION STUDY

Ablation study concludes with various changes of elements in the base CNN model and the results are recorded. Time complexity [50], and test accuracy are evaluated for each experimental configuration. Time complexity in theory can be defined as [51]:

$$O = \left\{ \sum_{j=1}^k n_{j-1} \cdot s_w \cdot s_h \cdot n_j \cdot m_w \cdot m_h \right\} \tag{26}$$

here, j indicates the index number of convolutional layers, and k refers to the total number of convolutional layers, n_{j-1} indicates the number of kernel or input channels in the
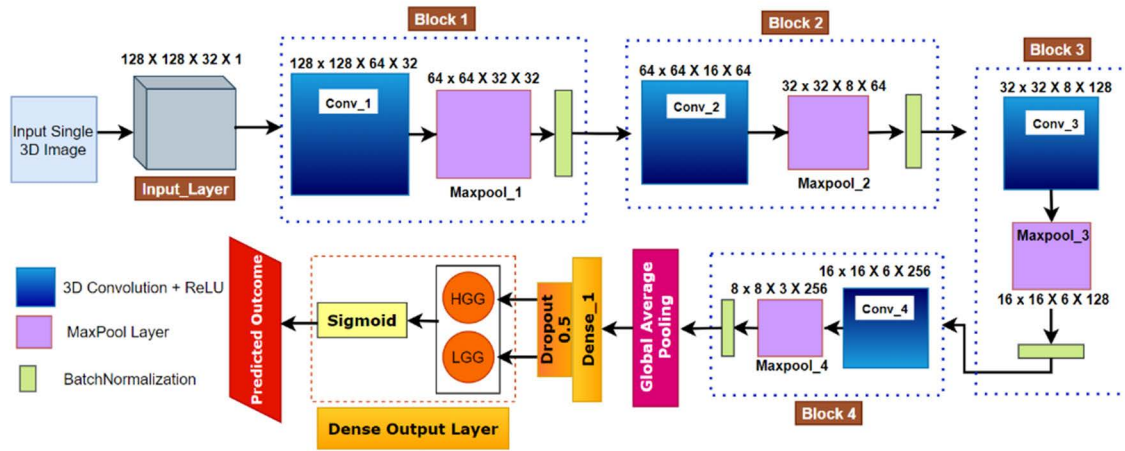
FIGURE 13. 3D CNN model architecture.

$j - 1^{th}$ convolutional layer, and n_j refers to the number of kernels in the jth layer, s_w and s_h indicates width and height of the kernels on j^{th} layer, and m_w and m_h indicates the width and height of the generated feature map correspondingly.

Results of the ablation studies conducted with the testing dataset (external) are presented in Table 3 and Table 4 where Table 3 concludes all the results relating to the model's layer configurations and activation functions and Table 4 showcases the results of tuning hyper-parameters, loss function and flatten layer.

Case study 1: changing convolution and maxpool layer

Here, the initial configuration of the base model mentioned above is kept as it is while changing the number of convolution and maxpool layer. Table 3 shows the performance for different configurations of the altered model architecture. Best performance is achieved from configuration 3 and 4 (Table 3) with a test accuracy of 95.06%. But regarding time complexity, configuration 3 containing four pairs of convolutional and maxpool layers shows a lower time complexity (116.32 million) than configurations 4. Taking this into consideration, configuration 2 was selected for the rest of the ablation case studies.

Case study 2: changing filter size

In this case study, experimentations with various kernel sizes of 3 x 3, 2 x 2 and 5 x 5 are carried out to observe the performance [52]. It is evident that changing filter size does not affect much on the overall performance (Table 3). However, the highest test accuracy, 95.91%, is acquired by employing the kernel size 2 x 2. Thus, configuration no 2 is chosen for further ablation case studies.

Case study 3: changing the number of filters

Initially, we started with a constant number of kernels [53] for all the four convolution layers (32, 32, 32 and 32). Later, number of features is increased to 64 and significant improvement in performance is found. However, we anticipated that gradually increasing might be a better approach. This is represented in configuration 3, 4 and 5 (Table 3). It is observed that

configuration 5 with filter numbers 64, 64, 64 and 128 for the four convolutional layers achieved the highest performance with a test accuracy of 97.75%. Therefore, we move forward with configuration 5.

Case study 4: changing the type of pooling layer

Two pooling layers, maxpool and average pool, are evaluated [50] where both pooling layers gained the same highest accuracy of 97.75% (Table 3). In this regard, Max pooling layer is chosen for further ablation studies.

Case study 5: changing the activation function

As various activation functions can have an impact on the performance of a model, choosing an optimal activation function is relevant in building an optimal model. Six activation functions, Linear, PReLU, ReLU, Leaky ReLU, Tanh and Exponential Linear Units (ELU) [54] are experimented with. It is found that the linear activation function performs best with a test accuracy of 98.33% (Table 3). This activation function was chosen for further ablation studies.

Case study 6: changing batch size

Batch size denotes the number of images utilized on each epoch to train the model. A larger batch size may result in the model taking a long time to accomplish convergence while batch sizes that are overly small can cause poor performance. Moreover, performance differs for different batch sizes for medical images due to the complex structure of medical images [55]. In this case, experimentation with four batch sizes are conducted and found that both the batch sizes of 16 and 32 achieved test accuracy of 98.33% similar to the previous case study (Table 4). In this regard, batch size of 16 is chosen for further ablation studies.

Case study 7: changing flatten layer

Multi-dimensional output of previous layers is taken by Flatten layer and results in a one-dimensional tensor. Experimentation with Global Max pooling and Global Average pooling are conducted, and results indicate that the previously used flatten layer yields the highest test accuracy of 98.33% (Table 4) while maintaining the lowest training time.

TABLE 3. Ablation study regarding layer configurations and activation functions.

