International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(8s) 2026 pp. 1130-1142
https://doi.org/10.70917/ijcisim-2026-4327
Article

A Multimodal Deep Learning Framework for Brain Tumor Classification Using EfficientNet and Vision Transformers

Lovenish Sharma[1],Dr Saroj Kumar Nanda[2]

[1]Ajeenkya DY Patil University, Pune, India, ORCID ID: 0000-0003-1705-6708, Email: lovenish2008@gmail.com
[2]Ajeenkya DY Patil University, Pune, India, ORCID ID: 0000-0002-0227-5492 , Email : sarojkumarnanda1979@gmail.com

Abstract: Brain tumors are becoming life-threatening issues when not treated at earlier stages. Accurate brain tumor detection is essential in order to support clinical decisions and diagnosis. The study uses the multimodal deep learning framework for brain tumor classification by effectively combining the EfficientNet-B7 network and vision transformer architectures. Modality specific feature extraction has been carried out using the EfficientNet-B7 network for both the Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). Then feature maps are converted into token sequence and passed through the transformer encoder layers in order to effectively model the global dependencies across the image. Then these features are fused and then classified using the fully connected layers whether it is healthy or tumor. Based on the experimental evaluation, the proposed model in this study such as MM-EffiFormer demonstrated the significant classification performance by achieving Accuracy of 0.990, Sensitivity of 0.987, Specificity of 0.995, Dice Similarity Coefficient (DSC) of 0.991, outperforms the existing model FCM-SVM.
Keywords: Brain Tumor, Multimodal, Deep Learning Framework, EfficientNet-B7, Vision Transformer, CT, MRI.

1. INTRODUCTION

The brain serves as the principal organ of the central nervous system, it is located in the head and surrounded by the skull. Any kind of abnormal or uncontrollable growth of the brain cells leads to the development of the tumor in the brain. If these are not treated or diagnosed at the earlier stages, it leads to significant neurological disorders in the human. In clinical applications, Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) techniques have been extensively used in order to identify brain tumors. CT provides the detailed analysis of bone structures, tumors that exist in the brain. While MRI further provides the more detailed anatomical information related to the brain such as soft tissues, tumors. Even though these both modalities have provided detailed anatomical information related to the brain by using the manual inspection techniques, which means it is a time-consuming process.

Recent advancements in the artificial intelligence techniques have shown the accurate performance in the medical imaging applications and it includes the tasks such as segmentation, identification and classification of various diseases and tumors in the human body (Cè et al., 2023). Even though using these techniques cannot completely replace the radiologist or medical experts, these techniques serve as a helpful diagnostic tool in medical applications especially brain tumors in an accurate and reliable way. Convolution Neural Network (CNN) has the ability to accurately analyze medical images such as CT, MRI for the application of prediction, segmentation and classification in medical applications. Conversely, transfer-learning also provides the advantages of using the pre-trained CNN models especially where data and computational resources are limited (Salehi et al., 2023). However, CNN models often face certain challenges in brain tumor detection and classification but they require large datasets and only capture the local relationships in the medical image (Maqsood et al., 2022). In addition to that CNN also fails to completely fail to model the global dependencies across the medical image data. Furthermore, many of the existing models have used single modality data. Although a single imaging modality is able to provide the detailed anatomical representation of the medical images, each of the imaging modalities has distinct advantages. CT is better at capturing the bone structure, swelling and tumors whereas MRI is better at capturing the soft tissues, tumor regions, ductal systems. Consequently, the model performance varies on the basis of imaging modalities which have been used.

To effectively address these challenges, this study has proposed a deep learning framework for brain tumor classification by effectively combining the EfficientNet-B7 with the vision transformers. The main objective of this study is to clearly capture the anatomical information from both the CT and MRI images and then enhance the feature representation by using the transformer architectures. Subsequently, all these steps enable the proposed MM-EffiFormer to accurately detect and classify the tumors.

1.1 Research Gap

Much of the existing research has been focused on employing the deep learning-based models for the application of brain tumor detection and classification by considering the Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). CTs are better at capturing the brain structures and failed to distinguish the tumor from the healthy brain tissue. On the other hand, MRI is better at highly relevant information related to the brain tumor and allows the health experts to clearly see the essential information related to the brain tumor. Much of the researchers either considered CT or MRI in order to detect the brain tumor, thereby restricting the availability of the complementary diagnostic information. In addition to that, neural network architecture is having the greater potential in capturing the local spatial pattern but often fails to model long-range dependencies in the medical image. Even though transformer-based models often address these limitations, standalone applications fail to exploit the modality-specific representation. This research addresses these limitations, utilizing the multimodal medical information to identify the brain tumor.

