RESEARCH PAPER

A Hybrid VGG16–ResNet50 Deep Learning Framework for Brain Tumor Detection and Segmentation from MRI

Santhosh M[1*], Nayana M[2], Madhura J[3], Kavyashree G M[4], Bharath B C[5], Shalini K B[5]
[1]*Assistant Professor, CSE, Dayananda Sagar University, Bengaluru, India*
[2]*Assistant Professor, MBA, Sai Vidya Institute of Technology, Bengaluru, India*
[3]*Associate Professor, ISE, Dayananda Sagar College of Engineering, Bengaluru, India*
[4]*Assistant Professor, ISE, Dayananda Sagar Academy of Technology and Management, Bengaluru, India*
[5]*Assistant Professor, ISE, Dayananda Sagar College of Engineering, Bengaluru, India*
Corresponding Author: Santhosh M*

Received: 15th Jul, 2026 | Revised: 25th Jul, 2026 | Accepted: 5th Aug, 2026 | Available Online: 12th Aug, 2026

ABSTRACT
Brain tumors are difficult to assess because diagnosis depends on both the presence of the lesion and its extent within the brain. MRI is commonly used for this assessment because it provides good soft-tissue contrast and shows the location and appearance of abnormal tissue. This study develops an automated framework for brain-tumor detection and segmentation from MRI. The framework combines transfer learning with VGG16 and ResNet50. The experiments described in the manuscript use a Kaggle MRI dataset containing approximately 3,000–4,000 images. The reported tumor-detection accuracy is above 90%.A web interface was also developed so that an MRI image can be submitted and analysed. The system returns the predicted tumor information together with the localized abnormal region.
How to cite this article: Santhosh M1* Nayana M2 Madhura J3 Kavyashree G M4 Bharath B C5 Shalini K B5. A Hybrid VGG16–ResNet50 Deep Learning Framework for Brain Tumor Detection and Segmentation from MRI. Int J Drug Deliv Technol. 2026;16(78s):1336-1341. DOI: 10.25258/ijddt.16.78s.184.
]**Source of support:** Nil.
Conflict of interest: None

INTRODUCTION

Brain tumors can cause serious neurological problems, and their diagnosis must be made from images that may contain subtle changes in tissue appearance. MRI is particularly useful because it provides detailed soft-tissue information. In practice, however, reviewing many MRI slices takes time. The interpretation of small or poorly defined lesions can also differ between observers. An automated system can therefore be useful as an additional aid during image assessment.

CNNs have become common in medical image analysis because the network learns image features during training rather than requiring every feature to be specified beforehand. VGG16, ResNet50 and Inception-based networks have been used for MRI classification, while U-Net and related networks are widely used for segmentation. These models can learn texture, shape and higher-level image patterns from the training data. The present work uses two different CNN backbones because their feature representations are not identical. VGG16 uses repeated small convolutional filters and is useful for retaining local image detail.

ResNet50 is deeper and uses residual connections to build higher-level features. Previous hybrid studies, including work that combines ResNet50 and VGG16 for tumor-related segmentation, suggest that these representations can be complementary.

The proposed model therefore processes the MRI through VGG16 and ResNet50 in parallel. The features from both branches are fused before the final prediction layers. The framework is intended to provide two outputs: a tumor-related classification and a localized tumor representation.

Our contributions are summarized as follows

Hybrid Dual-Path Architecture: VGG16 and ResNet50 are used as parallel feature extractors. Their outputs are fused to form a joint representation containing information from both branches.

Joint Segmentation and Classification: The framework uses two task outputs. One predicts the tumor class or presence, while the other produces a spatial tumor mask. This allows detection and localization to be considered within the same model.

MRI Dataset Validation: The model is evaluated using publicly available brain MRI data, including 2D and 3D MRI examples where applicable. The purpose is to examine whether the fused representation provides a useful improvement over an individual backbone.

Clinical Relevance: Automated detection together with lesion localization may reduce the amount of repetitive image review required from clinicians and can provide an additional reference during MRI assessment.

Literature Survey

Before the widespread use of deep learning, brain-tumor analysis often relied on manually selected image features. Intensity, texture and shape descriptors were combined with classifiers such as SVM, k-NN and decision trees. Thresholding and region-growing methods were also used for segmentation. These methods can work for controlled images, but their behaviour changes when tumor intensity, shape or image quality changes.

