Reimagining Spatiotemporal Models for Volumetric MRI: Classifying Brain Tumor with Spatiospatial Deep Learning

Dr. Vijaya J[1*], Harshvardhan Sharma[2†], Shantanu Gupta[3†], Avani Gajallewar[4†]

[1*]Department of DSAI, IIIT Naya Raipur, Raipur, 493661, Chhattisgarh, India.
[2,3,4]Department of DSAI, IIIT Naya Raipur, Raipur, 493661, Chhattisgarh, India.

*Corresponding author(s). E-mail(s): vijaya@iiitnr.edu.in;
Contributing authors: harshvardhan22102@iiitnr.edu.in;
shantanu22102@iiitnr.edu.in; avani22101@iiitnr.edu.in;
†These authors contributed equally to this work.

Abstract

This study focuses on the classification of brain tumors using advanced deep learning techniques applied to 3D MRI scans. Gliomas, highly invasive and life-threatening brain tumors, present significant challenges in accurate diagnosis and treatment planning. While traditional approaches rely on either 2D models that analyze individual MRI slices or 3D models that process entire volumetric data, this research explores a novel approach by leveraging spatiotemporal deep learning models adapted as spatiospatial models for tumor classification. By interpreting the slice dimension in MRI scans as a pseudo-temporal axis, these models efficiently capture both in-plane spatial features and inter-slice correlations. To distinguish high-grade gliomas from healthy brain tissue and categorize them into low-grade kinds, two architectures are used: ResNet (2+1)D and ResNet Mixed Convolution. Metrics including classification accuracy, computational efficiency, and the models' capacity to extract intricate characteristics from volumetric data are used to assess them. The study also looks at the advantages of model pre-training for transfer learning. This approach demonstrates the potential of spatiospatial models in medical imaging, offering a promising direction for improving brain tumor classification and supporting clinical decision-making.

Keywords: Convolutional Neural Network, ResNet, Spatiotemporal, Spatiospatial

1 Introduction

Brain tumors are caused by aberrant cell proliferation in the brain, which makes identification and therapy extremely difficult. These tumors are often classified as either benign or malignant depending on their pace of development and chance of recurrence after therapy. Benign tumors develop more slowly, are less likely to recur following therapy, and are not malignant. Conversely, malignant tumors consist of cancerous cells capable of invading nearby tissues or spreading to distant parts of the body via metastasis. Among malignant brain tumors, gliomas are particularly significant, arising from mutations in glial cells that lead to uncontrolled growth. Gliomas account for approximately 30% of all brain and central nervous system tumors and represent 80% of malignant brain tumors. These tumors are commonly associated with astrocytomas, which develop in the brain or spinal cord. The World Health Organization (WHO) classifies gliomas into Low-Grade Gliomas (LGG) and High-Grade Gliomas (HGG) based on their aggressiveness. LGG, being less invasive, are generally surgically treatable. In contrast, HGG are highly invasive and challenging to remove due to their tendency to infiltrate surrounding tissues, making accurate classification crucial for guiding treatment strategies. In the domain of medical imaging, tumor classification tasks often leverage deep learning models. Conventional approaches include 2D convolutional models, which process individual slices of volumetric MRI data, and 3D convolutional models, which operate on the full volumetric data to capture spatial patterns across all dimensions. While 3D convolutional networks are commonly used for volumetric classification tasks, their application is computationally intensive. Alternatively, spatiotemporal models, originally developed for video analysis, offer an efficient approach by combining spatial and temporal convolutional operations. These models can be adapted for volumetric data by treating the slice axis as a pseudo-temporal dimension, effectively transforming them into "spatiospatial" models. Spatiospatial models hold promise for tumor classification tasks, as they can exploit the advantages of 3D convolutional networks while reducing computational complexity. The use of 3D filters in these models enables the capture of intricate features across all dimensions, addressing the significant morphological variability of gliomas. On the other hand, 2D convolution filters excel in extracting spatial features from individual slices. This combination allows spatiospatial models to learn robust representations of tumor patterns both within slices and across volumes. This study investigates two spatiospatial deep learning models: ResNet (2+1)D and ResNet with Mixed Convolution. These models are compared against a baseline ResNet3D architecture to evaluate their effectiveness in classifying gliomas and distinguishing between high-grade, low-grade, and healthy brain tissue. The research aims to explore the potential of spatiospatial models in improving classification accuracy while addressing the challenges associated with volumetric medical imaging.

2 Motivation

One of the most serious and potentially fatal medical conditions in the world is brain tumors. Because of these tumors' complexity and heterogeneity, early diagnosis and precise classification are essential to effective treatment and patient survival. The ability to distinguish between benign and malignant tumor types is still a major challenge in clinical practice, despite advancements in medical imaging technologies. The main causes of this are the drawbacks of manual diagnosis and conventional image processing methods, which can be laborious and prone to human error. The main driving force behind this project is the need for more precise, effective, and scalable diagnostic techniques. A non-invasive, incredibly detailed imaging technique called Magnetic Resonance Imaging (MRI) is widely used to diagnose brain tumors. However, it can be challenging to analyze MRI scans to find minute variations in tumor growth and structure, particularly when tumors share features. Due to the inherent variability in brain tumor appearance, manual diagnosis heavily depends on the expertise of radiologists, and even experienced professionals can occasionally make inconsistent interpretations. The issue is made even more urgent in settings with limited resources and a shortage of specialized radiologists, which can result in delayed or inaccurate diagnoses. This calls for the creation of intelligent, automated systems that can help with accurate, early tumor detection. A potential remedy for this issue is the recent success of deep learning models in a number of medical imaging domains. In particular, Convolutional neural networks (CNNs) have demonstrated a remarkable ability to accurately analyze and classify medical images. Traditional CNNs, on the other hand, mostly concentrate on spatial features, which, while helpful, might not adequately represent the temporal changes that brain tumors undergo over time. An innovative step forward is the addition of spatiotemporal models, which allow the system to analyze the temporal dynamics and spatial structures of tumors. The model can offer more profound insights into tumor growth patterns and morphological changes by integrating temporal information. These features are crucial for distinguishing between benign and malignant tumors, as well as different grades of gliomas. The creation of this deep spatiotemporal model is in line with an expanding body of research in medical AI that aims to integrate various data modalities for more thorough diagnosis. This project was motivated by the realization that many of the diagnostic tools available today do not offer the required level of analysis depth. We hope to close this gap by providing a more dependable and robust approach to brain tumor classification with the suggested model. Better patient outcomes could result from the ability to automatically classify tumors based on MRI scans, which could also minimize human error, speed up diagnosis, and enhance treatment planning. Furthermore, this project could make high-quality diagnostic tools more accessible to all. This model can be used as a primary diagnostic tool in locations without access to specialized radiologists, or it can be used as a second opinion in healthcare systems, particularly in under-resourced areas. This could provide more equitable access to life-saving diagnostic technologies and lessen regional disparities in healthcare delivery.