Case study 1: changing convolution and maxpool layer					
Configuration No.	No. of convolution layer	No. of pooling layer	Time complexity (Million)	Test accuracy (%)	Finding
1	2	2	108.13M	92.50	Lowest accuracy
2	3	3	114.68M	95.06	Modest accuracy
3	**4**	**4**	**116.32M**	**95.45**	**Highest accuracy**
4	5	5	116.73M	95.45	Highest accuracy
5	6	6	116.83M	94.18	Modest accuracy
Case study 2: changing filter size					
Configuration No.	Filter size	Time complexity (Million)	Test accuracy (%)		Finding
1	3 x 3	41.87M	95.56		Accuracy improved
2	**2 x 2**	**18.61M**	**95.91**		**Highest accuracy**
3	5 x 5	116.32M	95.45		Previous accuracy
Case study 3: changing the number of filter					
Configuration No.	No. of kernel	Time complexity (Million)	Test accuracy (%)		Finding
1	32→32→32→32	18.61M	95.91		Previous accuracy
2	64→64→64→64	73.40M	97.64		Accuracy improved
3	32→32→64→64	20.44M	97.26		Accuracy improved
3	32→64→64→64	39.32M	97.47		Accuracy improved
4	**64→64→64→128**	**74.44M**	**97.75**		**Highest accuracy**
Case study 4: changing type of pooling layer					
Configuration No.	Type of pooling layer	Time complexity (Million)	Test accuracy (%)		Finding
1	**Max**	**74.44M**	**97.75**		**Highest accuracy**
2	Average	74.44M	97.75		Highest accuracy
Case study 5: changing activation function					
Configuration No.	Activation function	Time complexity (Million)	Test accuracy (%)		Finding
1	**Linear**	**74.44M**	**98.33**		**Highest Accuracy**
	PReLU	74.44M	98.17		Accuracy improved
2	Relu	74.44M	97.75		Previous accuracy
3	Leaky ReLu	74.44M	98.06		Accuracy improved
4	Tanh	74.44M	95.18		Accuracy dropped
5	ELU	74.44M	97.25		Accuracy dropped

Case study 8: Changing Loss Functions

Experimentation regarding various loss functions, namely Binary Cross-entropy, Categorical Cross-entropy, Mean Squared Error, Mean Absolute Error, Mean Squared Logarithmic Error and Kullback Leibler Divergence are carried out in order to select the appropriate loss function for the optimal model. While equipped with Binary Cross-entropy the model had a 98.68% (Table 4) test accuracy which is the highest result. Hence this is chosen.

Case study 9: Changing Optimizer

Experimentation with different optimizers namely Adam, Nadam, SGD, Adamax and RMSprop are carried out in order to identify the optimal optimizer. In this case the learning rates are set to 0.001. The best test accuracy of 98.84% (Table 4) was achieved with the Nadam optimizer. We select the Nadam optimizer for further ablation study.

Case study 10: Changing Learning rate

Experimentation with different learning rates 0.01, 0.001, and 0.0001 was conducted. The best test accuracy of 98.90% (Table 4) was recorded with the learning rate 0.0001 and the Nadam optimizer.

Visual representation of the gradual performance boost with different ablation study cases is shown in Fig. 14 for better understanding.

After completing the ablation studies on the base model, the proposed TD-CNN-LSTM model is acquired where a significant improvement in classification accuracy is observed. Configuration of TD-CNN-LSTM model is presented in Table 5.

C. PERFORMANCE COMPARISON WITH 3D CNN MODEL

Comparison of two approaches regarding various accuracy and error matrics is conducted in this section. In this regard, the training configuration of 3D CNN model is kept similar to TD-CNN-LSTM model where, Nadam optimizer is used with a learning rate of 0.0001, batch size of 16 and number of epochs 150. Classification results acquired from the TD-CNN-LSTM and 3D CNN models, which are trained on their respective datasets and tested on the respective external testing dataset are shown in Table 6. The results of our proposed model are bold on the tables.

TABLE 4. Ablation study regarding model hyper-parameters, loss function and flatten layer.

Case study 6: changing batch size				
Configuration No.	Batch size	Time complexity (Million)	Test accuracy (%)	Finding
1	**16**	**74.44M**	**98.33**	**Highest accuracy**
2	32	74.44M	98.33	Highest accuracy
3	64	74.44M	98.27	Near highest accuracy
Case study 7: changing flatten layer				
Configuration No.	Flatten layer type	Time complexity (Million)	Test accuracy (%)	Finding
1	**Flatten**	**74.44M**	**98.33**	**Highest accuracy**
2	Global Max pooling	74.44M	98.17	Accuracy dropped
3	Global Average pooling	74.44M	98.08	Accuracy dropped
Case study 8: Changing Loss Functions				
Configuration No.	Loss Function	Time complexity (Million)	Test accuracy (%)	Finding
1	**Binary Crossentropy**	**74.44M**	**98.72**	**Highest accuracy**
2	Categorical Crossentropy	74.44M	98.33	Previous accuracy
3	Mean Squared Error	74.44M	98.17	Accuracy dropped
4	Mean absolute error	74.44M	98.45	Accuracy improved
5	Mean squared logarithmic error	74.44M	96.72	Accuracy dropped
6	Kullback Leibler Divergence	74.44M	96.48	Accuracy dropped
Case study 9: Changing Optimizer				
Configuration No.	Optimizer	Time complexity (Million)	Test accuracy (%)	Finding
1	Adam	74.44M	98.72	Previous accuracy
2	**Nadam**	**74.44M**	**98.84**	**Highest accuracy**
3	SGD	74.44M	92.68	Accuracy dropped
4	Adamax	74.44M	95.75	Accuracy dropped
5	RMSprop	74.44M	90.82	Accuracy dropped
Case study 10: Changing Learning rate				
Configuration No.	Learning rate	Time complexity (Million)	Test accuracy (%)	Finding
1	0.01	74.44M	98.17	Accuracy dropped
3	0.001	74.44M	98.81	Previous accuracy
5	**0.0001**	**74.44M**	**98.90**	**Accuracy improved**

TABLE 5. In depth details of TD-CNN-LSTM model configurations.

MODEL	TD-CNN-LSTM
DATASET	COMBINED (FLAIR, T1, T1CE, T2)
OPTIMIZER	NADAM
LEARNING RATE	0.0001
LAYER ACTIVATION FUNCTION	LINEAR
LOSS FUNCTION	BINARY CROSSENTROPY
BATCH SIZE	16
EPOCH (HIGHEST ACCURACY)	79
TIME PER EPOCH	AROUND 2 SECONDS
CLASSIFICATION LAYER ACTIVATION FUNCTION	SIGMOID
DEPTH	11

TABLE 6. Training accuracy (T_Acc), Validation loss (Val_loss), validation accuracy (Val_Acc), test loss (Te_Loss) and test accuracy (Te_Acc) for all trained models.