1.2 Contributions of this study

The contributions of this study are as follows:

1. This research has developed a novel multimodal deep learning framework for the application of brain tumor classification by using both the imaging modalities such as the Computed Tomography (CT) and Magnetic Resonance Imaging (MRI).
2. In this study, EfficientNet-B7 feature extraction has effectively combined with the transformer encoder layers in order to enable the effective modelling of the both the local texture details along with the global spatial dependencies, to provide the more discriminative and the feature extraction for the application of the medical image analysis.
3. In this study, two individual EfficientNet-B7 networks have been utilized in order to effectively extract features from both the modalities such as CT and MRI by removing the final classification layer in this architecture.
4. Instead of individually analyzing the features, the proposed framework has fused both these features to effectively allow the model to capture the cross-modal relationships between the CT and MRI.
5. To assess the effectiveness of the proposed model, it has been compared with the other baseline models such as FCM-SVM. Based on the experimental evaluation on the same dataset, the proposed model has outperformed the FCM-SVM model across the performance metrics such as Sensitivity, Specificity, Accuracy and Dice Similarity Coefficient (DSC).

2. LITERATURE REVIEW

The rapid involvement of the artificial intelligence techniques in the medical applications has demonstrated the significant performance. Earlier classifier models such as Support Vector Machines (SVM) (Vankdothu & Hameed, 2022), Random Forest (RF) (Sarkar et al., 2023; Amin et al., 2024), ensemble models (Noreen et al., 2021) applied for the tumor classification have shown greater accuracy. When SVM is used for the tumor classification, they have shown significant classification performance with the accuracy of 96.19% (Latif et al., 2022). However, these model's performance majorly depends on the feature extraction.

CNN based models such as VGGNet (Jabbar et al., 2023), ResNet (Oladimeji & Ibitoye, 2023; Neamah et al., 2024) have demonstrated superior performance in the brain tumor classification and segmentation tasks. Because CNN has the capability of effectively capturing the local spatial regions in the image. Even though modes show promising models for medical classification performance but they use a single modality dataset and this limits the diagnosis accuracy.

Vision Transformers (ViTs) have emerged as one of the promising techniques in medical image analysis (Takahashi et al., 2024; Khan et al., 2025; Rane, 2023). Because they effectively capture global dependencies by using the self-attention mechanism. Many existing studies have effectively combined these transformer architectures with the CNN in order to further enhance the medical image analysis task and also achieved better performance (Sadu et al., 2025).

Recent studies indicate that they (Hassan & Ghadiri, 2025) indicated that even though the developed deep learning-based approaches demonstrated significant brain tumor classification performance. But their performance is limited in the real-world healthcare environments due to the issues like the training of the models has carried on the less number of samples and high computational resource requirement. This paper has applied the transfer learning capabilities such using the pre-trained models that have been trained on a wide range of large datasets. By incorporating these transfer learning capabilities, it significantly reduces the model training time, issues exited because of the limited medical data. Based on the experimental evaluation on the collected dataset such as the Masoud (i.e., combination of three datasets such as Figshare, SARTAJ, Br35H), the EfficientNetV2 shown the superior brain tumor classification performance and also achieved the accuracy of 0.9916.

Furthermore, table 1 shows the different types of models applied for the application of brain tumor classification and their performance has been compared across the metrics such as accuracy, precision, recall.

Reference Paper	Model	Accuracy	Precision	Recall
Akter et al., (2024)	Deep Convolutional Neural Network (Deep CNN)	0.987	0.988	0.987
dos Santos Felipe et al., (2023)	AI Technique	0.9893	0.98	0.99
Arumugam et al., (2024)	CSA-MLP	0.9856	0.9864	0.98
Ullah et al., (2022)	InceptionResNetV2	0.9891	0.9828	0.9975

Table. 1. Overview of different methodologies applied for the brain tumor classification

3. RESEARCH METHODOLOGY

3.1 Data Collection and Extraction

The dataset which has been used in this research for the application brain tumor detection and classification has been collected from the Kaggle repository and it has been downloaded by link: https://www.kaggle.com/datasets/murtozalikhon/brain-tumor-multimodal-image-ct-and-mri. This dataset mainly consists of data related to the multimodal medical images such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) scans of the brain. Further each of the images in this collected dataset has been labeled with the tumor and identified location in both the CT and MRI images. After the data has been collected, then the dataset organizes separate sections on the basis of class such as Healthy and Tumor for both the image modalities such as CT and MRI.

3.2 Preprocessing

Once the data has been collected, this data has undergone the data pre-processing stage. Because the collected may consist of the unnecessary information in the data such as noise, higher contra levels. When this data has been directly used for the training of the developed models, they may face certain challenges such as more training time, using more resources and being able to provide inaccurate results. To overcome these issues, data preprocessing is necessary to ensure consistent and stable model training.

Initially, resizing technique has been applied for each image in the two modalities such CT and MRI. This means all the images has been resized to a size of 224 224 pixels in order to effectively meet the input size of the deep learning models used in this research such as the EfficientNet-B7 network. ×

Once all the images are resized to the fixed resolution and then normalization technique has been applied to all these images. Then these resized grayscale images have been effectively converted to the three-channel format. After that Normalization technique's main objective is to effectively convert the image pixel values in the range of "-1" to "1", enabling the maintaining the feature consistency, faster and stable training of the CNN models. Together all this data pre-processing ensures that the data is clean and quality data. Figure 1 clearly shows how the raw data has converted to the clean and high-quality data along with their labels and image modalities.