3 Contributions

This research explores an innovative approach to classifying 3D volumetric MRI scans using spatiotemporal models traditionally applied to video analysis. These models typically process data along three dimensions, handling spatial information across two axes and temporal data along the third. Examples include architectures such as ResNet(2+1)D and ResNet Mixed Convolution. However, the application of these models to inherently 3D medical imaging data, like MRI scans, remains relatively unexplored. This study seeks to adapt spatiotemporal models by reinterpreting the third spatial dimension of MRI scans (the slice axis) as distinct from the other two, effectively transforming these frameworks into "spatiospatial" models tailored for medical image classification. The primary objective is to investigate the performance of these modified spatiotemporal models in the categorization of gliomas, which are brain tumors that are graded according to their aggressiveness. Using 3D MRI volumes, these spatiospatial models will classify the brain tumors into several glioma kinds, ranging from low-grade to highly aggressive glioblastomas, and also distinguish between healthy and tumor-affected brains. These models have the potential to capture small alterations in tumor architecture that span numerous slices by handling the slice dimension in MRI data differently. Unlike conventional 3D convolutional models like ResNet3D, which handle all three dimensions equally, this method takes a different approach. These spatiospatial models' performance will be compared to that of the pure 3D convolutional model (ResNet3D) in order to assess how effective they are.

4 Problem Definition

In medical imaging and neuro-oncology, accurately classifying brain cancers from 3D MRI data is a major difficulty. The bulk of malignant brain tumors are gliomas, which vary greatly in their composition, course, and degree of intensity. The World Health Organization (WHO) grading system they use to classify patients is crucial for predicting patient outcomes and choosing the best course of therapy. However, because to the intricacy of tumor shape, traditional diagnostic approaches are frequently subjective, time-consuming, and error-prone, which can result in delayed or erroneous diagnosis. Tumor classification challenges have made extensive use of both pure 3D convolutional models, such ResNet3D, and conventional 2D image analysis techniques. Although these techniques are promising, it's possible that they won't be able to capture fine structural information in several 3D MRI volume slices. In 3D convolutional models, treating all three dimensions equally can result in the loss of important information, particularly when dealing with complicated malignancies such as gliomas that exhibit variability in size, shape, and position across distinct brain planes. Furthermore, the inability of current methods to accurately distinguish between tumors of varying grades and healthy brain tissue emphasizes the need for more sophisticated methods. In order to overcome these issues, this study looks into the application of spatiotemporal models—which were first created for the classification of videos—in a new setting: the classification of brain tumors using 3D volumetric MRI data. The study investigates how to modify spatiotemporal models to become "spatiospatial" models by handling one dimension (the slice dimension) differently from the other two spatial

dimensions. By contrasting the outcomes of these spatiospatial models with those of traditional 3D convolutional models like ResNet3D, the study seeks to evaluate how well these models perform in categorizing various forms of gliomas and differentiating them from healthy brains. Furthermore, the models will be assessed both with and without pre-training in order to investigate the advantages of transfer learning in this field. Ultimately, the problem addressed in this project is how to develop a more accurate, efficient, and reliable method for classifying brain tumors from 3D MRI images, leveraging deep learning to improve diagnostic precision and treatment planning.

