Ganesh Aravind Harkude. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 12, Issue 8, August 2022, pp. 17-26

RESEARCH ARTICLE OPEN ACCESS

Deep Spatio-spatial Models for Classifying Brain Tumors in MR Images

Ganesh Aravind Harkude
Artificial Intelligence and Machine Learning
BMS Institute of Technology and Management

ABSTRACT
A brain tumour is a mass or cluster of abnormal cells in the brain that has the potential to spread to other parts of the body and pose a serious threat to the patient's life. For effective treatment planning, a precise diagnosis is necessary, and the main imaging technique for determining the extent of brain tumours is magnetic resonance imaging. The majority of this increase in Deep Learning techniques for computer vision applications may be attributed to the availability of a sizable quantity of data for model training and the advancements in model designs that produce better approximations in a supervised environment. The availability of free datasets with trustworthy annotations has significantly improved the classification of cancers using such deep learning techniques. These techniques often use either 3D models that employ 3D volumetric MRIs or even 2D models that take each slice into account independently. However, spatiotemporal models may be used as "spatial" models for this job by treating each spatial dimension individually or by seeing the slices as a succession of pictures through time [2]. These models may learn certain spatial and temporal correlations while using less processing power.
This study classifies several types of brain tumours using two spatiotemporal models, ResNet (2+1) D and ResNet Mixed Convolution. Both of these models outperformed the ResNet18 pure 3D convolutional model, it was determined. It was also shown that pre-training the models on a distinct, even unrelated dataset before training them for the objective of cancer classification enhances performance. Pre-trained ResNet Mixed Convolution, which had the lowest computational cost and a macro F1-score of 0.9545, was found to be the most accurate model in these studies. It also achieved a test accuracy of 98.98 percent.
Keywords: Brain Tumor, ResNet, MRI, Data, Convolution Neural Network, F1_Score, Accuracy

I. INTRODUCTION

The expansion of aberrant brain cells is known as a brain tumour. Based on their rate of development and likelihood of recurrence following therapy, brain tumours are categorised. They can be broadly classified into two groups: malignant and benign. The likelihood of a benign tumour returning is lower following therapy since it is not malignant and grows slowly [7]. The majority of malignant tumours, on the other hand, are composed of cancer cells; they can either locally infiltrate tissues or migrate to other areas of the body through a process known as metastasis. Mutations in glial cells cause malignancy in normal cells, which results in glioma tumours. They represent 30% of all brain and central nervous system tumours and 80% of all malignant tumours, making them the most prevalent forms of astrocytomas (brain or spinal cord tumours) [4]. Glioma tumours can have Astrocytomas, Oligodendrogliomas, or Ependymomas as its phenotypic composition. The World Health Organization (WHO) employs the following grading-based methodology to categorise each of these tumours depending on their1aggressiveness [8]:

- **Grade 1:** Tumors are often benign, meaning they are usually treatable, and they are frequently encountered in youngsters.
- **Grade 2:** contains three different tumour types: oligodendrogliomas, oligoastrocytomas, and oligoastrocytoma, which combines both2. Adults commonly experience them. All low-grade gliomas have the potential to develop into high-grade tumours over time3.
- **Grade 3:** Anaplastic Astrocytomas, Anaplastic Oligodendrogliomas, or Anaplastic Oligoastrocytomas are examples of tumours. They are sneakier and aggressive than grade 2.
- **Grade 4:** The WHO class glioma, often known as glioblastoma multiforme (GBM), is the most dangerous tumour.

Grades I and II gliomas are typically referred to as low-grade gliomas (LGG), whilst grades III and IV gliomas are referred to as high-grade gliomas (HGG) [4]. The benign tumours known as LGG can be removed by surgical excision. HGGs, on the

other hand, are malignant tumours that are challenging to remove by surgical means due to the degree of adjacent tissue invasion. An example MRI of LGG and HGG is shown in Figure 1.

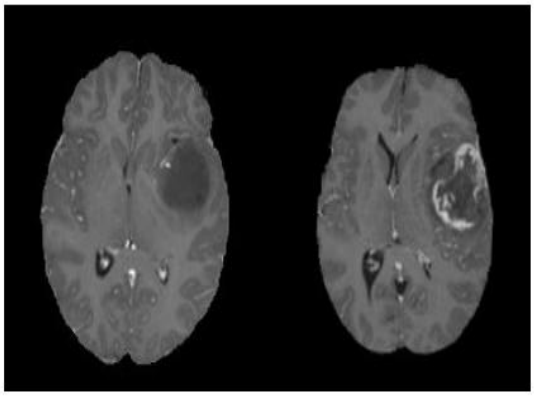

Figure No.1 : An illustration of an MRI showing a high-grade glioma (HGG) and a low-grade glioma (LGG), from BraTS 2019.