Figure. 1. Data Preprocessing

Then this high-quality dataset which includes both CT and MRI images have been effectively split into two sections such as 80% for the deep learning model training purpose, where the remaining 20% has been used testing the performance of the models on the unseen or new data. Once the data undergoes the pre-processing stage, random samples in the preprocessed dataset are visualized and it is shown in the figure 2.

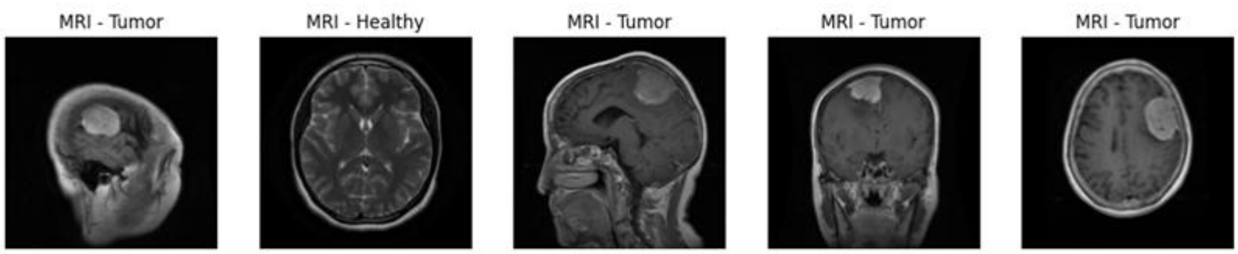

Figure. 2. Random Samples of preprocessed images with their labels and modalities

3.3 Feature Extraction

In this research, features have been extracted from the high-quality preprocess images of both CT and MRI by using the EfficientNet-B7 network. The proposed framework in this study mainly employs the two independent EfficientNet-B7 backbones. In both of these EfficientNet-B7 networks, the final classification layer has been removed. Because in this study, EfficientNet-B7 is mainly used to extract the most informative features from both the CT and MRI images. One EfficientNet-B7 network has been used for the CT images, whereas the second EfficientNet-B7 network has been applied to the MRI images. Subsequently, this type of the feature extraction from the both the modalities such as CT and MRI images, enables the modality specific representations that effectively capture the relevant and unique characteristics from each of the imaging techniques. The extracted features from the both the modalities have been represented as:

$$F\_CT = E\_CT\ (CT\ Image) \quad (1)$$

Where,

E_CT = EfficientNet-B7 feature extraction function for CT modalities

$$F\_MRI = E\_MRI\ (MRI\ Image) \quad (2)$$

Where,

E_MRI = EfficientNet-B7 feature extraction function for MRI modalities

3.4 Transformer-Based Feature Modeling

Although the EfficientNet-B7 features extracted layers can extract the features from both the CT and MRI modalities, these are only useful for capturing the local spatial relationships and failed to model the global dependencies across the image. To effectively overcome this issue, this study effectively integrated the transformer encoders. Because transformer-based architectures have demonstrated significant performance in learning the long-range spatial relationships, contextual feature interactions in the image.

In this study, the feature maps are effectively converted into sequence of tokens and then passed through the transformer encoder layers. Self-attention mechanism which is employed in the transformer architectures is able to effectively model the global dependencies across both the CT and MRI modalities. Because of this self-attention mechanism, it only models the most important and informative regions in the image instead of relying only on the local regions in the image. The self-attention mechanism has been represented by:

$$Attention(Q, K, V)=softmax((QK^T)/\sqrt{(d\_k)})V \qquad (3)$$

Where,

Q = Query matrice

K = Key matrice

V = Value matrice

d_k = Dimensionality of the key vector

Then the outputs obtained from the transformer encoder layers have been aggregated by using the mean pooling. This aggregation step using the mean pooling, ensures the compact and discriminative modality specific feature vectors.

3.5 Brain Tumor Classification by Fully Connected Layers

After transformer encoder layers effectively perform the feature modeling. Then transformer encoder outputs of both the CT and MRI modalities have been concatenated or fused. These fused representations effectively capture the complementary information of both the CT and MRI modalities. It has been represented as:

$$F\_Fusion=[F\_CT; F\_MARI] \qquad (4)$$

These fused vectors effectively capture the relevant information in both the modalities. From CT images, it captures the structural information whereas from the MRI image, it captures the information related to the soft tissues. Together the combination of both these results in detailed anatomical information related to the brain especially brain tumors.

Once the features have been effectively fused, it has passed through the fully connected layers. The main objective of the FC layers is to effectively predict the probability of each of the classes such as “Healthy” or “Tumor” using a SoftMax activation function.