5 Proposed Model

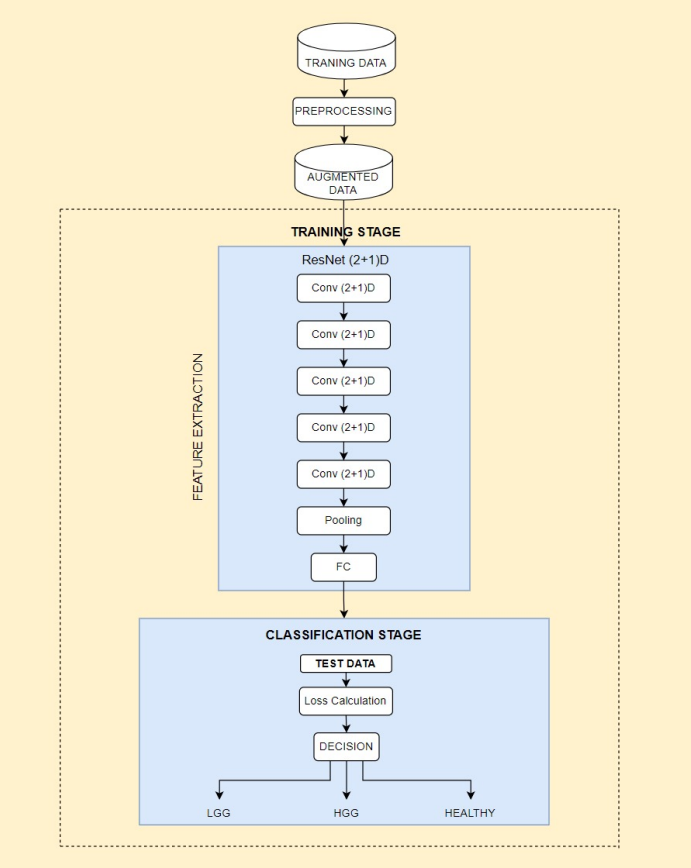

Fig. 1 Project Workflow

In contrast to conventional 3D convolution-based models, spatiotemporal models, which are frequently employed for video analysis, handle data in two spatial dimensions and one temporal dimension. Since 3D volumetric picture categorization does not involve a temporal component, 3D convolutional networks are frequently the best option. As an alternative, volumetric data may be divided into 2D slices, and each slice can be examined separately using 2D convolutional models. For tumor classification, particularly gliomas, the morphological complexity of these tumors makes 3D filters highly advantageous. These filters enable convolution kernels to discern intricate patterns across all dimensions, making them sensitive to tissue variations throughout the volume. In contrast, 2D filters focus on extracting spatial features within individual slices, offering complementary insights. Spatiotemporal models combine different convolutional approaches within a unified framework, presenting potential advantages such as simplifying model complexity or incorporating additional nonlinearity. When applied to volumetric medical data, these models can be adapted as spatiospatial models. In this adaptation, the slice axis of 3D MRI data is treated as a pseudo-temporal dimension, while the in-plane axes represent spatial dimensions. This approach leverages the strengths of spatiotemporal architectures while addressing the unique requirements of 3D medical imaging tasks, forming the foundation for their use in tumor classification.

6 Literature Review

REFERENCE	TITLE	AUTHORS	MODEL USED	ACCURACY
1	Brain Tumor Classification Using MRI Images and Convolutional Neural Networks	M. A. Hafeez, C. B. Kayasandik and M. Y. Dogan (2022)	Convolutional Neural Networks	Competitive performance achieved
2	Deep Learning Based Brain Tumor Detection and Classification	N. M. Dipu, S. A. Shohan and K. M. A. Salam (2021)	You Only Look Once(YOLO) and FastAI library	Varied across techniques
3	Classification of brain tumours in MR images using deep spatio-spatial models	Chatterjee, S., Nizamani, F. A., & Nürnberger, A. (2022)	Residual Networks (2+1)D, Mixed convolutions	State of the art results
4	Brain Tumor Classification of MRI Scans using Deep Learning Techniques	A. Agrawal and V. Maan (2024)	VGGNet-16, EfficientNet, ResNet-50	VGGNet-16 performed well among others
5	Deep Learning for Brain Tumor Classification from MRI Images	S. Arora and M. Sharma (2021)	SVM Classifier, Random Forest Classifier, VGG – 16, Inception-V3	Varied across methods
6	Deep Transfer Learning Based Multi-Class Brain Tumors Classification Using MRI Images	M. Mondal, M. F. Faruk, N. Raihan and P. Ahammed (2021)	VGG-19 transfer learning model	Outperformed in multi class classification than existing results
7	Classification and Detection of Brain Tumors from Magnetic Resonance Imaging Scans using Deep Transfer-Learning	BeebiNaseeba, S. B. Nikhil, N. S. Nair, AsutoshDoppalapudi, N. P. Challa and R. Katta, (2023)	VGG-19, Resnet101, Inception-Resnet-v2	Varied across methods
8	An Efficient Brain tumor classification using CNN and transfer learning	P. Sasikumar, S. Cherukuvada, P. Balmurugan, P. Vijay Anand, S. Brindasri and R. Nareshkumar	Convolutional Neural Networks and transfer learning	Transfer learning performed better

Fig. 2 Literature Survey of 8 recent Research Papers

The table provides a comprehensive overview of recent research on brain tumor classification using MRI images and deep learning techniques. It highlights the authors, models employed, and reported performance, showcasing the advancements and diverse methodologies in this critical area of medical imaging. The research collectively explores both traditional and state-of-the-art deep learning models, offering insights into their strengths and limitations for tumor classification. Convolutional Neural Networks (CNNs) have been a foundational approach in many studies. Hafeez et al. demonstrated the competitive performance of CNNs in classifying brain tumors from MRI scans. Similarly, Sasikumar et al. combined CNNs with transfer learning techniques, which outperformed traditional approaches. Transfer learning emerged as

a dominant trend across the studies, particularly when applied to sophisticated architectures like VGGNet, ResNet, and Inception models. Mondal et al., for instance, utilized VGG-19 with transfer learning and reported superior performance in multiclass brain tumor classification compared to conventional methods. Beebi Naseeba et al. further explored transfer learning with ResNet-101, VGG-19, and Inception-ResNet-v2, observing that its application consistently improved classification accuracy. Several studies investigated advanced architectures for enhanced classification. Chatterjee et al. employed residual networks (2+1D) and mixed convolutions, achieving state-of-the-art results in brain tumor detection and classification. Agrawal and Maan compared different architectures, including VGGNet-16, EfficientNet, and ResNet-50. Their findings indicated that VGGNet-16 outperformed others, emphasizing the importance of selecting appropriate model architectures for specific classification tasks. The performance of certain models varied depending on the techniques employed and the dataset characteristics. Dipu et al. utilized YOLO alongside the FastAI library, finding that accuracy differed across techniques. Arora and Sharma integrated traditional machine learning models, such as SVM and Random Forest, with deep learning approaches like VGG-16 and Inception-V3. Their results varied significantly, underlining the influence of combining traditional and modern methods on classification accuracy. Transfer learning's widespread success across studies highlights its capability to address challenges posed by limited datasets, a common issue in medical imaging. This technique enables the application of pre-trained models to new tasks, reducing training time while enhancing performance. The studies consistently demonstrate that leveraging transfer learning with deep neural networks not only improves classification accuracy but also reduces computational overhead. Overall, the table highlights significant progress in brain tumor classification through the adoption of deep learning models. Transfer learning, in particular, has emerged as a crucial strategy for achieving high accuracy and robust performance. The findings suggest that combining advanced architectures with transfer learning can yield significant improvements over traditional methods. Moreover, the observed variations in accuracy across different studies emphasize the importance of tailoring model selection and optimization techniques to the specific characteristics of the dataset and classification task. This body of research demonstrates the potential of deep learning to revolutionize medical diagnosis, offering efficient and reliable tools for brain tumor detection and classification. It also underscores the need for continued exploration of emerging architectures and hybrid approaches to further advance the field and address existing challenges.

