RESEARCH PAPER

Explainable Deep Learning Framework For Brain Tumor Detection And Pharmacological Treatment Planning Support Using MRI-Based CNN–Transformer Architecture

Dr.R.Senthilkumar[1]*, Nivaashini M[2], Dr.A.Ravikumar[3], Laxmi Raja[4], Dr.K.Sreenivasa Reddy[5], Rathika Prabhu[6]

[1]Associate Professor, Department of Computer Science Engineering, Hindusthan Institute of Technology, Coimbatore https://orcid.org/0000-0001-8787-4330, sentinfo@gmail.com

[2] Assistant Professor, Department of Artificial Intelligence & Data Science, Sri Eshwar College of Engineering, Coimbatore, http://orchid.org/0000-0002-4469-2737, nive19794@gmail.com

[3] Professor, Department of Computer Science and Engineering, Sridevi Women's Engineering College, Hyderabad, Telangana. https://orcid.org/0009-0004-2327-1396, aravikumar007@gmail.com,

[4] Assistant Professor, Department of Cyber Security, Faculty of Engineering, Karpagam Academy of Higher Education Coimbatore, https://orchid.org/0000-0001-6040-8794, laxmirajaphd@gmail.com

[5] Associate Professor, Department of CSE (DS), TKR College of Engineering and Technology, Medbowli, Meerpet, Balapur Mandal, Rangareddy Distict, Telangana, https://orchid.org/0009-0008-9455-5053, srinivas.reddy7963@gmail.com

[6] Assistant Professor, Department of Artificial Intelligence and Data Science, Dr.Mahalingam College of Engineering and Technology, Coimbatore, https://orcid.org/0009-0002-0554-179X, rathikarathinam@gmail.com

Corresponding Author

Dr.R.Senthilkumar

Associate Professor, Department of Computer Science Engineering, Hindusthan Institute of Technology, Coimbatore https://orcid.org/0000-0001-8787-4330, sentinfo@gmail.com

---

ABSTRACT

Magnetic Resonance Imaging is key to the diagnosis of brain tumors and is fundamental to early diagnosis and planning for therapy. However, much work remains to be done on the precise classification of tumors because the tumor morphology, texture, intensity distribution, and complicated MRI features are still ambiguous. Although some conventional models can also offer comparable effectiveness for medical image analysis, they lack interpretative ability and consequently, clinical transparency, which can greatly limit their use in medical image analysis applications. As mentioned, some traditionally proposed models can be good at medical image analysis, but these models have yet to exhibit interpretability and clinical transparency which can potentially restrict them to answering the practical question in medical image analysis tasks. To tackle these limitations, a Explainable Hybrid CNN–Transformer Framework for intelligent prediction of brain tumors and Clinical Decision Support has been proposed using MRI images. Suggested solutions follow the idea of using Transformer attention mechanism for context, CNN is used for spatial feature extraction. The task of experimental analysis was conducted on four types of lesions (microglial, meningioma, pituitary tumor and non tumor) from MRI images in the Kaggle Brain Tumor MRI dataset. For the proposed model, precision, accuracy, recall, and F1 scores were obtained as 98.42%, 98.11%, 97.94%, and 98.02% respectively making a result of 98.02% that fell within the acceptable range. From the results obtained, the approach suggested the present disclosure offered the ability to enhance the interpretability of the models; the ability to precisely categorize tumors and to guide medical decision making.

Keywords: Brain Tumor Classification, MRI Imaging, Explainable Artificial Intelligence, CNN–Transformer Framework, Deep Learning, Clinical Decision Support, Medical Image Analysis, Grad-CAM, SHAP Analysis, Healthcare AI.

How to cite this article: Senthilkumar R, Nivaashini M, Ravikumar A, Raja L, Sreenivasa Reddy K, Prabhu R. Explainable Deep Learning Framework For Brain Tumor Detection And Pharmacological Treatment Planning Support Using MRI-Based CNN–Transformer Architecture. Int J Drug Deliv Technol. 2026;16(54s): 1620-1631. DOI: 10.25258/ijddt.16.54s.151