The following tissue types are frequently present in a Glioblastoma Multiforme (GBM) [11] (Figure No. 2):

- Tumor Core: The malignant cells that are aggressively growing in this area of the tumour.
- Necrosis: The crucial difference between low-grade gliomas and GBM4 is the necrotic area. The cells and tissue in this area are either dying or have already passed away.
- Perifocaloedema: The accumulation of fluid around the tumour core, which raises the intracranial pressure, results in brain swelling; perifocaloedema is brought on by alterations in glial cell distribution [12].

The location, histological subtype, and tumour margins are only a few of the variables that affect a brain tumor's prognosis. Even after therapy, the tumourfrequently returns and advances to grade IV3. The site of the tumour may be determined using contemporary imaging techniques like MRI, which is then utilised to investigate tumour progression and arrange surgical procedures. Along with its hemodynamics, MR imaging is utilised to evaluate the anatomy, physiology, and metabolic activity of the lesion. As a result, MR imaging continues to be the major method for diagnosing brain tumours [3].

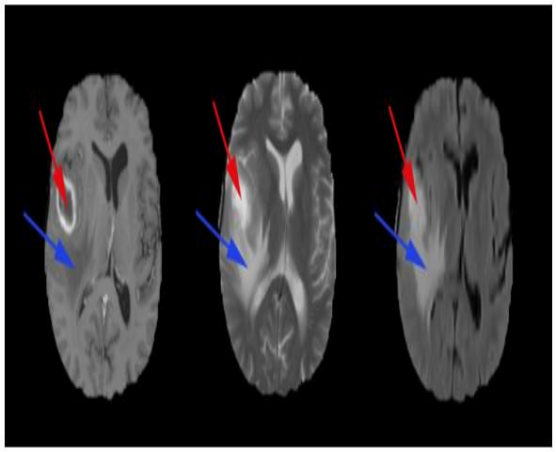

Figure No. 2 : From left to right, high-grade glioma structure on T1ce, T2, and FLAIR contrast images, necrotic core, perifocaloedemaBraTS 2019 as source

Early cancer identification in particular has the potential to alter how a patient is treated. Early diagnosis is essential because lesions that are detected earlier are more likely to be treatable; if action is taken, this might be the difference between life and death. Deep learning techniques can assist in automating the process of identifying and categorising brain lesions. By prioritising just malignant lesions, they can also lessen the radiologists' workload of analysing numerous pictures [1]. This can decrease diagnostic errors6 and eventually increase overall efficiency. Recent research has demonstrated that deep learning techniques in radiography have already surpassed human performance levels for several pathologies.

II. RELATED WORK

Recently, a number of deep learning-based approaches for classifying brain tumours have been presented. T1 contrast-enhanced images were used by **Mzoughi et al. [8]** to suggest a method for classifying high-grade and low-grade gliomas. **Pei et al. [9]** performed a similar study on the categorization of gliomas based on grading, segmenting the tumour first before classifying it as either HGG or LGG.

One MR contrast picture was utilised at a time for much of the research on the categorization and grading of glioma tumours, however **Ge et al.** [11] developed a fusion architecture that concurrently classified the tumour using T1 contrast-enhanced, T2, and FLAIR images.

The non-subsampled shearlet transform (NSST) was used by Ouerghi et al.11 to transform T1 images into low frequency (LF) and high frequency (HF) subimages, effectively separating principle information from edge information in the source

image. The images were then fused according to predefined rules to include the coefficients, resulting in the fusion of T1 and T2 or FLAIR images. The majority of the literature simply distinguishes between the various grades of tumours and does not include healthy brains as a separate category.

III. TECHNICAL BACKGROUND

One of the most effective network topologies for image identification tasks, ResNet or residual network, was proposed by **He et al. [**12] and addresses issues with deep networks, such as disappearing gradients. The identity mappings known as residual-links, or "skipped connections," are introduced in this study. Their outputs are appended to the outputs of the other stacked layers. These identification links enhance the training process without increasing network complexity. The spatiotemporal models for action recognition developed by **Tran et al.** [13] are essentially 3D convolutional neural networks built on ResNet.

Video data is three-dimensional since it has two spatial dimensions and one-time dimension. It is clear that utilising a network with 3D convolution layers is the best option for processing such data (such as an action detection job). ResNet (2+1) D and ResNet Mixed Convolution are two different types of spatiotemporal models that **Tran et al.** [14] presented. In the ResNet(2+1)D model, 2D and 1D convolutions are employed, with the 2D convolutions being used for the spatial component and the 1D convolutions being saved for the temporal component [13] . By utilising non-linear rectification, this provides a benefit of greater non-linearity and makes this type of mixed model more "learnable" than traditional complete 3D models.