7 Existing Technology

7.1 ResNet3D

3D-CNNs are advanced deep learning models designed for tasks like image segmentation, particularly in analyzing volumetric medical images and video sequences. By extending traditional 2D-CNNs into a third dimension, 3D-CNNs can process volumetric data, capturing both spatial and temporal dependencies. Unlike 2D-CNNs, which analyze individual slices, 3D-CNNs consider MRI scans as complete volumes, extracting spatial features along the x, y, and z axes. The process involves sliding

3D convolutional filters across the volumetric data to identify patterns and extract meaningful spatial features. This is followed by downsampling through pooling layers, which reduce dimensionality while retaining critical information. The extracted features are then combined through fully connected layers for classification. Several 3D-CNN architectures are available, with ResNet 3D being one of the most popular, alongside V-Net, U-Net, and DenseNet. ResNet 3D addresses challenges like vanishing and exploding gradients, enabling the training of much deeper networks. This capability is crucial for capturing complex spatial relationships in high-dimensional data. Pure 3D models demonstrate excellent performance in segmentation tasks, effectively leveraging spatial dependencies inherent in volumetric data. Their ability to comprehensively analyze spatial relationships makes them particularly suitable for medical imaging applications, where capturing fine details and 3D structures is essential.

7.2 ResNet Mixed Convolution

This model employs a hybrid approach, combining both 2D and 3D convolutional layers to enhance feature extraction. The architecture begins with a stem containing a 3D convolutional layer. Following the stem, the model integrates one 3D convolutional block, succeeded by three 2D convolutional blocks. Each convolutional block comprises two residual blocks, and each residual block contains a pair of convolutional layers. To ensure seamless spatial-temporal feature learning, the residual blocks in all convolutional blocks—except for the first—are separated by two 3D convolutional layers. Each convolutional layer, whether 2D or 3D, is followed by a 3D batch normalization layer and a ReLU activation function, which help stabilize training and improve non-linearity. This combination of 2D and 3D convolutions enables the model to capture diverse features: 3D convolutions extract spatial information across volumetric data, while 2D convolutions focus on extracting intricate patterns within individual slices. After the series of convolutional blocks, the architecture incorporates pooling layers, dropout, and fully connected layers for dimensionality reduction, regularization, and classification. This comprehensive approach ensures the model effectively processes MRI scans, leveraging both the volumetric and slice-level information to produce highly accurate results. The hybrid design offers a unique advantage by overcoming common challenges, such as skipped connections, and achieves a remarkable classification accuracy of approximately 96%. The integration of 3D and 2D convolutions ensures the model fully traverses the spatial dimensions of the MRI data while retaining fine-grained details from each slice. This combination not only enhances the representation of tumor-related features but also ensures robustness and generalization. The result is a model well-suited for medical imaging applications, providing an excellent analysis of MRI scans while effectively extracting and integrating features across multiple dimensions.

7.3 RNN-LSTM Based

Recurrent Neural Networks (RNNs) combined with Long Short-Term Memory (LSTM) networks are highly effective for analyzing sequential or temporal data. This combination is particularly advantageous in MRI-based brain tumor diagnosis, where

the model can process a series of 2D slices derived from volumetric MRI scans. By leveraging both spatial information within individual slices and sequential relationships across slices, the approach captures intricate patterns in the data. The process begins by converting MRI images into 2D slices. Spatial features from each slice are extracted using a Convolutional Neural Network (CNN), which serves as the backbone for identifying meaningful patterns in the images. These feature vectors, representing the spatial characteristics of the slices, are then fed into the LSTM network. The LSTM is designed to handle sequential data effectively by employing input, forget, and output gates. These gates manage the flow of information, allowing the model to retain or discard specific slice-to-slice relationships as needed, ensuring critical temporal patterns are captured. Once the sequence is processed, the LSTM outputs a representation that is passed through a fully connected layer, which performs the final classification. The model classifies the tumor as either a low-grade glioma (LGG) or a high-grade glioma (HGG). The overall accuracy achieved by this approach is approximately 91%, highlighting its effectiveness. However, the model's computational demands are significant, given the complexity of both the CNN-based feature extraction and the LSTM's sequential processing. Despite these challenges, the integration of CNNs and LSTMs offers a promising method for improving diagnostic accuracy in the medical imaging domain.

7.4 Support Vector Machines

Support Vector Machines (SVMs) are a fundamental machine learning algorithm widely used for both binary and multiclass classification tasks. In the domain of brain tumor classification, SVMs have proven to be highly effective due to their ability to handle datasets with both linear and nonlinear relationships. They achieve this by mapping the input data into a higher-dimensional feature space, where the separation between classes becomes more distinct and effective. To prepare the data for classification, Histogram of Oriented Gradients (HOG) is utilized to extract relevant features from the images. Since the high-dimensional feature data generated by HOG is unsuitable for direct input into an SVM classifier, Principal Component Analysis (PCA) is applied. PCA reduces the dimensionality of the data while preserving essential features, making the data more manageable for the classifier. Following this preprocessing step, the model classifies the input data into three categories: glioma, meningioma, or no tumor. The SVM-based approach demonstrates strong performance and is notably robust against overfitting, even when working with limited datasets. However, the computational cost increases significantly when dealing with large datasets due to the complexity of both feature extraction and the classification process. Despite these challenges, SVMs remain a reliable and versatile choice for brain tumor classification tasks, offering a balance between accuracy and generalization.