MODEL	DATASET	T_ACC	V_LOSS	V_ACC	TE_LOSS	TE_ACC
3D CNN	FLAIR	98.41	0.07	94.55	0.07	94.86
3D CNN	T1	98.66	0.06	97.56	0.06	96.31
3D CNN	T1CE	98.41	0.07	94.52	0.07	94.60
3D CNN	T2	92.84	0.50	89.73	0.05	89.85
TD-CNN-LSTM	**ALL**	**99.92**	**0.03**	**98.66**	**0.03**	**98.90**

In Table 7 training accuracy is referred to as T_Acc, validation accuracy as V_Acc, validation loss as V_loss and test accuracy as T_Acc and test loss as t_loss. It can be observed that the TD-CNN-LSTM model trained on the combined dataset had the highest test accuracy of 98.90% and test loss of 0.03%. Furthermore, it has the highest training and validation accuracy scores of 99.92% and 98.66% respectively, among all configured models. It also yields the lowest validation loss of 0.03. 3D CNN trained on T1 dataset
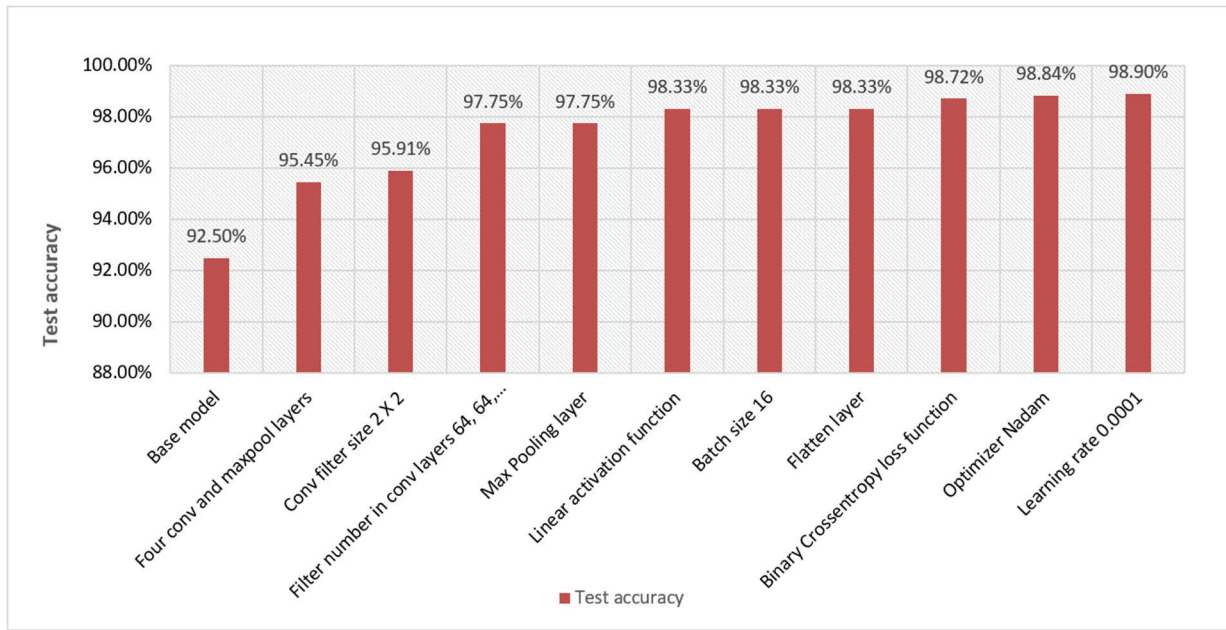
FIGURE 14. Result analysis of ablation study.

TABLE 7. Specificity (Spe), Precision (Pre), Recall (Rec), F1 score (F1) and AUC values for all trained models.

MODEL	DATASET	SPE	PRE	REC	F1	AUC
3D CNN	FLAIR	95.97	94.45	92.62	92.77	94.73
3D CNN	T1	98.23	96.54	96.22	98.72	96.45
3D CNN	T1CE	95.75	95.65	93.71	94.16	94.68
3D CNN	T2	89.86	90.02	88.58	88.25	89.80
TD-CNN-LSTM	**ALL**	**99.15**	**98.95**	**98.78**	**98.83**	**99.04**

FIGURE 15. MAE of all five configuration (3D CNN model trained on 4 datasets and TD-CNN-LSTM model trained on a combined datasets).

demonstrated satisfactory accuracy among other sequences. It achieved a training accuracy of 98.66% and a validation accuracy of 97.56% with a test accuracy of 96.31%. The T2 dataset trained on 3D CNN model resulted in the lowest performance with test accuracies below 89.9% and the lowest validation accuracies below 90%.

In Table 7 Specificity is referred as Spe, Precision as Pre, Recall as Rec, F1 score as F1 and Area under the curve value as AUC. Among all six configured models, TD-CNN-LSTM trained on the combined dataset had the most correct predictions with an AUC of 99.04%, Specificity of 99.15%, Precision of 98.95%, Recall of 98.78% and F1 score of 98.83%. The 3D CNN model showed moderate performance with the T1dataset, with AUC score of 98.25% and Specificity of 98.23%. It is evident that the combined dataset in combination with the TD-CNN-LSTM model yielded the most promising results over 3D CNN approach.
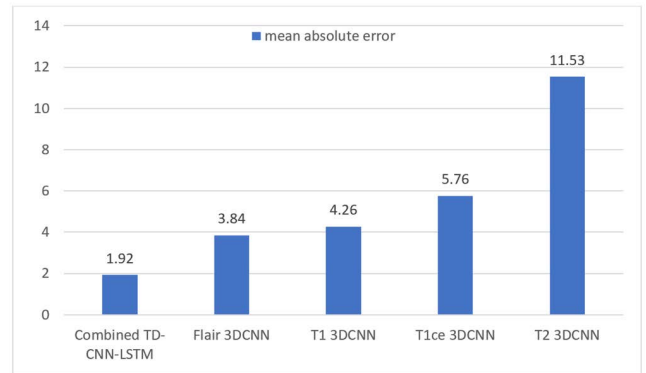
Fig. 15 portraits the MAE [46] for our proposed model and 3D CNN model. For *MAE*, the smaller the error value the better the performance of the classifier. It can be observed (Fig. 15) that the TD-CNN-LSTM model trained on the combined dataset had the smallest MAE (1.92). Datasets FLAIR, T1 and T1ce showed moderate performance in this regard having MAE values below 6.

The RMSE of all the datasets is shown in Fig. 16. Like MAE, the lower the value the better the classifier performs. It can be observed from Fig. 16 that the TD-CNN-LSTM model with the combined dataset had the lowest RMSE value (13.86). 3D CNN model performed moderately with the FLAIR, T1 and T1ce datasets.
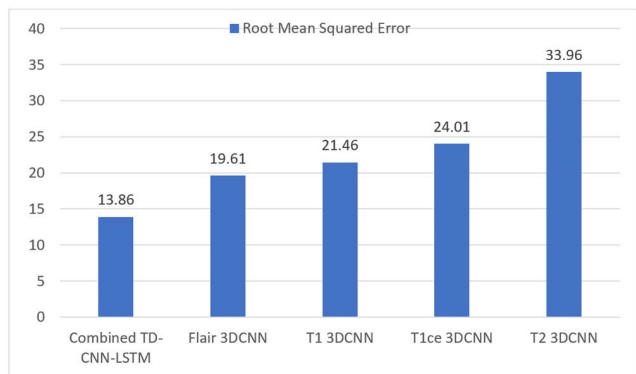
FIGURE 16. RMSE of all five configurations (3D CNN model trained on 4 datasets and TD-CNN-LSTM model trained on a combined dataset).