INTRODUCTION

Important early tumor diagnosis is key to treatment planning in brain tumors which can be done by Magnetic Resonance Imaging. However, definitive tumour characterisation remains challenging because of the overlap in tumour morphology, texture, tumour intensity distribution and complex MRI characteristics. While current deep learning approaches show promising success in medical image analysis, many existing models lack interpretability and clinical transparency, which is limiting their ability to be readily adopted in medical contexts. To address the above limitations, this study proposes an Explainable Hybrid CNN–Transformer Framework for intelligent brain tumor prediction and clinical decision support, using MRI images. The designed system utilizes Convolutional Neural Networks on spatial feature extraction and also uses Transformer-based attention mechanism to learn context. In addition, explainable AI approaches like Grad-CAM and attention heatmaps and SHAP-based interpretation are leveraged to improve the transparency of predictions and to validate the predictions with radiologists. MRI images from The Kaggle Brain Tumor MRI Dataset were used in the experiment to investigate MRI images of gliomas and meningiomas as well as pituitary tumours and non-tumors. The suggested framework achieved 98.42% accuracy, 98.11% precision, 97.94% recall, and 98.02% F1-score. The results show that the proposed framework not only makes the whole process clinically interpretable but also further enhances the tumor diagnosis accuracy and to intelligent support in the medical decision making sections.Brain tumors are among the most serious neurological diseases affecting the central nervous system and timely and accurate diagnosis enables the development of a successful therapy and prolongs patient lifeThe shape and nature of the neoplasms in the brain may vary widely, including gliomas, meningiomas and pituitary tumors. The time to receive a diagnosis of a type of tumor may be significant in determining the treatment options and can also make a clinical problem harder to treat. MRI is also used widely to diagnose brain cancers because it is excellent soft tissue imaging and is very adept at providing very detailed structural information from brain tissue. The MRI can help the radiologist determine the location of any structural abnormalities seen at the site of the tumour, areas of oedema and other changes. In fact, although the images analyzed by manual MRI have similar visual characteristics and complex anatomical structure, Manual MRI image processing is still a difficult task and is also labor intensive. The automatic processing of medical images has been very promising over the past few years using the techniques of Artificial Intelligence (AI) and deep learning. Convolutional neural networks are often used in medical imaging for their accuracy in learning from MRI images and their capability of spatial and textural extraction from these images [12] [18]. CNN-based networks enable discriminative visual patterns learning without the time consuming feature engineering approach, which would be hard to work out in typical classification applications thus enhancing the efficiency and accuracy of classification. Deep learning has been successfully applied in image classification systems that accomplished the brain tumor identification from the MRI data. Although these developments, most of the existing CNN based models have been primarily developed to study local spatial interdependency and cannot well capture the long-range interactions within the large-scale and complex 3D tumour microstructures.

Many CNN-based systems suffer from poor resilience and generalization due to variations in MRI acquisition quality, tumor size, tumor morphology, tumor intensity distribution and the class imbalance. However, these limitations have recently motivated the development of transformer architectures to be applied for processing medical images. In the field of transformer models, self-attention is used to glean global context from image data and improves the representation of spatial context in MRI scans [10, 19]. In comparison to traditional CNN architectures, Transformer-based models exhibit superior contextual learning abilities, especially for intricate medical imaging tasks that involve diverse tumor shapes. Meanwhile, training such autonomous models is frequently difficult due to large amount of data and high computing requirements. These processing requirements are not suitable for practical use in healthcare facilities having only a few annotated MRI datasets.Hybrid architectures of CNN and Transformers are gaining interest due to the

capacity of contextually learning information from Transformers, and locally extracting features from CNNs.

Most of these challenges are solved by deep learning models, however the complexity of these models makes them difficult to understand, resulting in effective health systems with regards to the tumor classification tasks. Deep learning architectures, and particularly using them as black boxes, are too costly to use, and require skills of validation and understanding beyond the medical practitioner's skill-set [1,3]. The transparency of medical diagnosis and prediction is the number one priority, allowing radiologist and other health care professionals to assess the accuracy of the AI diagnostic outcomes before real-world use. In the medical world, transparency and honesty in medical diagnosis and prediction is critical; it would be helpful if the doctors, such as radiologists, can review the accuracy of AI communication before they use it for the medical practice. Explanatory Artificial Intelligence techniques have been created, for example, in the field of medical imaging. Explanatory Artificial Intelligence techniques can be applied to Medical Imaging. Visual explanations such as Grad-CAM, SHAP and attention visualization [6, 8] can assist physicians in determining which parts of the image they should focus on in order to make predictions for the image.This clarity of interpretation and the increased trust and confidence of clinicians in AI-driven healthcare systems are provided by these interpretability techniques. Explainable models also help physicians gain an understanding of tumour localization from a clinical perspective and aid treatment planning decisions.

In modern health care environment, the development of complex clinical decision support systems is one of the most critical requirements. Helpful in reducing diagnostic strain, ensuring consistency and supporting early detection of disease, AI assisted decision support systems can assist radiologists [20] [21]. Many existing healthcare prediction systems focus on classification performance but have a relatively low level of interpretability and context learning ability. Some algorithms perform well on prediction accuracy but don't have the ability to provide the clinical scenarios to explain them. Furthermore, current medical imaging systems still face challenges such as having computational complexity, MRI variability, and tumor heterogeneity.