7.5 Autoencoder for Feature Learning

Autoencoders are powerful unsupervised deep learning models designed to learn compact representations of input data. In the context of brain tumor analysis, the process begins with preprocessing MRI scans, including normalization and segmentation, to

isolate and focus on tumor regions. The autoencoder consists of two main components: an encoder and a decoder. The encoder compresses the high-dimensional MRI data into a compact latent representation, effectively capturing meaningful and discriminative features that reflect the spatial and intensity characteristics of the tumor. This feature extraction occurs in an unsupervised manner, enabling the model to uncover underlying patterns in the data without requiring explicit labels. The decoder then reconstructs the original input from these latent features, ensuring the encoding retains essential tumor-related information. Once the autoencoder is trained, the latent features generated by the encoder are fed into a classifier, such as a neural network or Support Vector Machine (SVM). This classifier is tasked with categorizing the tumor as either a low-grade glioma (LGG) or a high-grade glioma (HGG). By leveraging the autoencoder's ability to handle complex, high-dimensional MRI data, this method provides a robust solution for extracting tumor-specific features. These features significantly enhance the accuracy of tumor classification, making the approach well-suited for medical imaging applications.

8 Experimental Setup

8.1 ResNet(2+1)D

The ResNet (2+1)D model is an advanced architecture originally designed for video data, where it separates spatial and temporal convolutions, making it distinct from traditional 3D convolutional networks. This model effectively processes data by decomposing 3D convolutions into two operations: a 2D convolution to capture spatial information within frames (or slices) and a 1D convolution to handle inter-frame (or inter-slice) relationships. This separation allows the model to learn spatial and slice-wise correlations independently, improving its ability to capture complex patterns while reducing computational complexity. The ResNet (2+1)D model used in this study begins with a "stem" containing a 2D convolutional layer with a kernel size of 7×7 and a stride of 2. This layer accepts a single input channel and produces 45 output channels, effectively extracting coarse spatial features. The channel depth is increased to 64 by passing the output through a 1D convolution with a kernel size of 3 and a stride of 1. Four convolutional blocks, each with two residual blocks, make up the architecture after the stem. A 1D convolution that records slice-wise associations is applied after a 2D convolution that processes spatial data in each residual block. A ReLU activation function and 3D batch normalization come after both convolution layers in each block, adding non-linearity and guaranteeing stability during training. Two 3D convolutional layers with a kernel size of 1 and a stride of 2 are used to divide the residual blocks in every convolutional block save the first one in order to downsample the feature maps. An adaptive average pooling layer at the end of the model reduces the output dimensions to $1 \times 1 \times 1$. For volumetric classification problems like brain tumor identification, the ResNet (2+1)D model is quite successful since it adds a dropout layer and a fully connected layer to generate class predictions.

Algorithm 1 Classification of Brain Tumor using ResNet(2+1)D for MRI Scans

Require: Input MRI volume $X \in \mathbb{R}^{D \times H \times W}$, where D is depth, H is height, and W is width. Number of classes C. ResNet(2+1)D model parameters θ. Number of epochs N_{epochs}. Learning rate η. Data loader DataLoader containing N samples.

Ensure: Trained model parameters θ^* and classification probabilities for each class.

Initialization:

- Initialize model parameters θ.
- Define Cross-Entropy Loss $\mathcal{L}$.
- Initialize Adam optimizer with learning rate η.

for epoch = 1 to N_{epochs} **do**
  Set model to training mode.
  Initialize epoch loss: epoch_loss $\leftarrow 0$.
  for each batch B in DataLoader **do**
    Extract inputs X_B and labels y_B from batch B.
    Move X_B and y_B to the computation device (e.g., GPU).
    Perform **forward pass**:

$$P_B = \text{Softmax}(\text{ResNet}(X_B; \theta)) \quad \text{where } P_B \in \mathbb{R}^{|B| \times C}.$$

    Compute loss $\mathcal{L}_B$ using Cross-Entropy Loss:

$$\mathcal{L}_B = -\frac{1}{|B|} \sum_{i=1}^{|B|} \sum_{j=1}^{C} y_{ij} \log(p_{ij}),$$

    where y_{ij} is the one-hot label and p_{ij} is the predicted probability.
    Perform **backward pass** to compute gradients:

$$\nabla_\theta \mathcal{L}_B.$$

    Update parameters θ using Adam optimizer:

$$\theta \leftarrow \theta - \eta \cdot \nabla_\theta \mathcal{L}_B.$$

    Accumulate batch loss into epoch loss:

$$\text{epoch\_loss} \leftarrow \text{epoch\_loss} + \mathcal{L}_B.$$

  end for
  Compute average loss for the epoch:

$$\text{epoch\_loss} \leftarrow \text{epoch\_loss}/\text{Number of Batches}.$$

  Print epoch loss.
end for

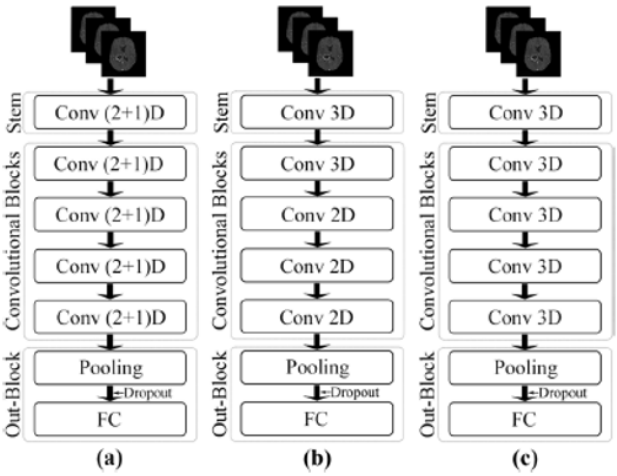

Fig. 3 Representation of network architectures: (a) ResNet(2+1)D, (b) ResNet Mixed Convolution, and (c) ResNet3D

8.2 Equations

8.2.1 Input Representation

Each MRI scan is represented as a 4D tensor:

$$\mathbf{X} \in \mathbb{R}^{C \times D \times H \times W}$$

where:

- C: Number of channels (e.g., 1 for grayscale MRI scans),
- D, H, W: Depth (slices), height, and width of the scan.