3.6 Pseudocode for Proposed Multimodal Deep Learning Framework

Input:

1. Collect dataset, pair the CT and MRI brain images
2. Class Labels (0 = Healthy, 1 =Tumor)

Output:

1. Predict the class label (“Healthy” or “Tumor”)

BEGIN

1. Data Collection and Initialization
   For each of the image in dataset:

Load CT Image, X_{CT}
Load MRI Image, X_{MRI}
Assign the label, Y = (0, 1)
Store the model (X_{CT},X_{MRI}, Y)
End for

2. Preprocessing
Load the stored dataset (X_{CT},X_{MRI}, Y)
    - Resize the each image to 224 × 224 pixels
    - Convert the grayscale images to three-dimensional format
    - Normalize the each pixels of image in then range [-1, 1]
    - Split the dataset
        a. 80% for training
        b. 20% for testing

3. Modality Specific Feature Extraction
    - Initialize the EfficientNet-B7 model for CT image
        a. Extract the features from CT image
    - Initialize the EfficientNet-B7 model for MRI image
        b. Extract the features from MRI image

4. Transformer-Based Feature Modeling
Initialize the transform encoder layers for both CT and MRI
    - Convert feature maps to token sequences (both the CT and MRI individually)
    - Apply mean pooling

5. Multimodal Feature Fusion
    - Fuse transformer encoder outputs of CT and MRI

6. Classification
Initialize the fully connected layers
    - Predict the class (Healthy, Tumor) by using the softmax activation function
7. Model Performance Evaluation
    - Sensitivity (Recall)
    - Specificity
    - Accuracy
    - Dice Similarity Coefficient (DSC)

END

Figure 3 shows the architectural workflow of how the CT, MRI images are processed and used for prediction using the EfficientNet-B7 networks and vision transformer architectures.

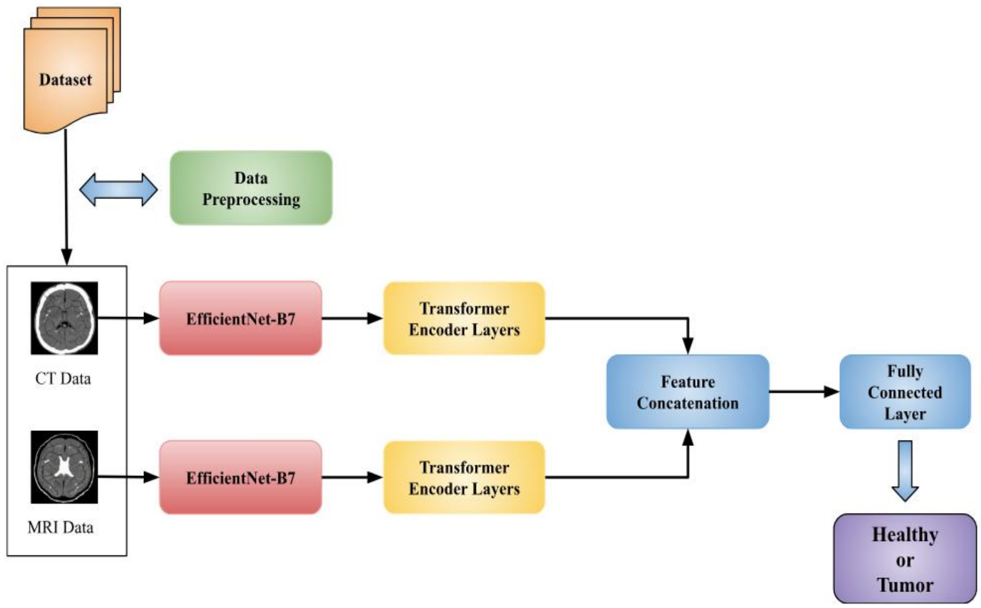

Figure. 3. Workflow of multimodal deep learning framework for brain tumor classification

4. PROPOSED RESULTS AND ANALYSIS

4.1 Experimental Setup

The experimentation of this study has been carried out using the hardware requirements such as a computer system having the windows/MAC operating system, GPU, intel i5/17 processor, 8 GB/ 16 GB RAM and internal storage of 512 GB/1TB or these experiments also conducted by using the cloud-based environments such as Google Colab. Table 2 shows the software requirements used for developing and assessing the model performance.

Requirements	Description
Programming Language	Python
Python Libraries	os, Numpy, Pandas, Matplotlib, Tensorflow, cv2, random, pathlib, collections

Table. 2. Software Requirements

Furthermore, the details of the dataset which is employed in this research is shown in table 3.

Parameter	Description
Dataset Collected from	Kaggle
Dataset Download Link	https://www.kaggle.com/datasets/murtozalikhon/brain-tumor-multimodal-image-ct-and-mri
Dataset Contains	CT and MRI scans of the brain tumor
Total Images	9620

Table. 3. Dataset Description

4.2 Performance Metrics

The performance of the models has been evaluated by using the performance metrics such as Sensitivity, Specificity, Accuracy, Dice Similarly index

1. **Accuracy:**

This metric is used to analyze model performance such correctness among all the predictions.

$$Accuracy=(TP + TN)/(TP + TN + FP + FN) \quad (5)$$

2. Sensitivity

This metric is useful in assessing the model performance in correctly identifying the tumor cases. Higher the value in Sensitivity indicates that the lower false negative rate.

$$Sensitivity=TP/(TP + FN) \quad (6)$$

3. Specificity

This metric is useful in assessing the model performance in correctly identifying the tumor cases. Higher the value in specificity significantly indicated that the healthy cases are not identified as the tumor cases.