Many recent research have attempted to combine explainability with deep learning analysis of MRI images [11, 13, 15]. However, very few research has focused on embedding CNN based spatial learning, Transformer based contextual representation of the image and explanation based visualization into a unified framework of Intelligent brain tumour prediction and clinical decision support. Unlike the typical CNN-based tumor classification systems, the proposed framework includes spatial feature extraction, contextual attention learning, and explainable AI mechanisms that improve diagnostic ability and the clinical interpretability of the model.In this research, we devised an Explainable Hybrid CNN–Transformer Framework to solve the aforementioned problems and challenges in the field of Intelligent Brain Tumor Prediction and Clinical Decision Support using MRI images.

This approach combines CNN feature extraction with CNN-T modelling in the context of MRI images, offering a more robust enhancement for tumor classification accuracy. The suggested approach is based on CNN feature extraction, enhanced by current modelling, to boost the effectiveness of the tumor classification process from MRI images. Furthermore, explainable AI techniques are embedded to provide a graphical explanation of the predictive outcomes and to improve the clarity of the diagnostics. The method is designed to achieve reliable tumor classification while offering clinically actionable and informative decision support for radiologists.

The main results of this research are as follows:

- Developing a Combination of CNN Model and Transformer to Classify Brain Tumors Using MRI images. A new feature is adding explanatory AI supervision to enhance user-friendliness.
- Inclusion of Transformer-based attention mechanisms to benefit from contextual information in MRI features.
- Design of a smart clinical decision support system for radiological support.
- Extensive performance evaluation on MRI brain tumor database along with many diagnostic markers.
- Due the current state of medical imaging, the proposed architecture aims at providing a reliable, understandable and clinically meaningful system for the intelligent diagnosis and treatment of brain tumors as part of healthcare.

RELATED WORKS

In recent years, medical image processing systems are now able to leverage AI in order to improve their performance and efficiency, especially when detecting brain tumors in MRI scans. Medical images often

require information from higher dimensions, such as computed tomography scans or MRI, where deep learning techniques are becoming increasingly popular in the medical field. However, many healthcare-specific models are not transparent, especially considering that accurate and clear predictions are essential to healthcare decision making. Explanatory Artificial Intelligence (XAI) is a relatively new research area aimed to overcome this challenges in medical imaging.This challenge in application of imaging in healthcare relates to the issue of Explanatory Artificial Intelligence (XAI). In this talk Van der Velden et al. [1] addressed the need for AI interpretability to build trust and gain user acceptance of medical image analysis solutions. In addition, in the health care domain, they have been observed to enhance learning results, like in medical knowledge—capturing associations with trans-former health care techniques as demonstrated by Rao et al. [2].

Borys et al. [3] emphasized that explainable AI has an important role for radiologists and visualization-based interpretation methods can help doctors to better understand the predictive behaviour when making a diagnosis. All of these studies agree that the clinical uptake of these AI health care systems must be supported in two ways – by high predictive value and by contextually appropriate interpretation. Many different libraries and researchers have explored the use of XAI in various sectors, such as medical imaging and healthcare analytics. The use of XAI has been extensively studied in medical imaging and health analytics. Mienye and Sun [4] explored many explainable AI techniques and pointed out that, when used in healthcare, they require this openness and dependability. Sadeghi et al. [5] discussed situations where explaining the generated ML models has been an issue in health care predictive systems, highlighting that one of the prominent difficulties in clinical settings is maintaining the credibility of the created ML models. Muhammad et al. [6] performed a systematic review on XAI techniques for medical image analysis, including Grad-CAM, SHAP and attention image visualization. These methods are used to elucidate medical images further. Bharati et al., [7] argue that explainability helps in making the decision making process transparent, while explainability gains consumer trust, especially in health care domain. Ghnemat et al. address the problem of medical images categorization in the context of medical image processing. [8] demonstrate the usefulness of expose analysis interpretation method of medical image classification models that can be seen during their use.I was investigated thoroughly about the usefulness of visible medical image classification models during classify medical image with anomaly visualization using medical image's explainability analysis interpretation method by Ghnemat et al. [8].

With the advent of MRI, which provides detailed information about tumor morphology and tissue abnormalities, deep-learning analysis of brain tumors has been receiving growing attention in recent years. In an aim to analyze and examine various deep learning methodologies to tackle the problem of brain tumor analysis, Dorfner et al. [9] studied the classification results of different brain tumor detection tasks on the basis of MRI images and came to the conclusion that CNNs showed strong classification performance in such tasks. However, the study noted that many of the CNN models have limited interpretability and lacked generalisation power for complex tumour architectures. Hosny and Mohammed [10] studied the notion of explainable vision. The Transformer models significantly outperform traditional CNN-based methods, with acronyms outperforming BacNet when evaluated on the MRI data's spatial span. When tested on the MRI data's spatial span, acronyms outperformed BacNet and Transformer structures were found to improve the contextual feature learning by capturing the long-range spatial correlations within MRI data. Transformer architectures are more likely to require larger datasets and computational power compared to CNN architectures for global feature representation.