D. OPTIMAL MODEL EVALUATION

In order to further evaluate the effectiveness of TD-CNN-LSTM model on the external dataset, a confusion matrix and ROC curve is generated. Moreover, experimentations with various K-folds configurations are conducted to observe the robustness of proposed model.

A confusion matrix for the results of TD-CNN-LSTM model tested on the external dataset is presented in Fig. 17. The row values represent actual labels of the test dataset and the column values represent predicted labels of the test dataset. In this case there is a binary classification with classes HGG and LGG. Diagonal values from top left to bottom right are TP and TN values. It can be seen that all HGG cases were predicted correctly and that there was just one misclassification for LGG. The TD-CNN-LSTM classifier trained on the combined dataset does not seem to be biased towards any one class which validates the robustness of the model.
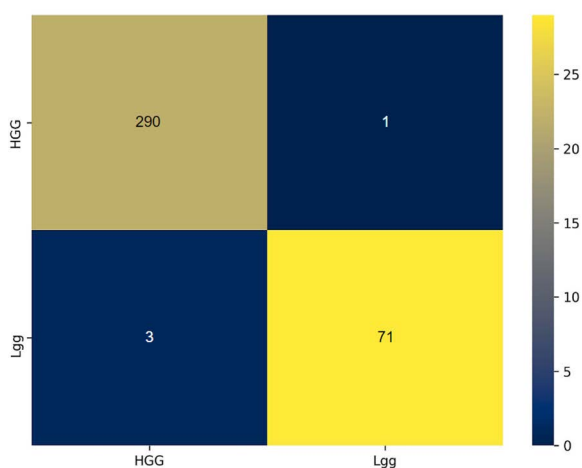

FIGURE 17. Confusion matrix of combined dataset trained on TD-CNN-LSTM model.

ROC probability curve is plotted, and the AUC value is derived (Fig. 18). The AUC is used as a summary of the ROC curve, representing the performance of a model in distinguish
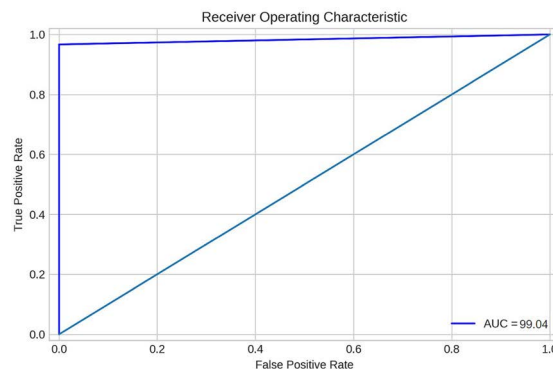
classes. An AUC value of 1 indicates that the model is able to detect all the TP and TN values flawlessly. Fig. 18 shows that the ROC curve touches almost at the pick of y-axis with a false positive rate close to 0 and a true positive rate closer to 1. The AUC value is **99.04%**.

In order to justify the consistency of the performance of the model even further, we test the model using a total of twelve K-fold cross validation configurations with various K values ranging between 3 to 30. Findings of each K-fold cross validation are presented in Fig. 19.
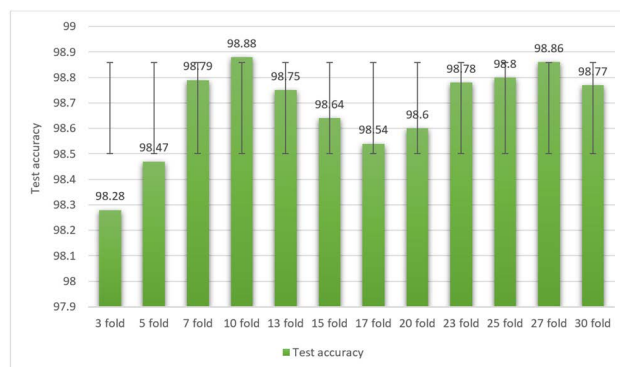

FIGURE 18. ROC curve of combined dataset trained on TD-CNN-LSTM.

FIGURE 19. Performance evaluation using K-fold cross validations with various K values ranging from 3 to 30.

It is evident that, for all the K-folds the model is able to yield good performance (>98%). The performance did not drop significantly for any fold, which further adds to the robustness and consistency of our model. It also indicates the superiority of the proposed approach (combining all four MRI sequences and using hybrid CNN and LSTM model) in 3D MRI classification problems where the model is able to generate such constant performance.

The best performing classifier TD-CNN-LSTM trained on the external combined dataset produced promising results across all evaluation measures. The TimeDistributed 2D convolutional layers in TD-CNN-LSTM model are capable of analyzing the combined 3D dataset. As discussed earlier, this combined dataset feeds a set of 3D images to the classifier simultaneously, taking less than 3 seconds per epoch

TABLE 8. Comparison of accuracy with some existing literatures.