8.2.2 ResNet(2+1)D Convolution

The key innovation in R(2+1)D is the factorization of a 3D convolution into:

- 2D spatial convolution: Operates on height and width.
- 1D temporal convolution: Operates on the depth (slices).

8.2.3 2D Spatial Convolution

Given a 3D kernel size (k_t, k_h, k_w), the spatial convolution is:

$$\mathbf{Y}_{\text{spatial}}(c, d, i, j) = \sum_{m=1}^{C} \sum_{u=-\frac{k_h}{2}}^{\frac{k_h}{2}} \sum_{v=-\frac{k_w}{2}}^{\frac{k_w}{2}} \mathbf{W}_{\text{spatial}}(c, m, u, v) \cdot \mathbf{X}(m, d, i+u, j+v)$$

where $\mathbf{W}_{\text{spatial}}$ represents the spatial kernel weights.

8.2.4 1D Temporal Convolution

The temporal convolution then operates on the output of the spatial convolution:

$$\mathbf{Y}(c, d, i, j) = \sum_{d'=-k_t/2}^{k_t/2} \mathbf{W}_{\text{temporal}}(c, d') \cdot \mathbf{Y}_{\text{spatial}}(c, d + d', i, j)$$

where $\mathbf{W}_{\text{temporal}}$ represents the temporal kernel weights.

8.2.5 Residual Connections

ResNet introduces skip connections to mitigate vanishing gradient issues:

$$\mathbf{Y}_{\text{res}} = \mathbf{Y} + \mathbf{X}$$

8.2.6 Downsampling and Feature Extraction

Pooling Downsampling reduces the feature map dimensions (spatial and/or temporal):

$$\mathbf{Y}_{\text{pooled}}(c, d', i', j') = \text{Max/Avg}\left(\mathbf{Y}(c, d, i, j)\right)$$

This step preserves significant features while reducing dimensionality.

8.2.7 Fully Connected Layer for Classification

After several R(2+1)D layers, the output is globally pooled:

$$\mathbf{z} = \frac{1}{D \cdot H \cdot W} \sum_{d=1}^{D} \sum_{i=1}^{H} \sum_{j=1}^{W} \mathbf{Y}(c, d, i, j)$$

The pooled output is then passed through a fully connected layer:

$$\mathbf{o} = \mathbf{W}_{\text{fc}} \cdot \mathbf{z} + \mathbf{b}_{\text{fc}}$$

where $\mathbf{W}_{\text{fc}}$ and $\mathbf{b}_{\text{fc}}$ are the weights and biases of the fully connected layer.

8.3 Dataset Description

Two datasets were used in this study: non-pathological pictures from the IXI dataset and diseased brain images from the Brain Tumor Segmentation (BraTS) dataset. While the IXI dataset contains healthy brain MRI scans, the BraTS dataset offers brain MRI images with four distinct contrasts. T1 contrast-enhanced (T1ce) images are the most commonly employed of the four MRI modalities found in the BraTS dataset for single-contrast tumor categorization. The BraTS dataset, which included 260 volumes of Glioblastoma Multiforme (high-grade glioma, or HGG) and 74 volumes of low-grade glioma (LGG), was used to select T1ce images of 334 participants for this study. Furthermore, to ensure that there were as many healthy cases as high-grade

glioma cases, 260 T1-weighted volumes from the IXI dataset were selected at random to represent healthy samples.. The combined dataset of pathological and healthy samples was then split randomly into three folds for training and testing, with an 8:2 ratio. This approach ensured a balanced and robust evaluation of the proposed models across diverse cases.

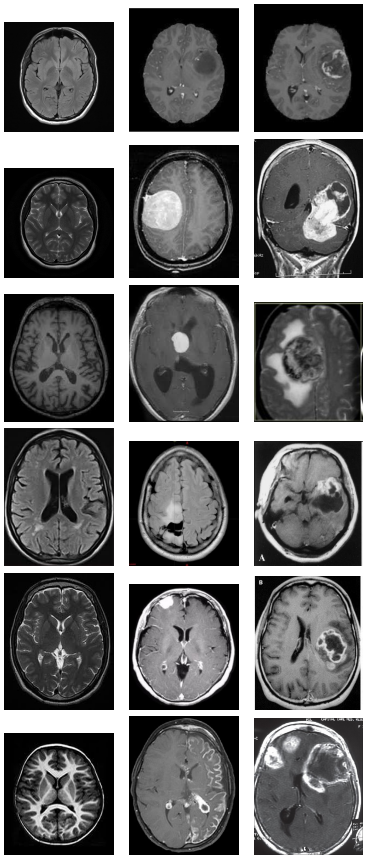

Fig. 4 MRI Images of Healthy Brain, LGG and HGG

8.4 Data Augmentation

Data augmentation was applied to the dataset prior to training to enhance model generalization, utilizing the TorchIO library for implementation. Two augmentation strategies were initially tested: light augmentation and heavy augmentation. Light augmentation involved the application of random affine transformations (scaling in the range of 0.9–1.2 and rotation up to 10 degrees) and random left-right flips with a probability of 0.25. Heavy augmentation incorporated all the transformations from light augmentation along with additional techniques, including elastic deformation and

random k-space transformations, such as motion artifacts, spike noise, and ghosting effects. Initial experiments revealed that heavy augmentation adversely affected model performance. Networks trained on heavily augmented data exhibited lower final accuracy, and the loss convergence was significantly slower compared to models trained with light augmentation. Based on these observations, only light augmentation was employed for all experiments in this study to achieve optimal training efficiency and model performance.