Srinivas, P and his colleagues [11] have developed an explainable brain tumour classification system using MRI and demonstrated the improved transparency of tumour prediction using visual interpretation. Iftikhar et al. [12] developed a CNN model that is interpretable for tumor detection and found that the Grad-CAM heat maps could highlight tumor areas that affect the prediction value. By employing explainable AI alongside hybrid deep learning models, Vamsidhar et al. [13] noticed improved accuracy in the classification of tumours through feature fusion techniques. Gundogan et al. [14] proposed a hybrid Dorfner et al. [9] analyzed and examined the classification results of some brain tumor detection tasks using the MRI image to better understand the problem of brain tumor analysis and to understand the various deep learning methods that can solve it. They concluded that CNNs have good classification capability in such tasks.

However, the research in recent years more and more puts an emphasis on hybrid designs, such as CNN combined with Transformer, to make better feature extraction of medical images. In MRI tumor classification, Abraham et al. [16] explored the role of explainable AI, which they determined pairs favor the combination of local spatial data and semantic

context.For MRI tumor classification, Abraham et al. [16] discussed the role of explainable AI, where they observed that hybrid frameworks ranks proved to be better suited for leveraging both local spatial variables and global contextual information to improve prediction reliability. To alleviate computational complexity and enhance the brain MRI classification performance, Mahadevan et al. [17] showed that a learnable and interpretable deep classification approach can successfully be applied to brain tumor classification by MRI, while being more efficient and yet is also more accurate. Transferring the features from existing CNN architectures to the current problem was done by Rahman et al. [18] with the aim of increasing the performance of the tumour identification problem by reusing the features effectively. The transformer-based ones demonstrate a good learning performance in the context while the CNN-based ones exhibit efficient computation for small datasets in medicine. Patel et al. [19] also confirmed using health information, the global information can be captured via attention-based Transformer topology, improving feature representation in healthcare analytics. The clinical decision support systems play principal role in intelligent health systems. In the paper Chen et al. [20] presented a deep learning based clinical decision making support system and emphasized the significance of an intelligent healthcare system to assist medical staff. Zhou et al. [21] proposed a multimodal explainable healthcare intelligence system emphasizing the improved interpretability of AI-driven healthcare predictions via explainability. Gupta and Seeja [22] performed a comparative evaluation of different XAI models, and factored to find interpretability of models as a critical factor for clinical use. Explainable Deep Learning: Methodology applied in classification of various type of brain tumour images and the incorporation and compatibility of the explainability algorithms with the deep feature extraction algorithms hastened the consistency in the diagnosis. Explainable AI models have the potential to increase the explainability and accuracy of the tumor predictions that physicians can make with their MRI diagnostic tools, bringing the ability to make more precise decisions. Talaat [15] found the same for explainable AI models, which enhance transparency and improve tumor prediction accuracy while enabling physicians to use MRI diagnostic scans.

Furthermore, CNN architectures have shown to struggle with long-range contextual knowledge capturing and Transformer-based algorithms require large datasets and high computational resources. Existing models are also sensitive in terms of model robustness and model generalisation to MRI image quality variations, tumor morphology and dataset imbalance. To address these disadvantages the proposed work is robotic system for intelligent brain tumor prediction and clinical decision support system using MRI images based on Explainable Hybrid CNN-Transformer architecture.This solution uses CNN in conjunction with AI that can be explained to improve explainability in diagnoses and the credibility of treatment decisions. This approach is a combination of CNN and Transformer for contextual learning, plus Explainable AI to enhance Trustworthiness of Therapeutic decisions and Transparency in Diagnosis.

PROPOSED SYSTEM

A. Overall Proposed Framework

The proposed Intelligent Brain Tumor Prediction using MRI-Intelligent EHCNN-Transformer is explained. This framework combines CNN's spatial feature extraction capabilities and the T's contextual representation capabilities, thereby improving multiclass tumor classification performance. Additionally, Explainable Artificial Intelligence methodologies are also adopted to improve the traceability and support the clinical understanding when making predictions, in the diagnostic process. This proposed scheme is shown in the figure 1 and is composed four steps viz, MRI image capturing, image preprocessing, CNN feature extraction in spatial domain and Transfor based contextual learning, Weighted Fusion of features, Interpretability Computation, and finally Tumor Classification.