By utilising non-linear rectification, this provides a benefit of greater non-linearity and makes this type of mixed model more "learnable" than traditional complete 3D models. The ResNet Mixed Convolution model, on the other hand, is built using a combination of 2D and 3D Convolution processes [20]. The model's first layers are constructed using 3D convolution techniques, whereas its subsequent layers use 2D convolutions. The justification for this setup is that since most motion-modelling takes place in the first few layers, using 3D convolution their better captures activity.

Transfer learning [14] is a method widely employed to boost the performance of the same network architecture in addition to trying to enhance the design itself. This method allows you to use a model that has already been trained to perform one job to perform another task entirely. Before beginning the training, the model parameters are typically initialised at random. Transfer learning, on the other hand, trains the model for task two using model parameters learnt from task one as the starting point (referred to as pre-training), rather than random values. Pre-training has proven to be a successful way to enhance the initial training process and subsequently increase accuracy.

IV. CONTRIBUTION

For three dimensional video classification applications, spatiotemporal models are frequently employed. Their potential for identifying "spatiospatial" models, such as 3D volumetric pictures like MRIs, has not yet been investigated. This examines the potential for using the spatiotemporal models ResNet(2+1)D and ResNet Mixed Convolution as "spatiospatial" models by treating the slice dimension of the three-dimensional volumetric pictures differently from the other two spatial dimensions. Using a single MR contrast, "Spatial" was used to classify brain tumours of various glioma kinds based on their grade as well as healthy brains from 3D volumetric MR Images. Their performances were compared to a pure 3D convolutional model (ResNet3D) [8]. For the purpose of evaluating the applicability of transfer learning for this task, the models will also be evaluated with and without pre-training.

V. METHODOLOGY

The network models utilised in this study are covered in depth in this part along with implementation information, pre-training and training methodologies, data augmentation approaches, dataset details, data pre-processing procedures, and evaluation metrics.

5.1: For tasks using video where there are two spatial and one temporal dimension, spatiotemporal models are typically employed. These models, as opposed to pure 3D convolution-based models, handle the spatial and temporal dimensions in various ways. A 3D convolution-based model is frequently used since 3D volumetric image classification tasks lack a time component. They are occasionally cut into 2D slices and subjected to 2D convolution-based models. In order to make the convolution kernels invariant to tissue discrimination in all dimensions and learn more complicated characteristics across voxels, 3D filters are used for the purpose of classifying tumours. 2D convolution filters will be used to capture the spatial representation inside the slices [17]. Spatial-temporal models can either reduce the complexity of the model or provide additional non-linearity by

combining two different forms of convolution into one model.

Considering the spatiotemporal models as "spatiospatial" models may make it feasible to take use of these benefits while working with volumetric data, which is why utilising such models for a tumour classification job is desirable. In this study, in-plane dimensions are taken as the spatial dimensions while slice-dimension is treated as the pseudo-temporal dimension of spatiotemporal models. The work of **Tran et al. [**13] served as the foundation for the spatiotemporal models employed here as spatial models.

ResNet (2+1)D and ResNet Mixed Convolution are two alternative spatiospatial models that are investigated in this article. Their results are contrasted with ResNet3D, a model that only uses 3D convolutions.

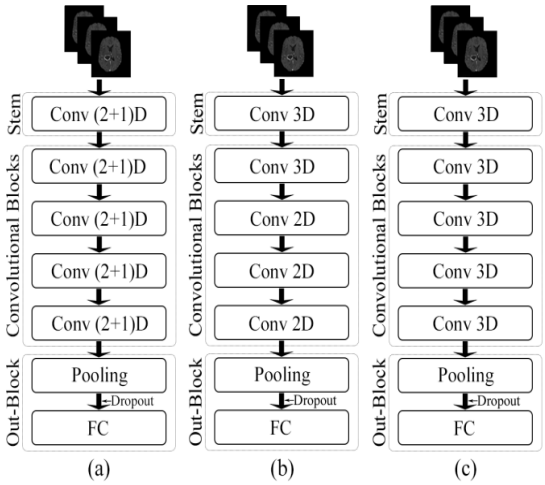

Figure No.3 : depictions in schematic form of the network architectures. ResNet (2+1) D, ResNet Mixed Convolution, and ResNet 3D are examples of neural networks.

5.1.1: ResNet (2 +1) D

Instead of using a single 3D convolution, ResNet (2+1) D employs a mixture of 2D convolution and 1D convolution. As opposed to utilising a single 3D Convolution, this design has the advantage of allowing an additional non-linear activation unit between the two convolutions. The network's overall number of ReLU units then rises as a result, enabling the model to learn ever more complicated functions. The ResNet(2+1)D employs a stem that consists of a 2D convolution with a kernel size of seven and a stride of two, accepting one channel as input and producing 45 channels as an output; this is followed by a 1D convolution with a kernel size of three and a stride of one, producing 64 channels as the final output [10].