Paper	Dataset for classification	Used sequence	Prediction Classes	Dimension	Classifier	Classification Accuracy
D. Jude et al.,2019 [15]	Abnormal MR brain tumor images from M/s. Devaki Scan Centre Total of 220 Brain MRI images	T1, T2 and T2 FLAIR	Metastasis, Meningioma, Glioma and Astrocytoma	2D	Modified Deep CNN	96.4%
Hiba et al., 2020 [16]	Multimodal Brain Tumor Segmentation Challenge 2018 (BraTS, 2018) Total of 284 Brain MRI images: 209 HGG and 75 LGG	T1-Gado	LGG HGG classification	3D	Deep CNN	96.49%
Linmin et al., 2020 [26]	CPM-RadPath 2019 validation and testing datasets.	---	LGG and HGG	3D	3D CNN	58.6%
Mohamed et al., 2020 [14]	The Cancer Imaging Archive (TCIA) 110 LGG images	T1 pre-contrast, FLAIR, and T1 post-contrast MRI sequences	Tumor grading	2D	Modified VGG-16	89%
Tommaso et al., [18]	Portoni Rossi Veterinary Hospital, Zola Predosa, Italy and Dick White Referrals, Six Mile Bottom, UK 56 meningioma images and 24 glioma images	Pre and post-contrast T1 and T2 sequences	meningiomas and gliomas	2D	GoogleNet	94%
Ying et al., 2020 [19]	The Cancer Imaging Archive (TCIA) low grade glioma (LGG) data 108 LGG images Multimodal Brain Tumor Image Segmentation (BraTS) Benchmark 2018 210 HGG images and 75 LGG images	T1, T1-Gd, T2, and FLAIR	HGG and LGG	3D	3DConvNet	97.1%
Thiruvenkadam et al., 2019 [20]	Multimodal Brain Tumor Segmentation Challenge 2013 (BraTS, 2013) and WBA data set contains eight volumes (V01-V08) of T2 images. Randomly selected 2D slices of 4500 images from 3D volumed images	T2	Normal and tumorous	2D	FLSCBN (CNN model with hyper parameter adjustment)	88.91%
Iram et al., 2018 [21]	Multimodal Brain Tumor Segmentation Challenge 2015 (BraTS, 2015) Selected Total of 60 Brain MRI images: 30 HGG and 30 LGG	FLAIR	HGG and LGG	2D	LSTM	84%
Rukmani et al., 2020 [12]	Multimodal Brain Tumor Segmentation Challenge 2018 (BraTS, 2018) Total of 300 Brain MRI images: 250 HGG and 50 LGG	FLAIR	High Grade and Low Grade	2D	AlexNet-LSTM	85%
Proposed approach	**Multimodal Brain Tumor Segmentation Challenge 2020 (BraTS, 2020) Total of 308 Brain MRI images: 234 HGG and 74 LGG**	**FLAIR, T1, T1ce, T2**	**HGG and LGG**	**3D**	**TD-CNN-LSTM**	**98.90%**

in training. The model is able to achieve highest test accuracy in under 80 epochs whereas a traditional 3D CNN model takes much longer to be trained (150 epochs) to achieve moderate accuracy. Since the four sequences in 3D MRI images contain features of different aspects of brain tumors, combining all these sequences into a single input makes the tumor region more defined which makes it easier to identify tumors. The entire approach of combining all the MRI sequences while training gives the model more definition of the tumor region that consequently boosts the classification performance while maintaining a low training time. Furthermore, the fact that the model is able to give a 98.90% correct prediction on an unseen external test dataset adds to the effectiveness of this approach on real world datasets. Moreover, the fact that TD-CNN-LSTM along with the combined dataset had similar performance across multiple K-fold configurations

is an indication of the robustness of the proposed approach.

E. COMPARISON WITH SOME EXISTING LITERATURE

In this section, the proposed TD-CNN-LSTM model is compared with some recent studies in brain tumour classification (section 3). Table 8 shows a comparison between these previous studies and our proposed approach based on accuracy. As stated before, the proposed TD-CNN-LSTM model, trained on the combined data set has the best results in our study, achieving an accuracy of 98.90%. More than half of the studies shown in Table 8 used similar 3D Brain MRI datasets and two studies used classifiers that are somewhat similar to our TD-CNN-LSTM model. Most results in Table 8 are in the range of 85%- 97%, which is lower than ours. Hiba *et al.* [16] used BraTS 2018 dataset achieving an accuracy of 96.49% with Deep CNNs. However, only T1-Gado MRI sequence was utilized. Unlike the previous work, Thiruvenkada *et al.*'s study [20] used 2D slices of the T2 sequence and managed to achieve an accuracy of 88.91%. Jude *et al.* [15] proposed a Modified Deep CNN classifier which can classify 2D MRI slices and demonstrated an accuracy of 96.4%. Mohammad at el. [17] used 2D slices from T1 pre-contrast, FLAIR, and T1 post-contrast sequences and using modified VGG 16, obtained an accuracy of 89%. A similar study was done by Tommaso *et al.* [18] where with GoogleNet model and 2Dslices of pre and post-contrast T1 and T2 sequences, they were able to achieve an accuracy of 94%. A fairly recent study conducted by Ying Zhuge [19] used a 3D CNN model and gained anaccuracy of 97.1% using T1, T1-Gd, T2, and FLAIR sequences. Linmin *et al.* [26] proposed a 3D CNN model that can train on 3D MRI images and they achieved an accuracy of 58.6% which is the lowest of all the studies shown in Table 8. Iram *et al.* [21] and Rukmani *et al.* [12] studies are more comparable to this paper. These studies used BraTS 2015 and BraTS 2018 respectively and used LSTM model which is a bit similar to our study. However, they used 2D slices for HGG and LGG classification. BraTS with LSTM, Iram *et al.* [21] achieved 84% accuracy whereas Rukmani *et al.* [12] achieved 85% accuracy score using the AlexNet-LSTM model.

Our proposed TD-CNN-LSTM model outperforms all these studies with a test accuracy of 98.90%. The idea of using four 3D images of a particular subject as a single input to a deep learning model contributed to the effectiveness of our mode. With this process more details of the brain tumour can be learned by the model than the other approaches described in literature. Training on this combined dataset makes the training process less time consuming and highly detailed 3D images make it easier for the classifier to distinguish between HGG and LGG tumors. Moreover, in this research, vigorous K-fold cross validation experimentations are performed which concludes in consistent performance of the model in all 12 K-fold configurations. A large external test dataset (265 samples) ensures that the model is tested properly for every ablation study cases in building the model that adds to sound evaluation of the robustness of proposed approach.

Furthermore, the optimal accuracy achieved from an external test dataset gives a glimpse of the model's absolute interpretation of 3D MRIs in real world application. With this proposed approach TD-CNN-LSTM trained on the combined dataset can outperform all other studies while also requiring fewer epochs for training. This demonstrates the potential of the proposed approach in classifying brain tumors from 3D MRI scans.

IX. CONCLUSION

The aim of this study is to mimic medical expert's diagnosis process of analysing all the MRI sequences of a single patient to determine cancer, into a deep learning approach with the highest accuracy and lowest computational complexity. A completely automated hybrid CNN model named TD-CNN-LSTM combining CNN with LSTM for classifying brain tumors into HGG and LGG using 3D volumetric MRI is proposed in this paper, the input layer is wrapped with TimeDistributed function so that all the four MRI sequences of a patient can be passed as single input data. In addition, a 3D CNN approach is employed to train with MRI sequences independently to compare the performance of our proposed model. Results show that the TD-CNN-LSTM network with optimal configuration outperforms 3D CNN model achieving the highest accuracy of 98.90% with optimizer Nadam and learning rate of 0.0001. Moreover, the results of K-fold cross validation method validate the robustness and consistency of our model's performance over different training scenarios. Our method of analysing all the MRI sequences together and developing a CNN model employing ablation study in order to achieve the highest accuracy while keeping the lowest time complexity, can be a useful approach in future research and real world tumor diagnosis system.