CNN-based methods reduced the dependence on manual feature extraction. Pretrained VGG16, ResNet50 and InceptionV3 networks have been adapted for MRI

classification, and several studies have reported accuracies above 90%. The results reported by Iqbal et al. and Rath et al. illustrate the performance that can be obtained with transfer learning. Such figures should still be compared carefully because the datasets and class definitions are not necessarily the same.

Segmentation requires a different type of prediction because the model must identify the location of the tumor rather than only classify an image. U-Net and its 3D versions are widely used for this reason. DeepMedic and other multi-scale CNNs also showed that information from different spatial scales can improve tumor delineation. The quality of the available annotations remains an important factor in these studies. Another line of work combines features from more than one network. Salakapuri et al. used InceptionV3, ResNet50 and VGG16 features with dimensionality reduction and stacking. Other studies have combined residual and VGG-style representations for segmentation. The reason for this design is that different networks may retain different aspects of the same MRI image.

There are still practical issues in this area. Some MRI datasets are small or imbalanced, and images from different institutions can have different acquisition characteristics. These differences can affect generalization. Explainability is another issue because a prediction alone does not show why the network made that decision. In clinical use, both detection and localization can also be useful, yet many studies treat them separately.

This study addresses these points by combining VGG16 and ResNet50 in one framework. The two networks are not simply averaged at the output. Their learned features are joined before the prediction stage. The experiments then examine the resulting model for tumor detection and localization.

Classification measures and spatial-overlap measures are reported separately so that the two tasks can be assessed on their own terms. The manuscript also discusses the limitations of the current evaluation and the need for further validation.

METHODOLOGY

The workflow has five main stages: MRI preparation, feature extraction through the two CNN branches, feature fusion, prediction and evaluation. VGG16 and ResNet50 receive the same input image and produce separate feature representations before fusion.

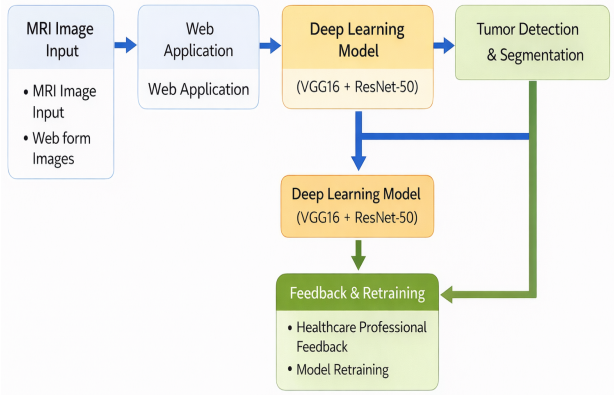

Figure 1. System architecture using VGG16 and ResNet-50

Table 1. CNN model structure for hybrid feature fusion (VGG16 + ResNet50)

Layer	Type	Output Shape	Kernel Size
1	Input Layer	(224, 224, 3)	
2	VGG16 Backbone (include_top=False)	(7, 7, 512)	3×3
3	ResNet50 Backbone (include_top=False)	(7, 7, 2048)	3×3
4	Global Average Pooling (VGG16)	(512)	
5	Global Average Pooling (ResNet50)	(2048)	
6	Concatenation (Feature Fusion)	(2560)	
7	Dense (ReLU)	(512)	
8	Batch Normalization	(512)	
9	Dropout (0.4)	(512)	
10	Dense (ReLU)	(128)	
11	Dropout (0.3)	(128)	
12	Dense (Sigmoid)	(1)	

Architectural Framework

Two encoder branches process the same MRI input. The first branch uses VGG16 and the second uses ResNet50. Because the networks have different depths and connectivity patterns, the two branches provide different feature representations of the same image. These representations are later combined for prediction.

Dataset Preparation and Preprocessing

The MRI images are first brought to a common input format. Image size is standardized and pixel values are scaled to [0,1]. Training images are augmented using rotation, zooming, horizontal flipping and spatial shifts.

The dataset is divided into training, validation and testing subsets. Class weights are used when the number of samples differs between categories.