In recent years, the diagnostic evaluation of brain tumors with state-of-the-art artificial intelligence models based on MRI has come to the fore. The challenges of intra- and inter-modality MRI variability, MRI noising and heterogeneity with regards to brain tumor prediction were respectively presented by Satushe et al. [23] in the context of Machine Learning and Deep Learning algorithms assessing the effectiveness of brain tumor analysis. To better estimate the clinical applications of expert AI systems, the researchers behind the study (24) developed an AI brain tumour classification and segmentation system for the MRI and made it easy-to-explain online. To apply the understanding of how systems in the healthcare domain can be explained, Aksoy et al. [24] created an application for an MRI image based web-based Explainable AI system that classifies and segments brain tumor. In another effort, Gomes et al. [25] tapped into the potential of using deep-learning methods to characterize a tumour in an MRI scan and concluded that the hybrid deep-learning methods (CNN–Transformer) outperformed the autonomous models in

the representation of information obtained from an image. Though there have been great strides in MRI studies that can help to predict brain cancers, many scientific questions still need to be resolved. The approaches developed in the past few decades are mainly focused on categorization metrics and provide limited information on radiological validation and clinical confidence.

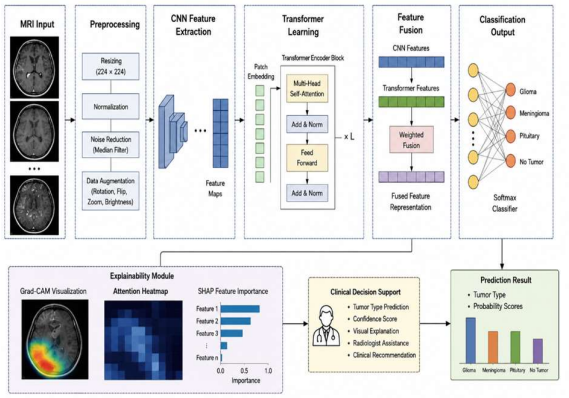

Fig. 1. Comprehensive workflow of the proposed explainable hybrid CNN–Transformer system for MRI-based brain tumor prediction and clinical decision-making support.

B. Data Preprocessing

MRI images may have imaging artifacts, noise, and unequal spatial dimensions which impact classifier performance.Hybrid CNN-Transformer architecture. You combine these technologies – CNN, Transformer and explainable AI – with a more explainable workflow and a more reliable diagnostic workflow. A user-friendly structure of the workflow that's easy to play and easy to understand. Its user-friendly, easy to execute and visually notable feature helps facilitate understanding and high levels of trust and transparency in diagnosis.

To standardize image distributions and to improve alignment of images for optimization, pixel intensity normalization was used. Median filtering was used to suppress imaging noise, and preserve the critical structure information of the tumor region. Brain images for the preliminary experiment come from a Kaggle competition brain tumor dataset that includes images of various types of brain tumor (glioma, meningioma, pituitary, and no tumor). This data set includes various MR images acquired under different imaging parameters, resulting in variations of intensity distribution, shape and appearance of the tumor and neighbouring tissues as a multidimensional space. These changes make it challenging to extract features needed to correctly diagnose cancer automatically, and require strong feature learning capabilities [23]. The analysis of these MRI scans shows the results of the scans in four categories, namely 'Glioma', 'Meningioma', 'Pituitary tumour' and 'Non tumour'. The data was split into 70%, 15%, and 15% for training, validation and testing models, respectively. The training set used to train the model, and the validation set used to determine the best set of model parameters to prevent over fitting. The test data set has been used for the ultimate performance evaluation.

Table 1. Distribution Of Mri Brain Tumor Images Used For Experimental Analysis

Class	Number of Images
Glioma	1321
Meningioma	1339
Pituitary	1457
No Tumor	1595

Proposed Hybrid CNN–Transformer Architecture

The processing was performed before the feature extraction and to enhance the overall image quality, thereby enhancing data coherence. To ensure consistent results with the model during the training, all the MRI images have been resized to 224×224 pixels. On the other hand, Transformer topologies, equipped with attention mechanisms, deliver better contextual understanding, but require more computational power [10, 19]. These limitations are addressed by proposing a framework that combines both architectures in one single feature learning model.

CNN-Based Feature Extraction

Enhancement of MRI images by techniques like, rotation, horizontal flipping, zooming, and brightness adjustments increased diversity in data and reduced over-fitting. The augmentation methods increased the variety of the data sets and improved both their robustness and their recovery in multiple imaging scenarios. For example, in healthcare imaging systems, MR pre-processing helps to ensure consistency of features, thereby enhancing the efficiency of learning [6] [23]. To preserve the complexity of the tumor while keeping the system lightweight, the convolutional layers use 3 × 3 kernels.