$$Specificity=TN/(TN + FP) \quad (7)$$

4. Dice Similarity Coefficient (DSC)

This metric measures the overlap between the predicted and reference labels.

$$Dice\ Similarity\ Index\ (DSC)=2TP/(2TP + FP + FN) \quad (8)$$

Where,

TP = True Positives

TN = True Negatives

FP = False Positives

FN = False Negatives

4.3 Performance Analysis

The proposed multimodal deep learning framework for brain tumor classification has demonstrated significant performance. This framework effectively combines the EfficientNet-B7 feature extraction ability along the vision transformer-based feature modeling. To assess the performance of this proposed MM-EffiFormer model, it has been tested on a properly processed and clean dataset. Based on the experimental evaluation, the proposed MM-EffiFormer has achieved the higher values in the key performance metrics such as Sensitivity of 0.987, Specificity of 0.995, Accuracy of 0.990 and Dice Similarity Coefficient (DSC) of 0.991. The figure 4 shows the performance metrics values obtained for the proposed MM-EffiFormer on the multimodal dataset.

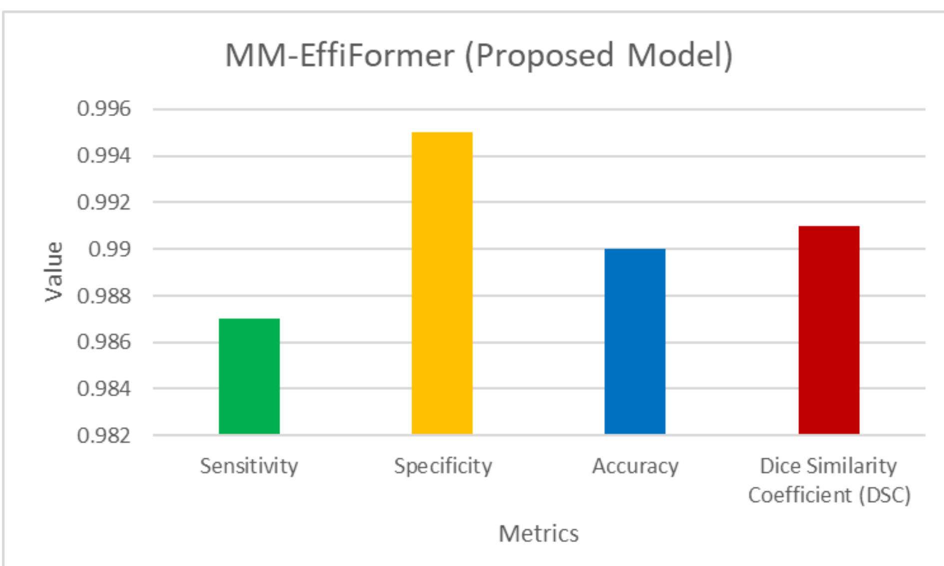

Figure. 4. Performance of MM-EffiFormer Model across the metrics

4.4 Comparative Analysis

To assess the effectiveness of the proposed MM-EffiFormer, it has been compared against the baseline existing model. The comparative analysis of both these models is shown in table 4.

Metric	**FCM-SVM**	**MM-EffiFormer (Proposed Model)**
Sensitivity	0.977	0.987
Specificity	0.979	0.995
Accuracy	0.982	0.990
Dice Similarity Coefficient (DSC)	0.961	0.991

Table. 4. Comparative Analysis of the Models Across the Metrics

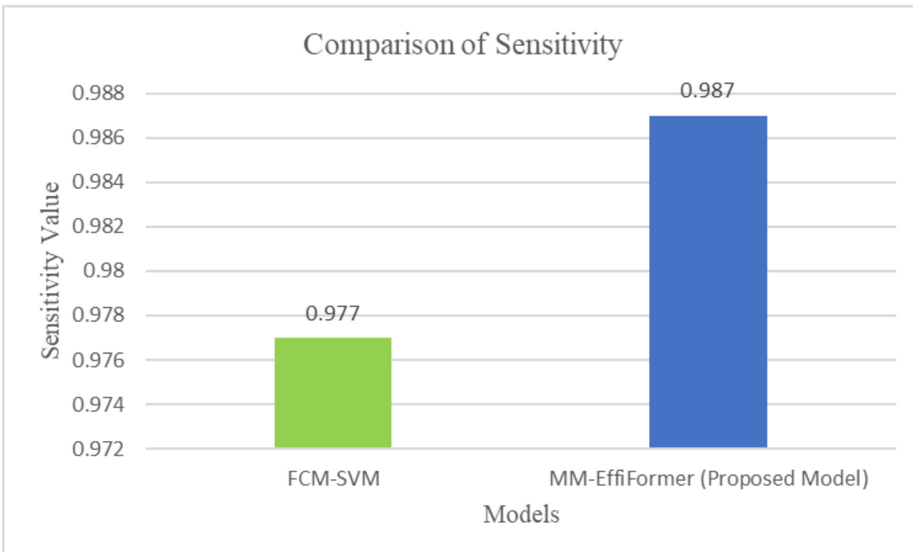

Figure. 5. Comparison of Sensitivity Metric across the models

Figure 5 shows that the proposed MM-EffiFormer has achieved the highest sensitivity value of 0.987 compared to the FCM-SVM model sensitivity value of 0.977. This significant increase in the sensitivity has indicated that the proposed MM-EffiFormer has demonstrated the significant detection capability with the minimal false-negative rate.