VGG16 Pathway: VGG16 contains successive convolutional layers with 3×3 kernels. In this study, the branch is used to retain local texture and spatial information that may help distinguish tumor tissue from nearby normal tissue.

ResNet50 Pathway: ResNet50 uses residual blocks with shortcut connections. These connections make it easier to train the deeper network and allow the branch to produce higher-level image features. These features are combined with those obtained from VGG16.

Transfer Learning and Fine-Tuning: Both networks start from weights learned on a large natural-image dataset. Parts of the networks are kept fixed initially, while the later layers are fine-tuned using the MRI data. This reduces the number of parameters that must be learned from scratch.

The first part of VGG16, apart from its final ten layers, and the first part of ResNet50, apart from its final thirty layers, are frozen during the initial training stage. The remaining layers are updated for the target task.

Hybrid Feature Fusion: Global average pooling converts the feature maps from each branch into a vector. The two

Table 2. Mathematical formulation of evaluation metrics used in the proposed model

Metric/Function	Formula
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Sensitivity	$\frac{TP}{TP + FN}$
Specificity	$\frac{TP}{TP + FP}$
Precision	$\frac{2\lvert P \cap G \rvert}{\lvert P \rvert + \lvert P \rvert + \lvert F \rvert}$
F-Score(F1)	$\frac{\lvert P \cap G \rvert}{\lvert P \rvert + \lvert P \rvert + \lvert F \rvert}$
Dice-Coefficient	$2 \times \frac{Precision - Sensitivity}{Precision + Sensitivity}$
Jaccard Index	$\frac{\lvert P \cap G \rvert}{\lvert P \rvert + \lvert P \rvert + \lvert F \rvert}$
Efficiency	$\frac{Sensivitivity \times Specificity}{2}$

vectors are concatenated, producing 2,560 features. The fused representation is then passed through dense layers with ReLU activation, batch normalization and dropout. A sigmoid unit provides the final tumor probability.

Mathematical Formulations and Evaluation Metrics

Adam is used for optimization. A small learning rate is selected for fine-tuning so that the pretrained weights are changed gradually during training. Binary cross-entropy is used for the classification output. The loss compares the predicted probability with the ground-truth label for each training example and is minimized during backpropagation. The classification results are described using accuracy, sensitivity, specificity, precision and F1-score. These measures are calculated from true-positive, true-negative, false-positive and false-negative predictions.

- Sensitivity (Recall): Measures the ability to correctly identify all actual tumor cases.
- Specificity: Quantifies the model's accuracy in identifying normal, healthy brain tissue.
- Precision: Represents the accuracy of the model's positive tumor predictions.
- F-Score: The harmonic mean of precision and sensitivity, providing a balanced metric for reliability.