8.5 Evaluation Metrics

To give a thorough assessment of the models' categorization skills, a number of criteria were used to evaluate their performance. The metrics listed below were calculated:

1. Precision: Measures the proportion of true positive predictions out of all positive predictions made by the model, highlighting the model's ability to avoid false positives.
2. Recall: Measures the proportion of true positive predictions out of all actual positive instances, indicating the model's capability to identify positive cases.
3. F1 Score: The harmonic mean of precision and recall, providing a balanced measure of a model's accuracy, especially in datasets with class imbalance.
4. Specificity: Reflects the model's ability to correctly identify negative cases, defined as the proportion of true negatives out of all actual negative instances.
5. Testing Accuracy: The overall proportion of correctly classified instances across all classes.

Additionally, class-wise performance was visualized using a confusion matrix, offering a detailed view of the classification results and highlighting potential misclassifications among the classes. This matrix provided insights into the strengths and weaknesses of each model in distinguishing between high-grade glioma (HGG), low-grade glioma (LGG), and healthy samples. These evaluation metrics, computed across three folds of the dataset, ensured a thorough comparison of the proposed spatiospatial models and the baseline ResNet3D architecture.

8.6 Loss Function

The weight value which is normalized for each class is calculated as:

$$W_{class} = \left[1 - \left(\frac{\text{samples}_{class}}{\sum \text{samples}_{total}}\right)\right]$$

The class loss is then calculated as:

$$\text{Loss}_{class} = W_{class}\left[-x_{class}\log(P(c))\right]$$

where x is the true distribution and P(c) is the probabilistic distribution for the class c. The total loss is the sum of CE loss for each class:

$$L_{\text{total}} = L_{c_1} + L_{c_2} + L_{c_3} + \cdots + L_{c_n}$$

9 Experiment Results and Analysis

The experimental results demonstrate that applying image slicing to 3D brain tumor MRI images significantly enhances the classification capabilities of deep learning models. The ResNet(2+1)D model's strong performance underscores the potential of this approach in clinical applications, paving the way for further advancements in automated brain tumor diagnostics. Future iterations of this work will focus on optimizing the model further and exploring additional techniques to improve classification accuracy and robustness.

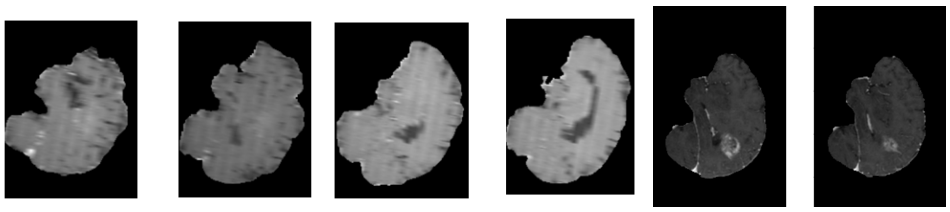

Fig. 5 Slicing of 3D MRI Images of Brain

Table 1 Resnet(2+1)D Model Evaluation

Brain Type	Precision	Recall	F1-Score	Support
Healthy	0.94	1.0	0.97	1743
HGG	0.77	0.89	0.83	777
LGG	0.52	0.43	0.47	314

Table 2 Resnet3D Model Evaluation

Brain Type	Precision	Recall	F1-Score	Support
Healthy	0.98	0.99	0.99	581
HGG	0.78	0.89	0.83	259
LGG	0.49	0.40	0.44	105

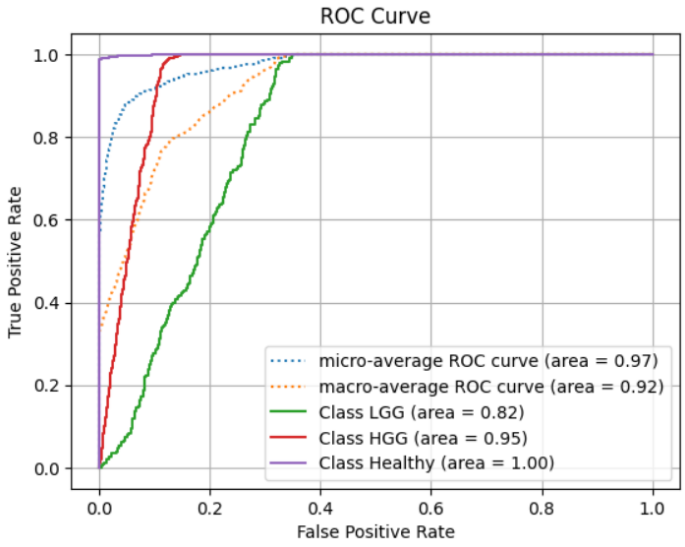

Fig. 6 ROC Curve

10 Acknowledgment

We would like to thank Dr. Vijaya J for giving us the opportunity to work on such an interesting topic in Deep Learning. Also, we would like to thank him for his excellent insights and guidance through out the semester because of which we were able to successfully complete our project. We would also like to extend our thanks to our colleagues who assisted us during our work.

11 Conclusion

In conclusion, we focused on gliomas of both low-grade (LGG) and high-grade (HGG) varieties, and we created a deep learning-based method for classifying brain tumors using MRI data. The appropriate T1-weighted pictures were recovered from the MRI scans in the first stage, which involved pre-processing the dataset. This made sure that the most pertinent data for the identification and categorization of tumors was separated out for additional examination. The complex 3D data was then divided into digestible 2D slices using the volumetric MRI pictures that had been cut into 2D forms. This conversion made processing simpler and gave us the opportunity to use well-established deep learning models—which are typically designed for two-dimensional data—in our work. Then, the neural network models for classification were fed these 2D slices. We used PyTorch to create a neural network architecture for model selection, concentrating on models appropriate for image classification applications. The processed photos were used to train the selected models, which included 3D convolutional networks and spatiospatial networks. In order to maximize the model's performance

and raise its accuracy in categorizing brain tumors into LGG, HGG, or healthy categories during training, loss calculations were carried out for each batch. Overall, the project's methodology contributed to improvements in automated medical diagnostics by successfully demonstrating the viability of classifying brain cancers from MRI scans using deep learning techniques.