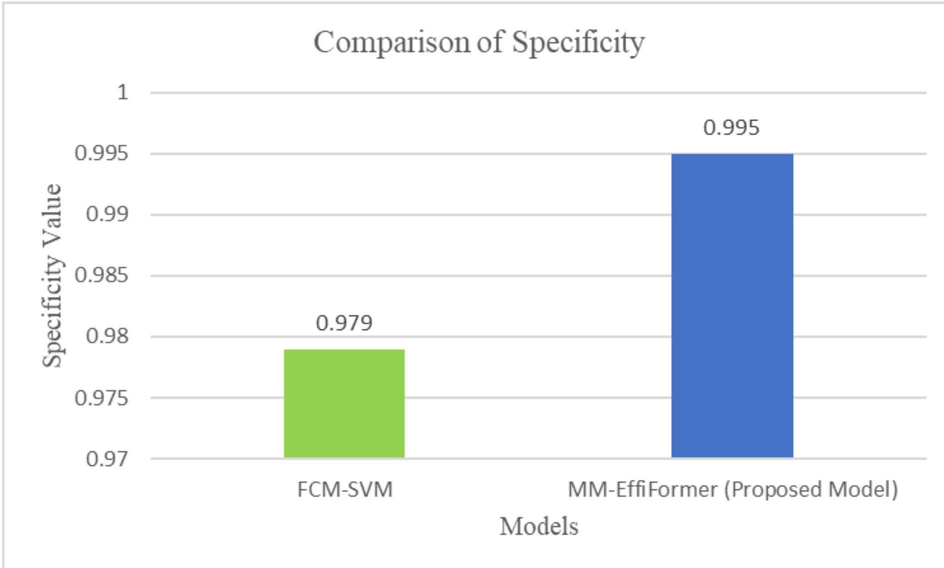

Figure. 6. Comparison of Specificity Metric across the models

Figure 6 shows that the proposed MM-EffiFormer has achieved the highest specificity value of 0.995 compared to the FCM-SVM model specificity value of 0.979. The rise in the specificity value indicates that the proposed MM-EffiFormer significantly reduced the false-positives rate and demonstrated better performance in correctly identifying the non-tumor or healthy cases compared to the base model FCM-SVM.

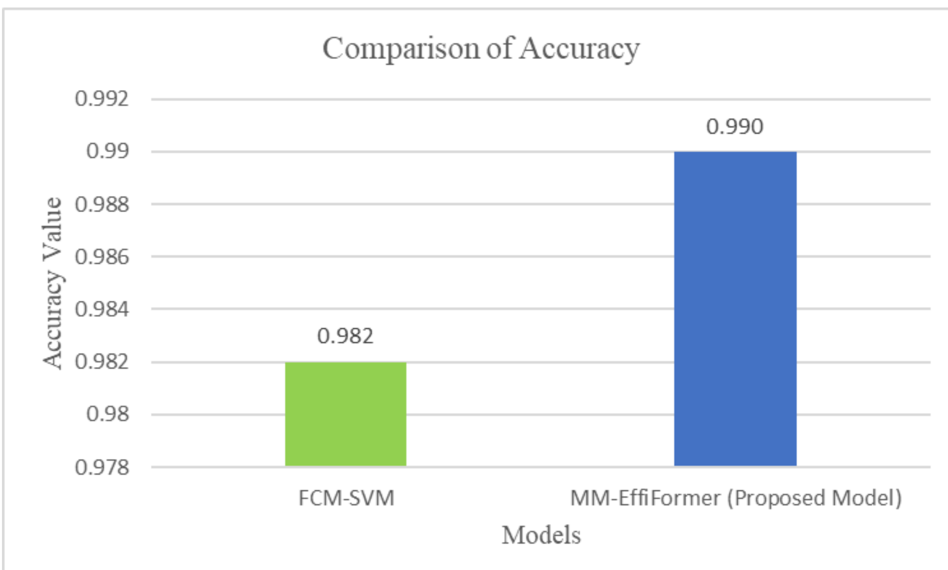

Figure. 7. Comparison of Accuracy Metric across the models

Figure 7 shows that the proposed MM-EffiFormer has achieved the highest accuracy value of 0.990 compared to the FCM-SVM model accuracy value of 0.982. These findings suggest that the proposed model MM-EffiFormer demonstrated superior performance in accurate prediction of tumor cases.