A 2D convolution with a kernel size of three and a stride of one is found in each residual block, followed by a 1D convolution with a kernel size of three and a stride of one. A 3D batch normalisation layer, followed by a ReLUactivation function, follows each convolutional layer in the model—both the 2D and the 1D versions. In order to down sample, the input by half, a pair of 3D convolution layers with a kernel size of one and a stride of two are used to separate the residual blocks inside the convolutional blocks, with the exception of the first convolutional block. The 1D convolutions are performed on the slice dimension, whereas the 2D convolutions are applied in-plane. An adaptive average pooling layer with an output size of one for all three dimensions has been introduced after the last convolutional block. To achieve the final output, a dropout layer, a fully connected layer, and n output neurons for n classes were added after the pooling layer. The ResNet (2+1) D architecture's schematic diagram is shown in Fig. 3(a).

5.1.2: RESNET MIXED CONVOLUTION

A mixture of 2D and 3D convolutions are used in ResNet Mixed Convolution. This model's stem has a 3D convolution layer with a kernel size of (3,7,7), a stride of (1,2,2), and a padding of (1,3,3); the first dimension is the slice dimension and the other two are the in-plane dimensions. This layer receives a single channel as input and outputs 64 channels. Three 2D convolution blocks come after the stem, then one 3D convolution block.All convolution layers, whether 3D and 2D, share the same three-kernel size and one-stride parameters. Each of these residual blocks has two convolution layers, and each of these convolution blocks has two residual blocks [15]. Similar to ResNet (2+1)D, a pair of 3D convolution layers with a kernel size of one and a stride of two are used to divide the residual blocks inside the convolutional blocks, with the exception of the first convolutional block, in order to ownsample the input by half. A 3D batch normalisation layer and a ReLU activation function are placed after each convolutional layer in the model, both 3D and 2D.

The rationale behind utilising both 2D and 3D modes of convolution is that although 2D can learn representation inside each 2D slice, 3D filters can learn the spatial properties of the tumour in 3D space. The final pooling, dropout, and fully connected layers follow the convolutional blocks and are the same as those in the ResNet (2+1)D architecture. The schematic illustration of this concept is shown in Fig. 3(b).

5.1.3: NESTNET3D

A pure 3D ResNet model is used as the benchmark to evaluate the performance of the spatiospatial models against (c). ResNet3D's architecture is nearly identical to ResNet Mixed Convolution's design [19], with the exception that this model only employs 3D convolutions. The main variation between both models stems from the usage of four 3D convolution blocks in this model as opposed to one 3D convolution block, followed by three 2D convolution blocks, in ResNet Mixed Convolution. A 3D ResNet18 model is created using this ResNet3D architectural setup.

2.1.4: SUMMARY AND COMPARISON

The network models' overall architecture may be broken down into the following sections: the stem receives input, followed by four convolutional blocks, the output block, which contains an adaptive pooling layer, the dropout layer, and lastly the fully connected layer. The stem of ResNet Mixed Convolution and ResNet 3D is identical and consists of a 3D convolutional layer with a kernel size of (3,7,7), a batch normalisation layer, and a ReLU. A different stem is used by ResNet (2+1) D, which divides the 3D convolution (3,7,7) used by the other models into two 2D and one 1D convolution layer (7,7) and (3), respectively (3). A batch normalisation layer and ReLU pair are followed by both 2D and 1D convolution inside of this stem [18]. The convolutional blocks in the ResNet3D and ResNet Mixed Convolution designs have the same structure: two residual blocks made up of two subblocks, each of which has a 3D convolution with a three-kernel size, followed by a batch normalisation layer and a ReLU.

As opposed to the 3D convolutional layers used by the other models, the initial convolutional block of the ResNet (2+1)D architecture utilises a pair of 2D and 1D convolutions with a three kernel size. The remainder of the building is identical. Because the 3D convolutions are divided into a pair of 2D and 1D convolutions, it is noted that this model has more nonlinearity than others. Additional pairs of batch normalisation and ReLU may have been utilised between the 2D and 1D convolutions. The second, third, and fourth convolutional blocks all contained a down sampling pair, which was composed of a 3D convolutional layer with a kennel size of one and a stride of two, followed by a batch normalisation layer. This down sampling pair was included in the first convolutional block, but not in the other three blocks (applicable to all three models). In the first convolutional block, this was absent.