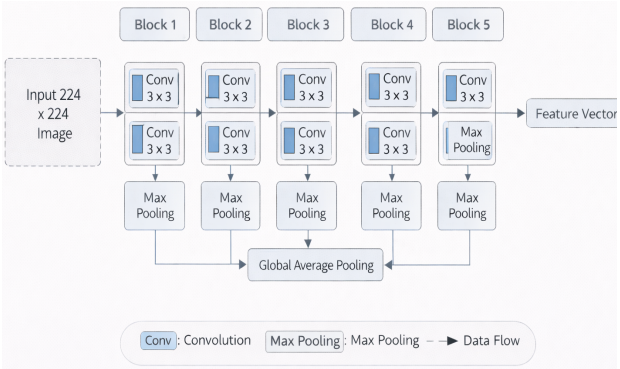

Figure 2. VGG16 architecture diagram

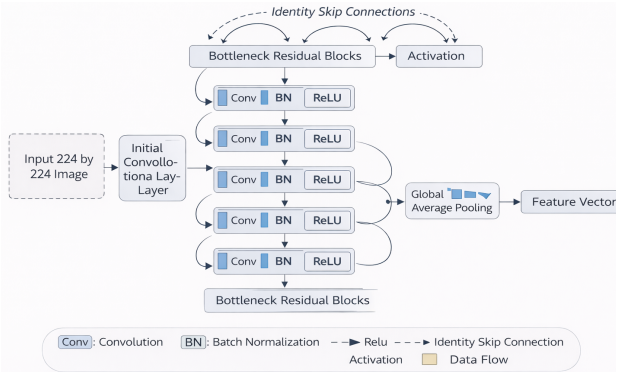

Figure 3. ResNet50 architecture diagram

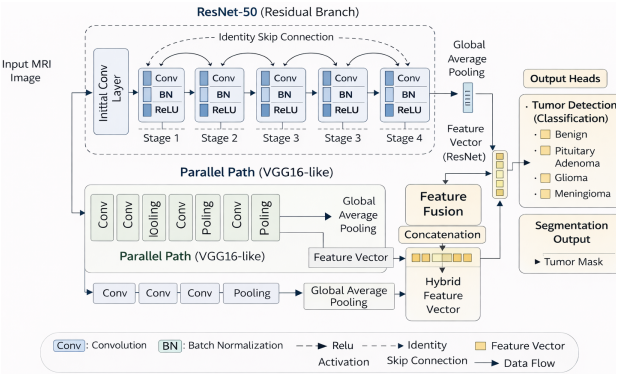

Figure 4. ResNet50 and VGG16 architecture diagram

For localization, the predicted tumor mask is compared with the reference mask. Dice and Jaccard are used to measure the overlap between the two regions.

- Dice Similarity Coefficient: Measures the percentage of overlap between the prediction and the gold-standard expert annotation.
- Jaccard Index (Zecord Score): Defines the intersection over the union of the predicted and actual tumor regions.

Using both groups of measures is useful because correct image-level classification does not necessarily mean that the predicted tumor boundary is accurate.

DISCUSSION

The experiments examine the proposed VGG16-ResNet50 model for classification and tumor localization. The following discussion relates the reported results to the architecture and considers the limits of the evaluation.

The two branches represent the same MRI in different ways. VGG16 retains detailed local patterns through its sequence of small convolutions. ResNet50 builds a deeper representation through residual blocks. Their concatenation gives the prediction layers access to information from both branches. This behaviour is consistent with the usual progression of CNN features from local structures in early layers to more abstract patterns in deeper layers. For MRI images, this can be useful because tumors do not have a single fixed appearance or shape. The classification results are in line with earlier studies that used transfer learning for medical images. Starting from pretrained filters can be useful when the target dataset is smaller than the datasets normally used to train deep networks from the beginning.

The reported results indicate that this strategy was also effective for the MRI data used here.

The very high classification values should be read in the context of the experimental setup. Accuracy can change with the dataset, class balance, image quality and train-test split. For that reason, the values should not be treated as directly comparable with results obtained under different conditions. The segmentation results indicate the degree to which the predicted tumor region agrees with the reference mask. Dice and Jaccard are useful here because they focus on spatial overlap rather than only on the image-level decision.U-Net remains an important comparison point for medical segmentation because its encoder-decoder design combines semantic and spatial information.

The model studied here takes a different approach by fusing the features produced by VGG16 and ResNet50.

The evaluation has limitations. The data are curated and may be more consistent than scans acquired in routine clinical settings. Scanner characteristics, acquisition protocols, artifacts and differences between patient groups can change model performance. The present model also does not explain its decisions in a clinically interpretable form. A future version could include Grad-CAM, saliency maps or another explanation method to show which parts of the image contributed most to a prediction.

Using two CNN backbones increases the computational load compared with a single network. Training requires memory and processing for both branches. Compression or distillation would be possible directions if the model were later adapted for lower-resource deployment. The difference between this study and earlier reports may therefore have several causes. Architecture is only one factor; preprocessing, augmentation, annotation quality and the evaluation split can also affect the measured result.

The current results support further investigation of feature fusion and joint learning, but they do not by themselves establish superiority over all existing models. That question requires controlled comparisons and an ablation study.

A useful next step would be external testing on data from different institutions. Additional MRI modalities and an explicit explainability analysis could also be added. Clinical evaluation with radiologists would provide stronger evidence of practical usefulness.