The mathematical modeling of a convolution operation is:

$$F(i,j) = \sum_m \sum_n I(i-m, j-n) K(m,n) \quad (1)$$

where Idenotes the input MRI image, Krepresents the convolution kernel, and F(i,j)indicates the generated

feature map. This operation permits to extract discriminative spatial representations related to tumor shape and tissue abnormalities.

Transformer-Based Contextual Learning

The extracted CNN feature-maps are further passed to the Transformer module for getting the contextual data. Generally, transformer models feature self-attention mechanisms which are able to transport long-range interactions among faraway areas of the MRI. The proposed model fuses the spatial information extracted from CNNs and the contextual information from transformers, improving the accuracy of the brain tumor detection in the magnetic resonance imaging MRI scans. Most of the traditional CNN based models are local feature extractors, and they do not have a global understanding of the spatial distribution of tumour regions.This contextual representation enhances the recognition of tumor groups that have similar appearance and different characteristics in tumor tissues [10].

The self-attention mechanism is represented as:

$$Attention(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

where Q, K, and Vrepresent query, key, and value matrices, respectively, while d_kdenotes the scaling factor. The attention mechanism emphasizes those regions that are rich with diagnostic information in the MRI, and enhances the interpretation of the MRI.

Feature Fusion Strategy

Then, a weighted feature fusion approach is used to combine the spatial features from the CNN module and the contextual representations obtained by the Transformer module. The proposed framework combines spatial local information and global contextual dependencies via the adaptive weighting scheme, which is different from conventional concatenation methods. This fusion technique increases the discriminability of features with less redundancy in the representations.

Description of feature fusion method:

$$F_{fusion} = \alpha F_{CNN} + \beta F_{Transformer} \quad (3)$$

where F_CNNand F_Transformerrepresent spatial and contextual feature representations, while αand βdenote adaptive weighting coefficients. The complementary feature representation of CNN and Transformer learning modules is improved by the fusion technique that is weighted, as demonstrated in Fig 1.

Classification Layer

All the features are encapsulated in the feature representation layer and the classification layer provides a number of tumor predictions. The final dense layer is for four outputs representing the four classes. Each tumor class is associated with a set of probability distribution via softmax activations. A method of optimization for categorical cross-entropy with the Adam optimizer that minimises classification loss. The framework was trained for 100 epochs with Adam optimizer and a learning rate of 0.001 and a mini batch size was taken as 32. The regularization was combined with dropout for regularization, reduction of over-fitting and improvement in the generalization capacity.

The classification loss function is given as:

$$Loss = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) \quad (4)$$

where y_irepresents the actual class label and y^_idenotes the predicted probability score. The effect is that by reducing the loss function they not only make the classification more accurate, but also ensure the stability of the convergence of the model in train.

Explainable AI Module and Clinical Decision Support

In order to make the approach clinically interpretable, the proposed approach comprises Explainable Artificial Intelligence understanding tools that consist of Grad-CAM visualization and attention heatmap analysis. In the healthcare sector, doctors must see the actual data to confirm that an AI algorithm is correct, before forming a diagnosis [1,3]. The explainability module generates heatmaps of the MRI regions significantly contributing to the prediction of the tumor, which help the radiologist to confirm the diagnosis.

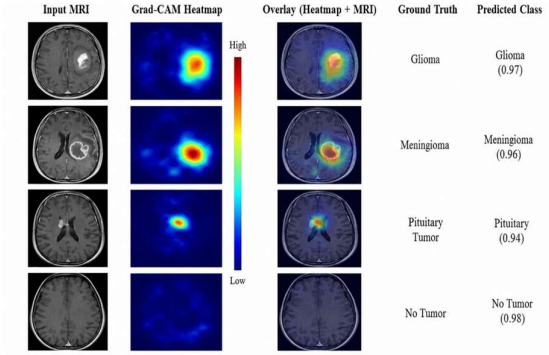

Fig. 2. Grad-CAM-based explainability visualization highlighting tumor regions influencing MRI classification results.

As it is seen clearly in the identify the tumor module based on Grad-CAM, by using it, the tumor locations in the brain which affect the predictions of the proposed CNN–Transformer system are highlighted and hence enhancing the transparency of the CNN–Transformer

system in healthcare decision making, as can be seen in Fig. 2. However, transformer-based designs enhance contextual learning capability which comes at the expense of an increased cost (more number) of computing expenses in the training of the model. However, to resolve this issue, the adaptive pooling method and the increase of feature dimensions were applied to reduce memory complexity as well as retain the classification performance. The proposed model enables combining CNN-based spatial learning, Transformer-based contextual representation, the integration of weighted feature fusion and explainable visualization mechanisms within one framework to achieve reliable and clinically interpretable intelligent prediction of brain tumors and support for healthcare decisions.