The number of input features to the first block is 64, while the number of output features to the fourth (and final) block is 512. The convolution blocks of each of the three models multiply the input features by two. In the last stage of each of these models, an adaptive average pooling layer imposes a 1x1x1 output shape with 512 distinct features. Before providing them to a fully connected linear layer that outputs n classes, a dropout with a probability of 0.3 is used to add regularisation and avoid over-fitting. These models have similar width and depth, but the number of trainable parameters varies based on the kind of convolution employed, as shown in Tab. 1. It is interesting that computational costs decrease with the number of trainable parameters [12]. A model with fewer parameters would require less computing resources (RAM and GPU), and it would also be less complicated, which would lower total computational costs for both training and inference. Additionally, fewer trainable parameters would lessen the chance of overfitting.

Model	Number of Parameters
RestNet3D	34,261,634
RestNet (2 + 1) D	32,398,366
RestNetMoxed Convolutions	22,583,975

Table No. 1 : Total Number of Models' Trainable Parameters

5.2: IMPLEMENTATION AND TRAINING

The models were created by altering the Torchvision models using PyTorch18, and they were trained with a batch size of 1 on an Nvidia RTX 4000 GPU with 8GB of RAM. Models with and without pre-training were contrasted. On Kinetics-40020, all models with pre-training had been trained, with the exception of the stems and completely linked layers. The 3D volumetric MRIs only have one channel, but the RGB images from the Kinetics dataset include three channels. As a result, the stem that had been trained on the Kinetics dataset was unable to be applied and was initialised at random [17]. The fully linked layer was additionally initialised with random weights because Kinetics-400 has 400 output classes whereas the job at hand only has three (LGG, HGG, and Healthy).

Trainings were carried out with the use of the Nvidia Apex library 22 and mixed-precision21. To minimise the under-representation of classes with fewer samples during training, the loss was

calculated using the weighted cross-entropy loss function, and it was optimised using the Adam optimiser with a learning rate of 1e-5 and a weight decay coefficient of =1e-3.

5.2.1: WEIGHTED CROSS ENTROPY LOSS

Each class's normalised weight value (Wc) is determined by:

$$Wc = \left[1 - \left(\frac{Sample_c}{\sum Sample_t}\right)\right]$$

samplest is the total number of samples from all classes, and samplesc is the number of samples from class c. This equation's normalised weight values are then utilised to scale the corresponding class loss's cross-entropy loss.

$$loss_c = W_c[-x_c log(P_c)]$$

Where P(c) is the estimated distribution for class c and xc is the real distribution for that class. The sum of the individual class losses is the overall cross-entropy loss.

$$Loss_{total} = loss_{c1} + loss_{c2} + \cdots \ldots + loss_{cn}$$

5.3: DATA AUGMENTATION

Before training the models, different data augmentation techniques were applied to the dataset, and TorchIO23 was utilised for that. Light and heavy augmentation were used in the initial experiments, with light augmentation consisting solely of random affine transformations (scale 0.9-1.2, degrees 10) and random flips (L-R, probability 0.25) and heavy augmentation consisting of the latter two as well as elastic deformation and random k-space transformations (motion, spike, and ghosting) [13]. In addition to having poor final accuracy, it was shown that the loss took substantially longer to converge when the network was trained using heavily augmented input. Therefore, in this study, relatively minimal augmentation was applied.

5.4: DATASET

In this study, two different datasets were used: the non-pathological images were taken from the IXI Dataset26, and the pathological images were taken from the Brain Tumour Segmentation (BraTS) 2019 dataset, which includes images with four different MR contrasts (T1, T1 contrast-enhanced, T2, and FLAIR). T1 contrast enhanced (T1ce) is the contrast that is most frequently employed when doing single-contrast tumour classification among the four types of MRIs that are available. Consequently, 332 participants' T1ce images from the BRaTS collection were used in this study: 259 volumes of high-grade glioma (HGG) and 73 volumes of low-grade glioma (LGG) [9]. To have the same number of individuals as HGG, 259 T1 weighted volumes were chosen at random from the IXI dataset as healthy samples. The resulting merged dataset was then arbitrarily split into three 7:3 training and testing halves.

5.5: DATA PRE-PROCESSING

The brain extraction tool (BET2) of FSL28, 29 was used as the initial step in the pre-processing of the IXI pictures. As the BraTS photos are already skull stripped, this was done to maintain consistency throughout the input data. As employed by Isensee et al.30, the intensity values of all the volumes from the combined datasets were also normalised by scaling intensities to the [0.5,99.5] percentile. Finally, the volumes were re-sampled with the same 2 mm isotropic voxel-resolution [1].

5.6: EVALUATION METRICS

Using precision, recall, F1 score, specificity, and testing accuracy, the models' performances were compared. A confusion matrix was also utilised to demonstrate class-wise accuracy.