REFERENCES

[1] D. N. Louis, A. Perry, G. Reifenberger, A. Von Deimling, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, O. D. Wiestler, P. Kleihues, and D. W. Ellison, "The 2016 World Health Organization classification of tumors of the central nervous system: A summary," *Acta Neuropathol.*, vol. 131, no. 6, pp. 803–820, Jun. 2016, doi: 10.1007/s00401-016-1545-1.

[2] V. K. Y. Ho, J. C. Reijneveld, R. H. Enting, H. P. Bienfait, P. Robe, B. G. Baumert, and O. Visser, "Changing incidence and improved survival of gliomas," *Eur. J. Cancer*, vol. 50, no. 13, pp. 2309–2318, Sep. 2014, doi: 10.1016/j.ejca.2014.05.019.

[3] M. Faranoush, M. Torabi-Nami, A. Mehrvar, A. A. HedayatiAsl, M. Tashvighi, R. R. Parsa, M. A. Fazeli, B. Sobuti, N. Mehrvar, A. Jafarpour, R. Zangooei, M. Alebouyeh, M. Abolghasemi, A.-H. Vahabie, and P. Vossough, "Classifying pediatric central nervous system tumors through near optimal feature selection and mutual information: A single center cohort," *Middle East J. Cancer*, vol. 4, no. 4, pp. 153–162, 2013. [Online]. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed11&AN=2013707966%5Cnhttp://sfx.leidenuniv.nl:9003/sfx_local?sid=OVID:Embase&issn=2008-6709&isbn=&volume=4&issue=4&spage=153&date=2013&pid=%3Cauthor%3EFaranoush+M.%3C

[4] H. A. Khalil, S. Darwish, Y. M. Ibrahim, and O. F. Hassan, "3D-MRI brain tumor detection model using modified version of level set segmentation based on dragonfly algorithm," *Symmetry*, vol. 12, no. 8, p. 1256, Jul. 2020, doi: 10.3390/SYM12081256.

[5] S. Banerjee, S. Mitra, F. Masulli, and S. Rovetta, "Deep radiomics for brain tumor detection and classification from multi-sequence MRI," pp. 1–15, 2019, *arXiv:1903.09240*.

[6] G. S. Tandel, A. Balestrieri, T. Jujaray, N. N. Khanna, L. Saba, and J. S. Suri, "Multiclass magnetic resonance imaging brain tumor classification using artificial intelligence paradigm," *Comput. Biol. Med.*, vol. 122, Jul. 2020, Art. no. 103804, doi: 10.1016/j.compbiomed.2020.103804.

[7] J. Liu, M. Li, J. Wang, F. Wu, T. Liu, and Y. Pan, "A survey of MRI-based brain tumor segmentation methods," *Tsinghua Sci. Technol.*, vol. 19, no. 6, pp. 578–595, Dec. 2014, doi: 10.1109/TST.2014.6961028.

[8] C. Chung, U. Metser, and C. Ménard, "Advances in magnetic resonance imaging and positron emission tomography imaging for grading and molecular characterization of glioma," *Seminars Radiat. Oncol.*, vol. 25, no. 3, pp. 164–171, Jul. 2015, doi: 10.1016/j.semradonc.2015.02.002.

[9] H. Saleem, A. R. Shahid, and B. Raza, "Visual interpretability in 3D brain tumor segmentation network," *Comput. Biol. Med.*, vol. 133, Jun. 2021, Art. no. 104410, doi: 10.1016/j.compbiomed.2021.104410.

[10] E. Irmak, "Multi-classification of brain tumor MRI images using deep convolutional neural network with fully optimized framework," *Iranian J. Sci. Technol., Trans. Electr. Eng.* vol. 45, no. 3, pp. 1015–1036, 2021, doi: 10.1007/s40998-021-00426-9.

[11] M. M. Badža and M. C. Barjaktarović, "Classification of brain tumors from MRI images using a convolutional neural network," *Appl. Sci.*, vol. 10, no. 6, p. 1999, 2020, doi: 10.3390/app10061999.

[12] S. R. Devi and P. Selvaraju, "Classification of brain tumor image based on high grade and low grade using CNN with LSTM," *Int. J. Adv. Sci. Technol.*, vol. 29, no. 7, pp. 3008–3017, 2020.

[13] R. A. Zeineldin, M. E. Karar, J. Coburger, C. R. Wirtz, and O. Burgert, "DeepSeg: Deep neural network framework for automatic brain tumor segmentation using magnetic resonance FLAIR images," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 15, no. 6, pp. 909–920, Jun. 2020, doi: 10.1007/s11548-020-02186-z.

[14] E. I. Zacharaki, S. Wang, S. Chawla, D. S. Yoo, R. Wolf, E. R. Melhem, and C. Davatzikos, "Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme," *Magn. Reson. Med.*, vol. 62, no. 6, pp. 1609–1618, Dec. 2009, doi: 10.1002/mrm.22147.

[15] D. J. Hemanth, J. Anitha, A. Naaji, O. Geman, D. E. Popescu, and L. H. Son, "A modified deep convolutional neural network for abnormal brain image classification," *IEEE Access*, vol. 7, pp. 4275–4283, 2018, doi: 10.1109/ACCESS.2018.2885639.

[16] H. Mzoughi, I. Njeh, A. Wali, M. B. Slima, A. BenHamida, C. Mhiri, and K. B. Mahfoudhe, "Deep multi-scale 3D convolutional neural network (CNN) for MRI gliomas brain tumor classification," *J. Digit. Imag.*, vol. 33, no. 4, pp. 903–915, Aug. 2020, doi: 10.1007/s10278-020-00347-9.

[17] M. A. Naser and M. J. Deen, "Brain tumor segmentation and grading of lower-grade glioma using deep learning in MRI images," *Comput. Biol. Med.*, vol. 121, Jun. 2020, Art. no. 103758, doi: 10.1016/j.compbiomed.2020.103758.

[18] T. Banzato, M. Bernardini, G. B. Cherubini, and A. Zotti, "A methodological approach for deep learning to distinguish between meningiomas and gliomas on canine MR-images," *BMC Veterinary Res.*, vol. 14, no. 1, pp. 1–6, Dec. 2018, doi: 10.1186/s12917-018-1638-2.