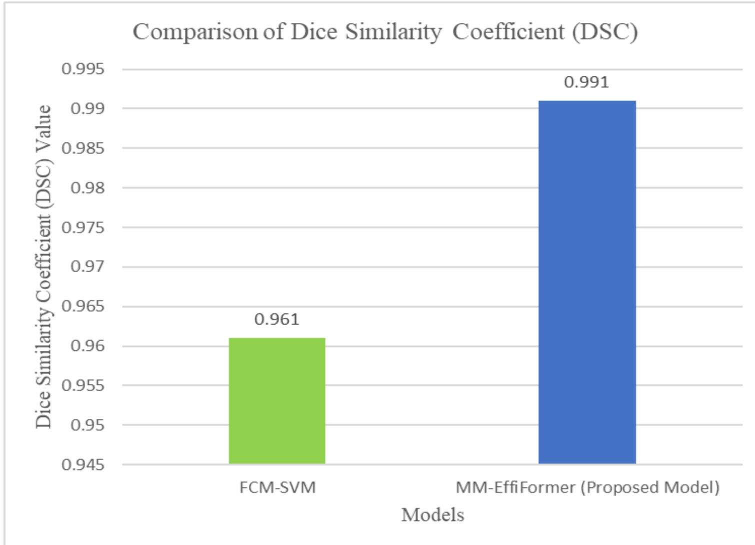

Figure. 8. Comparison of Dice Similarity Coefficient (DSC) metric across the models

Figure 8 shows that the proposed MM-EffiFormer has achieved the highest Dice Similarity Coefficient (DSC) value of 0.991 compared to the FCM-SVM model DSC value of 0.961. These findings suggest that the proposed model MM-EffiFormer demonstrated superior performance in higher agreement between the predicted outputs and ground truth

4.5 Theoretical Implications

This research contributed towards the classification of the brain tumor by extracting the essential information from both the medical image modalities and it includes the CT, MRI. This study integrates the neural network based local feature extraction capabilities with the transformer global dependency modeling each other effectively. Furthermore, this study fuses the features in order to capture the heterogeneous medical information and these fused features used for the brain tumor classification.

4.6 Practical Implications

The proposed framework in this study such as MM-EffiFormer demonstrated significant performance in brain tumor classification and also achieved better results in the key performance metrics like Accuracy of 0.990, Specificity of 0.995, Sensitivity (Recall) of 0.987, Dice Score (DSC) of 0.991 as illustrated in the table 2. These findings indicated that the proposed framework could help the medical assists in early diagnosis of the brain tumor and also significantly reduces manual interpretation time.

5. CONCLUSION AND FUTURE SCOPE

5.1 Conclusion

This research concludes that it proposed a multimodal deep learning framework for brain tumor classification by effectively integrating the EfficientNet-B7 networks with the vision transformer architectures. The novelty of work lies in the use of the multimodal dataset instead of using the single modalities data. Initially, this study collected the dataset from the Kaggle repository and it consists of data related to the brain tumor in both the CT and MRI modalities. Then data pre-processing steps are applied on this dataset and it includes the resizing of the image to 224224 pixels in order to match the input size of deep learning models, converting the grayscale images to three-dimensional format, normalization techniques to make the pixel value lie in the range of -1 to 1. Together all these data pre-processing ensures the clean and high-quality dataset. By using the two EfficientNet-B7 networks, one for the CT images and another for the MRI images, features have been extracted from both the CT and MRI. Then these features are converted into token sequence and passed through the transformer encoding layers in order to model the global dependencies across the image. Then transformer encoder outputs are fused together in order to learn the complex and essential information in the images and this fusion strategy has significantly enhanced the classification performance. ×

Based on the experimental evaluation, the proposed MM-EffiFormer has demonstrated the significant performance in accurately classifying the tumor and healthy cases and also achieved the higher performance values such as specificity, sensitivity, accuracy and Dice Similarity Coefficient (DSC), outperforms the existing model such FCM-SVM.

5.2 Limitations

Even though the proposed framework demonstrated significant brain tumor classification performance on the Kaggle dataset, certain limitations have limited its applicability in real-world clinical applications. First, its performance is evaluated on the single dataset which limits its generalizability across the diverse clinical datasets. Second, the proposed framework currently performs the binary classification of whether tumors exist and often fails to provide the sub-types of the tumor like glioma, meningioma and pituitary tumors. Third, lack of trust and transparency of the model decisions. Which means MM-EffiFormer (proposed framework) identifies whether the brain tumor exists or not and fails to provide the reasons behind why the certain decision has been made. Subsequently, this kind of the limited explainability nature of the proposed framework can reduce the trust and limits its applicability in the real-world healthcare environments.

5.3 Future Scope

To further enhance the performance of the proposed framework, future aspects should focus on the below aspects:

1. Even though the proposed model demonstrated significant performance brain tumor classification by using the multimodal data. Future works should focus on incorporating the other modalities such as the Positron Emission Tomography (PET) and functional MRI (fMRI) to further improve the diagnostic accuracy and classification.
2. Although the proposed framework accurately classifies the tumor classes, still decisions taken by the deep learning models are "black-box" in nature, which limits the trust and transparency of the model decisions. Future work should focus on inserting the Explainable AI techniques such as Grad-CAM, to visually show which sections are considered for taking the particular decisions, which enhances the user trust on the model decisions.