RESULTS AND DISCUSSION

A. Experimental Setup

In assessing the proposed Explainable Hybrid CNN-Transformer Framework, we analyzed MRI images of three types of brain cancers (Glioma, Meningioma, Pituitary tumor) with non-tumor pictures from the Kaggle Brain Tumor MRI dataset. Studies were conducted in a deep learning environment, where computations were accelerated by GPU computation to assist to train and infer the models. This is due to the flexibility and optimization of TensorFlow and Keras frameworks for tackling healthcare image analysis tasks [19, 21]. All MRI images were then resized to 224x224 image size, and were divided into training, validation and testing sets of 70%, 15% and 15%, respectively. Therefore, many studies have been carried out on multiclass tumour classification, to ensure consistency and not due to performance bias. The training time of the model is significantly shortened when using the GPU and the accuracy of optimization of the Transformer based contextual learning is enhanced. The proposed method has been successfully applied to the deep learning framework with the use of GPUs to train and test the models more quickly. Please look at Table 2 below.

B. Training Parameters

The adaptive learning capability and efficient convergence rate of the Adam optimizer made it the best choice for training, especially in medical image processing applications [18]. Batch normalization and dropout regularization were applied during the network training to deal with over fitting problem and to assure proper generalization.

Table 2. Hardware And Software Configuration Used For Experimental Evaluation

Parameter	Value
Processor	Intel Core i7 12th Gen
GPU	NVIDIA RTX 3060 (12 GB)
Framework	TensorFlow / Keras
Python Version	Python 3.10

An increased validation accuracy and reduced loss graphing were conducted during the training process. Early halting also to prevent unnecessary computation and to ensure model stability. The proposed structure was able to achieve comparable performance in learning, while the validation loss decreased slowly while optimizing the structure. In order to enhance convergence and classification performance, optimal learning parameters of the proposed model were used during training as obtained in Table 3 below.

Table 3. Hyperparameter Configurations Utilized For Training The Proposed Hybrid Cnn–Transformer Framework

Parameter	Value
Epochs	100
Batch Size	32
Learning Rate	0.001
Optimizer	Adam

C. Evaluation Metrics

A number of health care categorization criteria, including accuracy, precision, recall, F1-score, sensitivity and specificity [5] and [7] were tested for suitability of the proposed approach. Proposed measurements allows for a full assessment of the confidence of classification and consistency of diagnosis within a healthcare prediction system based on MRI.

- Accuracy evaluates the ability of the overall framework to correctly classify MRI images to their respective tumor types.
- Precision: It represents the ratio of correctly classified "tumor" cases compared to all cases in which the model detected a tumor.
- Recall provides an insight into the model's validity for detecting true tumor positive MRI images.
- F1 score is an excellent metric for multi-class healthcare classification problems because it brings together precision and recall together.
- Sensitivity represents the ability of the framework in detecting if there is a tumor or not in the MRI images.

• Specificity refers to its ability to accurately identify MRI images that do not contain a tumour and prevent false-positive results.

D. Performance Analysis

For multi-class categorization, the Explainable Hybrid CNN-Transformer Framework delivered good results on all the parameters of evaluation. Net Fusion between CNN and the Transformer model was performed to better classify the features in spatial learning for various types of tumors and contextual learning for contextual representation. An additional complimentary learning was improved via the weighted feature fusion method that was added.

Table 4. Classification Performance Of The Proposed Explainable Hybrid Cnn–Transformer Framework

Metric	**Value (%)**
Accuracy	98.42
Precision	98.11
Recall	97.94
F1-Score	98.02
Sensitivity	97.88
Specificity	98.67

The accuracy, precision, recall and F1 score results for the proposed framework on healing brain tumour with the multiclass brain tumour classification are described in Table 4 which presented good accuracy, precision, recall and F1 score values.Results demonstrated that the proposed framework had a good diagnostic MRI feature capture and possessed NIR classification performance on the huge number of tumor types. High specificity value assures its reliable recognition of MRI non-tumorous images and high sensitivity value assures very good abilities of tumor recognition. Standard CNN-based architectures were complemented by the proposed approach achieving improved contextual-knowledge of MRI tumor structures with the help of attention learning models using Transformers [10, 19].