VI. RESULTS

Comparisons were made between the models' performances with and without pre-training. Figures 4, 5, and 6 provide, for ResNet (2+1)D, ResNet Mixed Convolution, and ResNet 3D, respectively, the average accuracy across 3-fold cross validation using confusion measures [3].

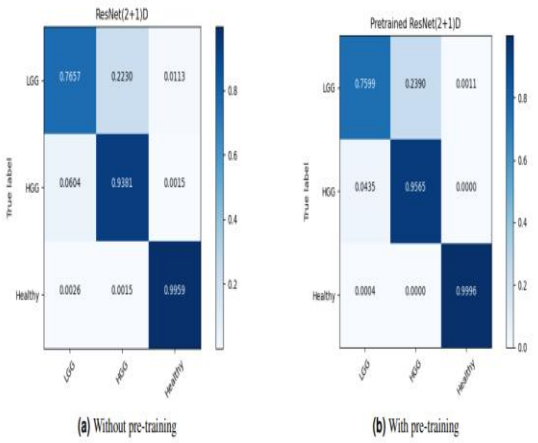

Figure No.4: Confusion Matrix for Pre-trained ResNet(2+1)D using 3-fold Cross-Validation

6.1: COMPARISON OF THE MODELS

The performance of the various models was compared using the mean F1-score across a 3-fold cross-validation. The results of the various models for the classes LGG, HGG, and Heathy are displayed in Tables 2, 3, and 4, respectively. Table 5 also displays the total scores. ResNet Mixed Convolution with pre-training obtained the highest

F1 score of 0.8949 with a standard deviation of 0.033 for low-grade glioma (LGG). With 0.8739 0.033, the pre-trained ResNet(2+1)D is not far behind [7].

The pre-trained ResNet Mixed Convolution model has the greatest F1 score for the high-grade glioma (HGG) class, 0.9123 0.029. This is greater than the F1 score of the top model in the LGG class. This is to be expected given the disparity in class between LGG and HGG. The Pre-trained ResNet(2+1)D with the F1 score of 0.8979 0.032 is also the second-best model for high-grade glioma, just like it was for low-grade glioma.

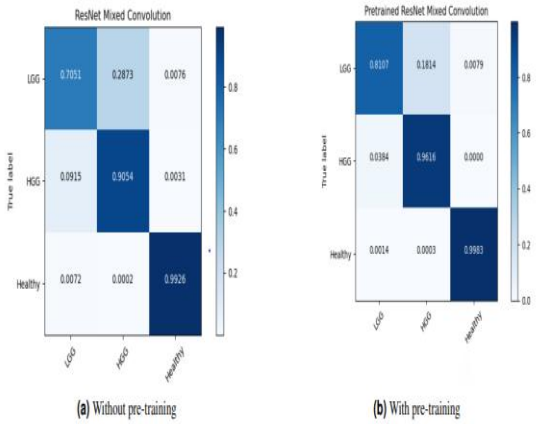

Figure No. 5 : Confusion Matrix for ResNet Mixed Convolution 3-fold Cross-Validation

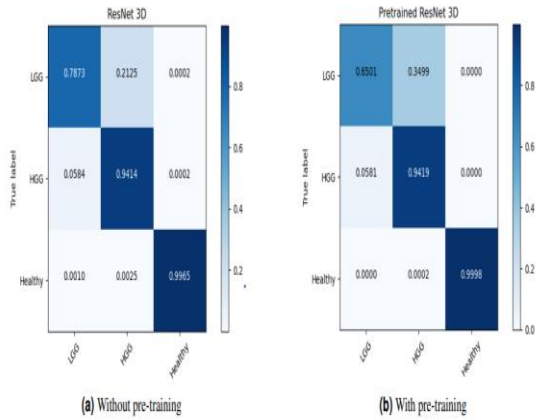

Figure No. 6 : Confusion Matrix for ResNet3D18 3-fold Cross-Validation

glioma of low grade	
Model	Mean F1_Score
ResNet 3D	0.8673±0.051
Pre – Trained ResNet 3D	0.8145±0.047
ResNet (2+1) D	0.8456±0.020
Pre – Trained ResNet (2 + 1)D	0.8738±0.041
ResNet Mixed Convultion	0.7782±0.032
Pre-trained ResNet Mixed Convolutions	0.9049±0.033

Table No.2 :Comparison of low-grade glioma models (* indicates the overall top model)