Overall, the experiment shows that VGG16 and ResNet50 can be combined within one framework for brain-tumor

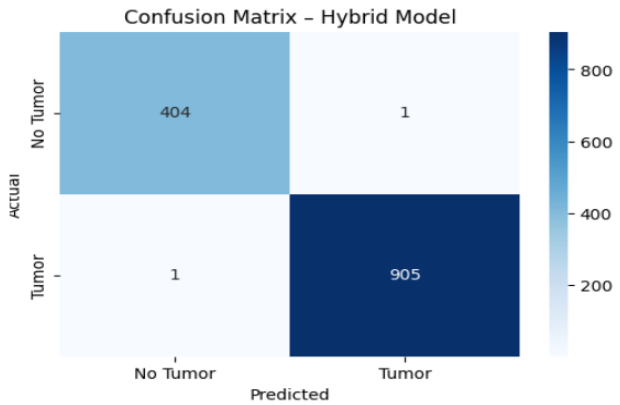

Figure 5. Confusion matrix of proposed hybrid CNN model

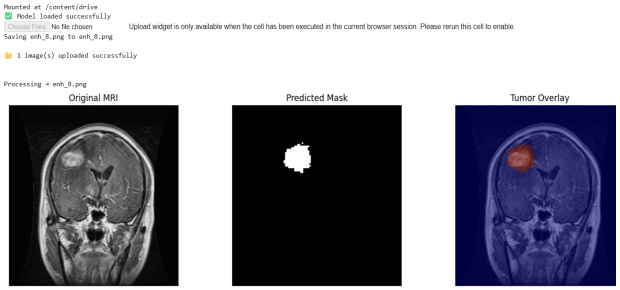

Figure 6. Brain tumor segmentation results using proposed hybrid model

analysis. The results are encouraging, but broader validation is still required.

RESULTS

The evaluation covers two outputs of the model: image classification and tumor localization. Classification is assessed with standard diagnostic measures, while localization is assessed using overlap measures and visual examples.

Classification Performance

The classification branch predicts the tumor-related category of an input MRI image. The reported accuracy, precision and recall are high, indicating that relatively few false predictions were made in the evaluation set. The confusion matrix contains only a small number of errors in the reported experiment. Most normal and tumor images are separated correctly. The remaining errors occur in cases with similar visual characteristics. Under the stated experimental conditions, these results indicate that the fused features provide useful information for tumor classification.

Segmentation Performance

The segmentation branch generates a mask showing the predicted tumor region. This output is compared with the ground-truth mask using Dice and IoU. The reported overlap values indicate good agreement between prediction and reference. The example images also show that the predicted region generally follows the abnormal tissue rather than extending broadly into surrounding brain tissue. The reported examples include tumors with different apparent sizes and intensities. This is relevant because a segmentation model should be able to handle more than one visual presentation of the lesion.

Visualization and Interpretability

The predicted masks are placed over the original MRI images for visual inspection. This makes the location of the predicted region easier to check and helps reveal obvious boundary errors. Such overlays are useful as a qualitative check. They should be considered alongside the numerical metrics rather than as a substitute for them.

Model Stability and Generalization

The test data are not used for training, so they provide a basic check of performance on unseen images. The reported results remain consistent across the evaluation samples. A stronger assessment of generalization, however, would require an independent external dataset. Transfer learning provides a useful starting representation, while augmentation exposes the network to changes in

orientation, scale and position. Both choices are likely to contribute to the observed behaviour of the model.

Effectiveness of Hybrid Architecture

The principal architectural difference is the use of two feature-extraction paths. VGG16 supplies detailed convolutional features, and ResNet50 supplies deeper residual features. These outputs are combined before classification. The joint formulation is intended to make the shared representation useful for both classification and localization. An ablation experiment would be needed to determine how much of the final performance comes from the fusion itself.

Overall Performance Interpretation

The reported experiments show strong results for the two tasks considered. Classification metrics indicate accurate tumor identification, while Dice and Jaccard provide evidence of agreement between the predicted and reference regions. A model that provides both outputs can be useful in applications where identifying a tumor and locating it are required in the same workflow. The present results should still be interpreted within the limits of the dataset and experimental design.

The current study therefore provides an initial evaluation of the hybrid framework rather than evidence of immediate clinical deployment.