12 References

1. P. Sasikumar, S. Cherukuvada, P. Balmurugan, P. Vijay Anand, S. Brindasri and R. Nareshkumar, "An Efficient Brain tumor classification using CNN and transfer learning," 2024.
2. BeebiNaseeba, S. B. Nikhil, N. S. Nair, AsutoshDoppalapudi, N. P. Challa and R. Katta, "Classification and Detection of Brain Tumors from Magnetic Resonance Imaging Scans using Deep Transfer-Learning," 2023.
3. M. Mondal, M. F. Faruk, N. Raihan and P. Ahammed, "Deep Transfer Learning Based Multi-Class Brain Tumors Classification Using MRI Images," 2021.
4. N. K. Kasabov, M. G. Doborjeh and Z. G. Doborjeh, "Mapping, Learning, Visualization, Classification, and Understanding of fMRI Data in the NeuCube Evolving Spatiotemporal Data Machine of Spiking Neural Networks," in IEEE Transactions on Neural Networks and Learning Systems, April 2017.
5. X. Huang, X. Li, Y. Ye, S. Feng, C. Luo and B. Zhang, "On Understanding of Spatiotemporal Prediction Model," in IEEE Transactions on Circuits and Systems for Video Technology, July 2023.
6. R. R. Sharmily, B. Karthik and T. Vijayan, "Brain Tumour Detection and Classification using Deep Learning And Transfer Learning Techniques," 2023.
7. S. Arora and M. Sharma, "Deep Learning for Brain Tumor Classification from MRI Images," 2021.
8. A. Agrawal and V. Maan, "Brain Tumor Classification of MRI Scans using Deep Learning Techniques," 2024.
9. N. M. Dipu, S. A. Shohan and K. M. A. Salam, "Deep Learning Based Brain Tumor Detection and Classification," 2021.
10. A. D, R. R, S. V, V. N and R. R, "Automated Brain Tumor Classification and Detection Using Modified Convolutional Neural Networks for Early Diagnosis," 2023.
11. Nadim Mahmud Dipu, Sifatul Alam Shohan, K. M. A. Salam. "Deep Learning Based Brain Tumor Detection and Classification," 2021 International Conference on Intelligent Technologies (CONIT).
12. Tom Philip Prie, Roshan Jahan, Preetam Suman. "Review of Brain Tumor Segmentation, Detection and Classification Algorithms in fMRI Images," 2018 International Conference on Computational and Characterization Techniques in Engineering Sciences (CCTES).
13. R. Meena Prakash, R. Shantha Selva Kumari. "Classification of MR Brain Images for Detection of Tumor with Transfer Learning from Pre-trained CNN Models," 2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET).

14. Muhammad Adeel Hafeez, Cihan Bilge Kayasandik, Merve Yusra Dogan. "Brain Tumor Classification Using MRI Images and Convolutional Neural Networks," 2022 30th Signal Processing and Communications Applications Conference (SIU).
15. Uvaneshwari M, M. Baskar. "Machine Learning for Brain Tumour Detection: A Systematic Review of Segmentation and Classification Methods," 2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS).
16. T. S. Kumar, K. Rashmi, S. Ramadoss, L. K. Sandhya and T. J. Sangeetha, "Brain tumor detection using SVM classifier," 2017
17. S. K. Baranwal, K. Jaiswal, K. Vaibhav, A. Kumar and R. Srikantaswamy, "Performance analysis of Brain Tumour Image Classification using CNN and SVM," 2020
18. A. Hussain and A. Khunteta, "Semantic Segmentation of Brain Tumor from MRI Images and SVM Classification using GLCM Features," 2020
19. F. P. Polly, S. K. Shil, M. A. Hossain, A. Ayman and Y. M. Jang, "Detection and classification of HGG and LGG brain tumor using machine learning," 2018
20. L. McIntyre and E. Tuba, "Brain Tumor Segmentation and Classification using Texture Features and Support Vector Machine," 2023
21. A. R. Mathew and P. B. Anto, "Tumor detection and classification of MRI brain image using wavelet transform and SVM," 2017
22. E. V. Babaferi, T. M. Fagbola and C. S. Thakur, "DeepCAI-V3: Improved Brain Tumor Classification from Noisy Brain MR Images Using Convolutional Autoencoder and Inception-V3 Architecture," 2024
23. P. Jyothi, S. Dhanasekaran and R. A. Singh, "Multiclass Classification of Brain Tumor for MR Images Using Shallow Autoencoder Based Neural Network," 2022
24. D. Agalya and S. Kamalakkannan, "Detecting Brain Tumor Stages Using Convolutional AutoEncoder (CAE) with Hybrid Deep Learning Method," 2024
25. S. K. Rajeev, M. P. Rajasekaran, R. Krishna Priya and A. Al Bimani, "A Review on Magnetic Resonance Spectroscopy for Clinical Diagnosis of Brain Tumour using Deep Learning," 2021
26. M. Aledhari and R. Razzak, "An Adaptive Segmentation Technique to Detect Brain Tumors Using 2D Unet," 2020
27. M. Suganthi, K. Bagadi, S. Karthiga, S. S. Dash, K. Manjulaadevi and N. Rajkumar, "Enhanced Multi-Class Brain Tumor Classification in MRI Scans Using WPCA-LSTM Network," 2024
28. S. K. Rajeev, M. P. Rajasekaran, K. Ramaraj, G. Vishnuvarthanan, T. Arunprasath and V. Muneeswaran, "A Hybrid CNN-LSTM Network For Brain Tumor Classification Using Transfer Learning," 2023
29. M. S. Fasihi and W. B. Mikhael, "Brain tumor grade classification Using LSTM Neural Networks with Domain Pre-Transforms," 2021
30. I. Shahzadi, T. B. Tang, F. Meriadeau and A. Quyyum, "CNN-LSTM: Cascaded Framework For Brain Tumour Classification," 2018