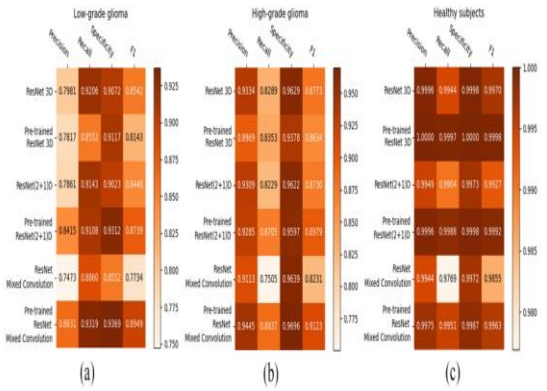

Figure No.7 :Heatmaps comparing Precision, Recall, Specificity, and F1-score to depict the class-wise performance of the classifiers: (a) Healthy, (b) HGG, and (c) LGG

glioma of high grade	
Model	Mean F1_Score
ResNet 3D	0.8845±0.041
Pre – Trained ResNet 3D	0.8741±0.039
ResNet (2+1) D	0.8852±0.029
Pre – Trained ResNet (2 + 1)D	0.8980±0.037
ResNet Mixed Convultion	0.8331±0.028

Pre-trained ResNet Mixed Convolutions	0.9249±0.034

Table No.3 : High-grade glioma model comparison (the asterisk (*) designates the overall top model)

Healthy Brain	
Model	Mean F1_Score
ResNet 3D	0.9972±0.005
Pre – Trained ResNet 3D	0.9998±0.001
ResNet (2+1) D	0.9928±0.002
Pre – Trained ResNet(2 + 1)D	0.9855±0.005
ResNet Mixed Convultion	0.9963±0.002
Pre-trained ResNet Mixed Convolutions	0.9985±0.003

Table No. 4 : Comparison of the healthy brain models (* indicates the overall top model)

The absence of any lesion in the MR images made it much easier for the model to learn and identify it from the brain MRIs with disease, which is why the healthy brain class got the highest F1 score of 0.9998, 0.0002 using the pre-trained ResNet 3D model [4]. It is challenging to choose a clear victor since, despite the fact that the pre-trained ResNet 3D model had the highest mean F1 score, other pre-trained models had similar F1 scores, i.e., all the mean scores are more than 0.9960.

Consolidated Score		
Model	Mean F1_Score	weighted F1 score
ResNet 3D	0.9096	0.9270
Pre – Trained ResNet 3D	0.8952	0.9272
ResNet (2+1) D	0.9045	0.9325
Pre – Trained ResNet (2 + 1)D	0.9328	0.9494
ResNet Mixed Convultion	0.8652	0.8891
Pre-trained ResNet Mixed Consolidated Score	0.9654	0.9770

Table No. 5: Combined evaluation of the models (the asterisk (*) designates the overall winner model)

ResNet Mixed Convolution with pre-training emerged as the top model for both classes with pathology (LGG and HGG) and earned a similar score to the other models when categorising healthy brain MRIs. This model was also the clear overall winner based on macro and weighted F1 scores. The spatiospatial models performed better with pre-training, but ResNet 3D performed better without pre-training, as can also be shown [3].

VII. COMPARISON AGAINST LITERATURE

This subsection compares seven more research articles that categorised LGG and HGG tumours against the top model from the preceding subsection (ResNet Mixed Convolution with pre-training). Since mean test accuracy was the most often used statistic in those articles, it was utilised as the metric to compare the outcomes [7].

Beginning with Shahzadi et al.31, who employed T2-FLAIR images from the BraTS 2015 dataset and LSTM-CNN to distinguish between HGG and LGG. Their research focused on employing a smaller sample size, and they were successful in achieving an accuracy rate of 84.000%31. Pei et al.9, who used all of the contrasts in the BraTS dataset and segmented their data using a model akin to the U-Net before doing classification, nonetheless only managed to reach a classification accuracy of 74.9 percent. **Ge et al** strategy of concurrently training several streams utilising multiple contrasts is new. On all the contrasts, their model had an overall accuracy of 90.87 percent, and on T1ce, it had an accuracy of 83.73 percent. Deep convolutional neural networks

were used by **Mzoughi et al.** [8] to reach 96.59 percent on T1ce pictures. It is challenging to compare their conclusions to other research because their study only provides the overall accuracy of their model as a metric for their findings. Using pre-trained GoogLeNet on 2D pictures, **Yang et al. [7]** carried out subsequent research, attaining an overall accuracy of 94.5 percent. Although they did not utilise the BraTS dataset, the goal of their work was to categorise glioma tumours according to LGG and HGG grading.

In their article, Ouerghi et al.11 employed a variety of machine learning techniques to train on fusion pictures, including the random forest technique, on which they were able to classify high-grade and low-grade gliomas with an accuracy of 96.5 percent. Finally, **Zhuge et al.** [3] surpassed the suggested model by 0.12 percent and reached an outstanding 97.1 percent utilising Deep CNN for classification of glioma based on LGG and HGG grading. This discrepancy can be attributed to two factors: 1) their use of BraTS 2018 in conjunction with an extra dataset from The Cancer Imaging Archive, and 2) their use of four distinct contrasts, both of which greatly expand the training set. Furthermore, their publication has no reports of cross-validation. The entire comparison data are displayed in Table 6.