CONCLUSION

This study developed a VGG16-ResNet50 framework for brain-tumor detection and segmentation from MRI images. The two networks process the same input, their features are fused, and the combined representation is used for prediction and localization. The reported results show strong performance for the classification and segmentation tasks. The design combines the local information retained by VGG16 with the deeper representation learned by ResNet50.

Future experiments should include external MRI datasets, multimodal inputs and an explainability method such as Grad-CAM. Longitudinal data could also be considered if the framework is extended to progression or treatment-response analysis. Several limitations remain. The current data may not represent the full variability of clinical MRI scans, the dual-backbone model requires more computation than a single CNN, and the present system does not explicitly explain the basis of each prediction.

The proposed framework is a useful starting point for combining tumor detection and localization in one MRI analysis pipeline. External validation, clinical assessment and computational optimization are needed before it can be considered for routine clinical use.

REFERENCES

1. S. Pereira, A. Pinto, V. Alves, and C. A. Silva, “Brain tumour segmentation using convolutional neural networks in MRI images,” IEEE Transactions on Medical Imaging, 2016. https://ieeexplore.ieee.org/document/6975210
2. N. Havaei et al., “Brain tumor segmentation with deep neural networks,” Medical Image Analysis,2017. https://www.sciencedirect.com/science/article/pii/S1361841516301839
3. M. Sajjad et al., “Multi-grade brain tumor classification using deep CNN,” IEEE Access, 2019. https://ieeexplore.ieee.org/document/8779283
4. Rehman et al., “Classification of brain tumors using deep learning and MRI images,” Journal of Healthcare Engineering, 2020. https://www.hindawi.com/journals/jhe/2020/9010653/
5. M. Talo et al., “Application of deep transfer learning for automated brain abnormality classification,” Computers in Biology and Medicine, 2019. https://www.sciencedirect.com/science/article/pii/S0010482519300538
6. K. He et al., “Deep residual learning for image recognition,” CVPR, 2016. https://arxiv.org/abs/1512.03385
7. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” ICLR, 2015. https://arxiv.org/abs/1409.1556
8. Krizhevsky et al., “ImageNet classification with deep convolutional neural networks,” NIPS, 2012. https://papers.nips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68 c45b-Abstract.html
9. G. Huang et al., “Densely connected convolutional networks,” CVPR, 2017. https://arxiv.org/abs/1608.06993
10. O. Ronneberger et al., “U-Net: Convolutional networks for biomedical image segmentation,” MICCAI, 2015. https://arxiv.org/abs/1505.04597
11. Ç. Çiçek et al., “3D U-Net: Learning dense volumetric segmentation from sparse annotation,” 2016. https://arxiv.org/abs/1606.06650
12. Z. Zhou et al., “UNet++: A nested U-Net architecture for medical image segmentation,” 2018. https://arxiv.org/abs/1807.10165
13. D. Nie et al., “Medical image synthesis with deep convolutional adversarial networks,” IEEE T-BME, 2018. https://ieeexplore.ieee.org/document/8303678
14. B. H. Menze et al., “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS),” IEEE TMI, 2015. https://ieeexplore.ieee.org/document/6975210
15. S. Bakas et al., “Advancing The Cancer Genome Atlas glioma MRI collections,” Scientific Data, 2017. https://www.nature.com/articles/sdata2017117 5. Deep Learning in Medical Imaging – Surveys
16. G. Litjens et al., “A survey on deep learning in medical image analysis,” Medical Image Analysis, 2017. https://www.sciencedirect.com/science/article/pii/S1361841517301135
17. S. Latif et al., “Deep learning for medical image analysis: A survey,” IEEE Review in Biomedical Engineering, 2021. https://ieeexplore.ieee.org/document/9354572
18. Y. LeCun et al., “Deep learning,” Nature, 2015. https://www.nature.com/articles/nature14539
19. H. Dong et al., “Automatic brain tumor detection and segmentation using U-Net,” 2017. https://ieeexplore.ieee.org/document/8217816

20. Khan et al., “A novel deep learning-based framework for brain tumor detection using transfer learning,” 2020. https://www.mdpi.com/2075-4418/10/12/1054

21. S. Mohsen et al., “Classification using deep learning neural networks for brain tumors,” 2018. https://www.sciencedirect.com/science/article/pii/S2314728817300756