References

1. Cè, M., Irmici, G., Foschini, C., Danesini, G. M., Falsitta, L. V., Serio, M. L., ... & Cellina, M. (2023). Artificial intelligence in brain tumor imaging: a step toward personalized medicine. *Current Oncology*, *30*(3), 2673-2701.
2. Salehi, A. W., Khan, S., Gupta, G., Alabduallah, B. I., Almjally, A., Alsolai, H., ... & Mellit, A. (2023). A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope. *Sustainability*, *15*(7), 5930.
3. Maqsood, S., Damaševičius, R., & Maskeliūnas, R. (2022). Multi-modal brain tumor detection using deep neural network and multiclass SVM. *Medicina*, *58*(8), 1090.
4. Latif, G., Ben Brahim, G., Iskandar, D. A., Bashar, A., & Alghazo, J. (2022). Glioma Tumors' classification using deep-neural-network-based features with SVM classifier. *Diagnostics*, *12*(4), 1018.
5. Noreen, N., Palaniappan, S., Qayyum, A., Ahmad, I., & Alassafi, M. O. (2021). Brain Tumor Classification Based on Fine-Tuned Models and the Ensemble Method. *Computers, Materials & Continua*, *67*(3).
6. Vankdothu, R., & Hameed, M. A. (2022). Brain tumor segmentation of MR images using SVM and fuzzy classifier in machine learning. *Measurement: Sensors*, *24*, 100440.
7. Sarkar, A., Maniruzzaman, M., Alahe, M. A., & Ahmad, M. (2023). An effective and novel approach for brain tumor classification using AlexNet CNN feature extractor and multiple eminent machine learning classifiers in MRIs. *Journal of Sensors*, *2023*(1), 1224619.
8. Amin, J., Sharif, M., Raza, M., & Yasmin, M. (2024). Detection of brain tumor based on features fusion and machine learning. *Journal of Ambient Intelligence and Humanized Computing*, *15*(1), 983-999.
9. Jabbar, A., Naseem, S., Mahmood, T., Saba, T., Alamri, F. S., & Rehman, A. (2023). Brain tumor detection and multi-grade segmentation through hybrid caps-VGGNet model. *IEEE Access*, *11*, 72518-72536.
10. Oladimeji, O. O., & Ibitoye, A. O. J. (2023). Brain tumor classification using ResNet50-convolutional block attention module. *Applied Computing and Informatics*.
11. Neamah, K., Mohamed, F., Waheed, S. R., Kurdi, W. H. M., Taha, A. Y., & Kadhim, K. A. (2024). Utilizing deep improved resnet50 for brain tumor classification based MRI. *IEEE Open Journal of the Computer Society*, *5*, 446-456.
12. Takahashi, S., Sakaguchi, Y., Kouno, N., Takasawa, K., Ishizu, K., Akagi, Y., ... & Hamamoto, R. (2024). Comparison of vision transformers and convolutional neural networks in medical image analysis: A systematic review. *Journal of Medical Systems*, *48*(1), 84.
13. Khan, A., Rauf, Z., Khan, A. R., Rathore, S., Khan, S. H., Shah, N., ... & Gwak, J. (2025). A recent survey of vision transformers for medical image segmentation. *IEEE Access*.

14. Rane, N. (2023). Transformers for medical image analysis: Applications, challenges, and future scope. *Challenges, and Future Scope (November 2, 2023)*.
15. Sadu, V. B., Bagam, S., Naved, M., Andluru, S. K. R., Ramineni, K., Alharbi, M. G., ... & Khadhar Moideen, R. (2025). Optimizing the early diagnosis of neurological disorders through the application of machine learning for predictive analytics in medical imaging. *Scientific Reports*, *15*(1), 22488.
16. Hassan, E., & Ghadiri, H. (2025). Advancing brain tumor classification: A robust framework using EfficientNetV2 transfer learning and statistical analysis. *Computers in Biology and Medicine*, *185*, 109542.
17. Akter, A., Nosheen, N., Ahmed, S., Hossain, M., Yousuf, M. A., Almoyad, M. A. A., ... & Moni, M. A. (2024). Robust clinical applicable CNN and U-Net based algorithm for MRI classification and segmentation for brain tumor. *Expert Systems with Applications*, *238*, 122347.
18. dos Santos Felipe, C., Alva, T. A. P., Winck, A. T., & Becker, C. D. L. (2023, September). An approach in brain tumor classification: The development of a new convolutional neural network model. In *Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)* (pp. 28-42). SBC.
19. Arumugam, M., Thiyagarajan, A., Adhi, L., & Alagar, S. (2024). Crossover smell agent optimized multilayer perceptron for precise brain tumor classification on MRI images. *Expert Systems with Applications*, *238*, 121453.
20. Ullah, N., Khan, J. A., Khan, M. S., Khan, W., Hassan, I., Obayya, M., ... & Salama, A. S. (2022). An effective approach to detect and identify brain tumors using transfer learning. *Applied Sciences*, *12*(11), 5645.