[19] Y. Zhuge, H. Ning, P. Mathen, J. Y. Cheng, A. V. Krauze, K. Camphausen, and R. W. Miller, "Automated glioma grading on conventional MRI images using deep convolutional neural networks," *Med. Phys.*, vol. 47, no. 7, pp. 3044–3053, Jul. 2020, doi: 10.1002/mp.14168.

[20] T. Kalaiselvi, T. Padmapriya, P. Sriramakrishnan, and V. Priyadharshini, "Development of automatic glioma brain tumor detection system using deep convolutional neural networks," *Int. J. Imag. Syst. Technol.*, vol. 30, no. 4, pp. 926–938, Dec. 2020, doi: 10.1002/ima.22433.

[21] I. Shahzadi, T. B. Tang, F. Meriadeau, and A. Quyyum, "CNN-LSTM: Cascaded framework for brain tumour classification," in *Proc. IEEE-EMBS Conf. Biomed. Eng. Sci. (IECBES)*, Dec. 2018, pp. 633–637, doi: 10.1109/IECBES.2018.8626704.

[22] (2018). *Brain Tumor Segmentation 2018 Dataset*. [Online]. Available: https://www.kaggle.com/datasets/andrewmvd/brain-tumor-segmentation-in-mri-brats-2015

[23] (2019). *Brain Tumor Segmentation 2019 Dataset*. [Online]. Available: https://www.kaggle.com/datasets/andrewmvd/brain-tumor-segmentation-in-mri-brats-2015

[24] (2020). *Brain Tumor Segmentation 2020 Dataset*. [Online]. Available: https://www.kaggle.com/awsaf49/brats20-dataset-training-validation

[25] K. Kamnitsas, C. Ledig, V. F. 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.

[26] L. Pei, L. Vidyaratne, M. M. Rahman, and K. M. Iftekharuddin, "Context aware deep learning for brain tumor segmentation, subtype classification, and survival prediction using radiology images," *Sci. Rep.*, vol. 10, no. 1, pp. 1–11, Dec. 2020, doi: 10.1038/s41598-020-74419-9.

[27] A. I. Tawfik and W. H. Kamr, "Diagnostic value of 3D-FLAIR magnetic resonance sequence in detection of white matter brain lesions in multiple sclerosis," *Egyptian J. Radiol. Nucl. Med.*, vol. 51, no. 1, pp. 1–9, Dec. 2020, doi: 10.1186/s43055-020-00247-6.

[28] R. Bitar, G. Leung, R. Perng, S. Tadros, A. R. Moody, J. Sarrazin, C. McGregor, M. Christakis, S. Symons, A. Nelson, and T. P. Roberts, "MR pulse sequences: What every radiologist wants to know but is afraid to ask," *RadioGraphics*, vol. 26, no. 2, pp. 513–537, Mar. 2006, doi: 10.1148/rg.262055063.

[29] D. Daimary, M. B. Bora, K. Amitab, and D. Kandar, "Brain tumor segmentation from MRI images using hybrid convolutional neural networks," *Proc. Comput. Sci.*, vol. 167, pp. 2419–2428, Jan. 2020, doi: 10.1016/j.procs.2020.03.295.

[30] S. Rathore, H. Akbari, and J. Doshi, "Radiomic signature of infiltration in peritumoral edema predicts subsequent recurrence in glioblastoma: Implications for personalized radiotherapy planning," *J. Med. Imag.*, vol. 5, no. 2, p. 1, Mar. 2018, doi: 10.1117/1.jmi.5.2.021219.

[31] 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, pp. 1–17, Dec. 2021, doi: 10.1038/s41598-021-90428-8.

[32] S. Sung, H. S. Kim, and J. W. Kwon, "MRI assessment of sacroiliitis for the diagnosis of axial spondyloarthropathy: Comparison of fat-saturated T2, STIR and contrast-enhanced sequences," *Brit. J. Radiol.*, vol. 90, no. 1078, pp. 7–9, 2017, doi: 10.1259/bjr.20170090.

[33] P. Ding, J. Li, L. Wang, M. Wen, and Y. Guan, "HYBRID-CNN: An efficient scheme for abnormal flow detection in the SDN-based smart grid," *Secur. Commun. Netw.*, vol. 2020, pp. 1–20, Aug. 2020, doi: 10.1155/2020/8850550.

[34] A. Crimi, *Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, vol. 2. Cham, Germany: Springer, 2018.

[35] T. T. Ho, T. Kim, W. J. Kim, C. H. Lee, K. J. Chae, S. H. Bak, S. O. Kwon, G. Y. Jin, E.-K. Park, and S. Choi, "A 3D-CNN model with CT-based parametric response mapping for classifying COPD subjects," *Sci. Rep.*, vol. 11, no. 1, pp. 1–12, Dec. 2021, doi: 10.1038/s41598-020-79336-5.

[36] A. Lecler, L. Duron, D. Balvay, J. Savatovsky, O. Bergès, M. Zmuda, E. Farah, O. Galatoire, A. Bouchouicha, and L. S. Fournier, "Combining multiple magnetic resonance imaging sequences provides independent reproducible radiomics features," *Sci. Rep.*, vol. 9, no. 1, pp. 1–8, Dec. 2019, doi: 10.1038/s41598-018-37984-8.

[37] S. Arif, J. Wang, T. Ul Hassan, and Z. Fei, "3D-CNN-based fused feature maps with LSTM applied to action recognition," *Future Internet*, vol. 11, no. 2, p. 42, Feb. 2019, doi: 10.3390/fi11020042.

[38] H. Qiao, T. Wang, P. Wang, S. Qiao, and L. Zhang, "A time-distributed spatiotemporal feature learning method for machine health monitoring with multi-sensor time series," *Sensors*, vol. 18, no. 9, p. 2932, Sep. 2018, doi: 10.3390/s18092932.

[39] R. L. Gollub, S. N. Murphy, R. L. Robertson, and P. E. Grant, *Brain Age Estimation Using LSTM on Children's Brain MRI*, S. He, J. D. Perez, S. Prabhu, and R. Pienaar, Eds. Boston, MA, USA: Massachusetts General Hospital, 2020, pp. 420–423.

[40] A. A. Novikov, D. Major, M. Wimmer, D. Lenis, and K. Buhler, "Deep sequential segmentation of organs in volumetric medical scans," *IEEE Trans. Med. Imag.*, vol. 38, no. 5, pp. 1207–1215, May 2019, doi: 10.1109/TMI.2018.2881678.

[41] G. Ercolano and S. Rossi, "Combining CNN and LSTM for activity of daily living recognition with a 3D matrix skeleton representation," *Intell. Service Robot.*, vol. 14, no. 2, pp. 175–185, Apr. 2021, doi: 10.1007/s11370-021-00358-7.