CONCLUSION

This study demonstrates how ResNet (2+1) D and ResNet Mixed Convolution, acting as spatio-spatial models, might enhance the classification of brain tumour grades (i.e. low-grade and high-grade glioma), as well as categorising brain pictures with and without tumours, while lowering the computing costs. The performance of the spatio-spatial models was compared to a pure 3D convolution model using a 3D ResNet18 model. To examine the efficacy of pre-training in this configuration, each of the three models was trained from scratch as well as utilising weights from pre-trained models that were trained on an action recognition dataset. Three fold cross-validation was used to produce the final findings. Despite having fewer trainable parameters, it was shown that the spatio-spatial models outperformed a pure 3D convolutional ResNet18 model in terms of performance.

Further observation reveals that pre-training enhanced the models' functionality. Overall, the pre-trained ResNet Mixed Convolution model was shown to be the best model in terms of F1-score, attaining 0.8949 and 0.9123 F1-scores for low-grade glioma and high-grade glioma, respectively, and a macro F1-score of 0.9345 and a mean test accuracy of 96.98 percent. This research demonstrates the potential of spatio-spatial models to outperform a fully 3D convolutional model.
This was only demonstrated here for one job, the categorization of brain tumours, and one dataset, BraTS. In the future, these models should be contrasted for different tasks to reach an agreement on the spatio-spatial models. This study's use of solely T1 contrast-enhanced pictures for tumour classification, which previously produced high accuracy, is a drawback. The model may perform even better if it incorporates any one or more of the four accessible picture types (T1, T1ce, T2, T2-Flair).

REFRENCE

[1]. Goodenberger, M. L. et al. Genetics of adult glioma. Cancer genetics 205, 613–621 (2012).
[2]. Engelhorn, T. et al. Cellular characterization of the peritumoral edema zone in malignant brain tumors. Cancer science 100, 1856–1862 (2009).
[3]. Rajpurkar, P. et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the chexnext algorithm to practicing radiologists. PLoS medicine 15, e1002686 (2018).
[4]. Pei, L. et al. Brain tumor classification using 3d convolutional neural network. In International MICCAI brainlesion workshop, 335–342 (2019).
[5]. Ge, C. et al. Deep learning and multi-sensor fusion for glioma classification using multistream 2d convolutional networks. In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 5894–5897 (2018).
[6]. Sarasaen, C. et al. Fine-tuning deep learning model parameters for improved super-resolution of dynamic mri with prior-knowledge. Artif. Intell. Medicine 121, 102196 (2021).
[7]. Pallud, J. et al. Quantitative morphological magnetic resonance imaging follow-up of low-grade glioma: a plea for systematic measurement of growth rates. Neurosurgery 71, 729–740 (2012).
[8]. Pérez-García, F. et al. Torchio: a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. Comput. Methods Programs Biomed. 106236 (2021).
[9]. Bakas, S. et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and

overall survival prediction in the brats challenge. arXiv preprint arXiv:1811.02629 (2018).

[10]. Isensee, F. et al. nnu-net: Self-adapting framework for u-net-based medical image segmentation. arXiv preprint arXiv:1809.10486 (2018).

[11]. Zhuge, Y. et al. Automated glioma grading on conventional mri images using deep convolutional neural networks. Med. physics 47, 3044–3053 (2020).

[12]. Ixi dataset. https://brain-development.org/ixi-dataset. (Accessed on 15th December 2021).

[13]. Yang, Y. et al. Glioma grading on conventional mr images: a deep learning study with transfer learning. Front. Neuroscience 12, 804 (2018).

[14]. Isensee, F. et al. nnu-net: Self-adapting framework for u-net-based medical image segmentation. arXiv preprint arXiv:1809.10486 (2018).

[15]. Engelhorn, T. et al. Cellular characterization of the peritumoral edema zone in malignant brain tumors. Cancer science 100, 1856–1862 (2009).

[16]. Rajpurkar, P. et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the chexnext algorithm to practicing radiologists. PLoS medicine 15, e1002686 (2018).

[17]. Pei, L. et al. Brain tumor classification using 3d convolutional neural network. In International MICCAI brainlesion workshop, 335–342 (2019).

[18]. Ge, C. et al. Deep learning and multi-sensor fusion for glioma classification using multistream 2d convolutional networks. In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 5894–5897 (2018).

[19]. Pallud, J. et al. Quantitative morphological magnetic resonance imaging follow-up of low-grade glioma: a plea for systematic measurement of growth rates. Neurosurgery 71, 729–740 (2012)

[20]. Menze, B. H. et al. The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging 34, 1993–2024 (2014).