[42] W. Lu, J. Li, Y. Li, A. Sun, and J. Wang, "A CNN-LSTM-based model to forecast stock prices," *Complexity*, vol. 2020, pp. 1–10, Nov. 2020, doi: 10.1155/2020/6622927.

[43] X. H. Le, H. V. Ho, G. Lee, and S. Jung, "Application of long short-term memory (LSTM) neural network for flood forecasting," *Water*, vol. 11, no. 7, p. 1387, Jul. 2019, doi: 10.3390/w11071387.

[44] S. Montaha, S. Azam, A. K. M. R. H. Rafid, P. Ghosh, M. Hasan, M. Jonkman, and F. D. Boer, ''BreastNet18: A high accuracy fine-tuned VGG16 model evaluated using ablation study for diagnosing breast cancer from enhanced mammography images,'' *Biology*, vol. 10, no. 12, p. 1347, 2021.

[45] C. de Vente, L. H. Boulogne, K. V. Venkadesh, C. Sital, N. Lessmann, C. Jacobs, C. I. Sánchez, and B. van Ginneken, ''Improving automated COVID-19 grading with convolutional neural networks in computed tomography scans: An ablation study,'' 2020, *arXiv:2009.09725*.

[46] R. Meyes, M. Lu, C. W. de Puiseau, and T. Meisen, ''Ablation studies in artificial neural networks,'' 2019, pp. 1–19, *arXiv:1901.08644*.

[47] E. S. N. Joshua, D. Bhattacharyya, M. Chakkravarthy, and Y.-C. Byun, ''3D CNN with visual insights for early detection of lung cancer using gradient-weighted class activation,'' *J. Healthcare Eng.*, vol. 2021, pp. 1–11, Mar. 2021, doi: 10.1155/2021/6695518.

[48] C. Feng, A. Elazab, P. Yang, T. Wang, F. Zhou, H. Hu, X. Xiao, and B. Lei, ''Deep learning framework for Alzheimer's disease diagnosis via 3D-CNN and FSBi-LSTM,'' *IEEE Access*, vol. 7, pp. 63605–63618, 2019, doi: 10.1109/ACCESS.2019.2913847.

[49] P. A. T. Hapsari, J. R. Dewinda, V. A. Cucun, Z. F. Nurul, S. Joan, D. S. Anggraini, M. A. Peter, E. P. I. Ketut, and H. P. Mauridhi, ''Brain tumor classification in MRI images using en-CNN,'' *Int. J. Intell. Eng. Syst.*, vol. 14, no. 4, pp. 437–451, Aug. 2021, doi: 10.22266/ijies2021.0831.38.

[50] L. Lu, Y. Yang, Y. Jiang, H. Ai, and W. Tu, ''Shallow convolutional neural networks for acoustic scene classification,'' *Wuhan Univ. J. Natural Sci.*, vol. 23, no. 2, pp. 178–184, Apr. 2018, doi: 10.1007/s11859-018-1308-z.

[51] F. Lei, X. Liu, Q. Dai, and B. W.-K. Ling, ''Shallow convolutional neural network for image classification,'' *Social Netw. Appl. Sci.*, vol. 2, no. 1, pp. 1–8, Jan. 2020, doi: 10.1007/s42452-019-1903-4.

[52] S. Yu, S. Wu, L. Wang, F. Jiang, Y. Xie, and L. Li, ''A shallow convolutional neural network for blind image sharpness assessment,'' *PLoS ONE*, vol. 12, no. 5, pp. 1–16, 2017, doi: 10.1371/journal.pone.0176632.

[53] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, ''Lung pattern classification for interstitial lung diseases using a deep convolutional neural network,'' *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1207–1216, Feb. 2016, doi: 10.1109/TMI.2016.2535865.

[54] C. Banerjee, T. Mukherjee, and E. Pasiliao, ''An empirical study on generalizations of the ReLU activation function,'' in *Proc. ACM Southeast Conf.*, Apr. 2019, pp. 164–167, doi: 10.1145/3299815.3314450.

[55] I. Kandel and M. Castelli, ''The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset,'' *ICT Exp.*, vol. 6, no. 4, pp. 312–315, Jan. 2020, doi: 10.1016/j.icte.2020.04.010.

[56] K. Mahmud, S. Azam, A. Karim, S. Zobaed, B. Shanmugam, and D. Mathur, ''Machine learning based PV power generation forecasting in Alice springs,'' *IEEE Access*, vol. 9, pp. 46117–46128, 2021, doi: 10.1109/ACCESS.2021.3066494.

A. K. M. RAKIBUL HAQUE RAFID received the bachelor's degree in computer science and engineering from Daffodil International University, Dhaka, Bangladesh. His research interests include computer vision, health informatics, machine learning, deep learning, and artificial intelligence-based systems.

MD. ZAHID HASAN (Member, IEEE) received the Master of Science (M.Sc.) degree in information and communication engineering from the University of Rajshahi, in 2009, and the Ph.D. degree in computer science and engineering from Jahangirnagar University. He is currently working as an Associate Professor with the Department of Computer Science and Engineering, Daffodil International University, Bangladesh. His research interests include computer vision, health informatics, machine learning, artificial intelligence-based systems, and decision theory.

ASIF KARIM (Member, IEEE) is currently a Research Active Lecturer with Charles Darwin University, Australia. Alongside being an Active Researcher, he has considerable industry experience in IT, primarily in the field of software engineering. His research interests include machine intelligence, health informatics, and smart contracts.

SIDRATUL MONTAHA received the bachelor's degree in computer science and engineering from Daffodil International University, Dhaka, Bangladesh. Her research interests include computer vision, health informatics, machine learning, deep learning, and artificial intelligence-based systems.

SAMI AZAM is currently a Leading Researcher and a Senior Lecturer with the College of Engineering and IT, Charles Darwin University, Casuarina, NT, Australia. He has several publications in peer-reviewed journals and international conference proceedings. His research interests include computer vision, signal processing, artificial intelligence, and biomedical engineering.

ASHRAFUL ISLAM received the M.S. degree in computer science from the University of Louisiana at Lafayette (UL Lafayette), USA, in 2020, where he is currently pursuing the Ph.D. degree with the Center for Advanced Computer Studies. Since 2015, he has been publishing around 30 scientific articles in renowned venues and journals. His research interests include human–computer interaction, assistive technologies, applied machine learning, health informatics, and relational agent. He regularly acts as a PC Member and an international advisor board member for several conferences and a reviewer for top-tier venues and journals, including *CHI*, *Journal of Medical Internet Research*, and *Journal of Neuroscience Methods*.

. . .