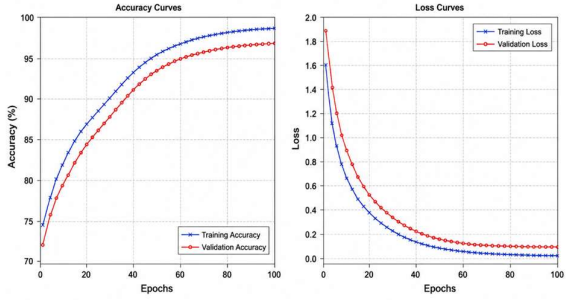

Fig. 3. Accuracy and loss curves for training and validation acquired during the optimization of the proposed framework.

In addition, there was a reduction in the amount of glioma and meningioma tumour type misclassification according to further confusion matrix analysis. Most of the classification errors were gained for MRI data having overlapping texture features, and non-uniform tumor contours. However the proposed system showed a stable prediction performance with various MRI imaging parameters. From the accuracy curve plot in training and validation (Fig. 3), it can be concluded that the training and validation accuracy curves slightly improved with decreasing loss values during the optimization. Dropout regularisation, augmentation operations and adaptive feature fusion were very helpful in the areas of convergence and preventing over fitting.

E. Comparative Analysis

The effectiveness of the categorization was conducted by incorporating the traditional architecture model CNN, ResNet, EfficientNet and Vision Transformer in the proposed model. Comparative results prove that proposed hybrid framework is able to learn simultaneously both spatial objects and objects in context, thus, it better surpasses the others. Table 5 shows that the proposed hybrid CNN–Transformer architecture achieved better performance compared to CNN, ResNet, EfficientNet and Vision Transformer.

Table 5. Comparative Classification Efficacy Of Deep Learning Models For Mri Analysis Of Brain Tumors

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**
CNN	93.84	93.26	92.97
ResNet	95.42	95.11	94.86
EfficientNet	96.31	95.88	95.76
Vision Transformer	97.08	96.94	96.52
Proposed Hybrid CNN–Transformer	98.42	98.11	97.94

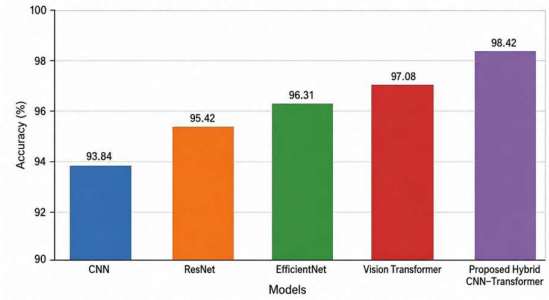

Fig. 4. Comparative accuracy analysis of CNN, ResNet, EfficientNet, Vision Transformer, and the proposed hybrid CNN–Transformer model.

The comparative results exhibit that lone CNN designs can precisely extract spatial tumor data, but have improper contextual awareness of the MRI area. Vision Transformer models vastly improve over global context understanding, but require more resources and complex training. The proposed framework was a hybrid, where the complementary strength of both architectures was used through a simple weighted feature fusion, helping to enhance the performance. The best classification accuracy of all the models tested was achieved by the suggested framework as shown in Fig. 4. The combination of CNN-based spatial representations and Transformer-based contextual learning has the greatest effect on the improvement of performance, mainly due to their enhancement of the identification of similar tumor types in the image.

F. Explainability Analysis

The Explanation study was conducted using Grad-CAM architecture visualization, attention heatmaps, and Shapley value interpretability of features with a view to obtaining more transparency to the predictions and better clinical interpretability [1,3]. In healthcare systems, visual explanation methods play a crucial role, as the medical community requires understandable content to aid in the process of diagnosis, which is supported by AI. The heatmaps generated from the proposed architecture using the Grad-CAM showed the regions in the MRI that made significant contributions to the results of predicting tumors. An activation map generated was consistent with the clinically important tumour site seen on radiological examination, enhancing the transparency of diagnosis and reliability of prediction.

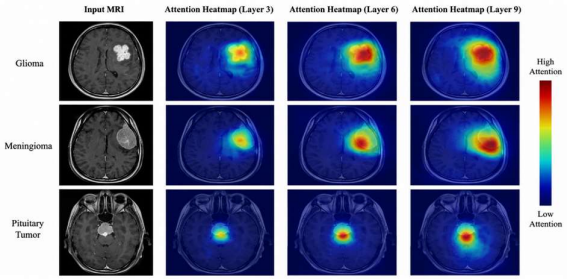

Fig. 5. Attention heatmap visualization showing diagnostically significant MRI regions identified by the Transformer module.

The heatmap analysis of the attention mechanism for the classification of tumors using the proposed Transformer based framework, shows the MRI regions that were more significant to the proposed model in the MRI classification task as can be seen on Fig. 5. The attention visualization also revealed that focusing on diagnostically relevant portions of the MRI was possible during contextual learning thanks to the Transformer module.

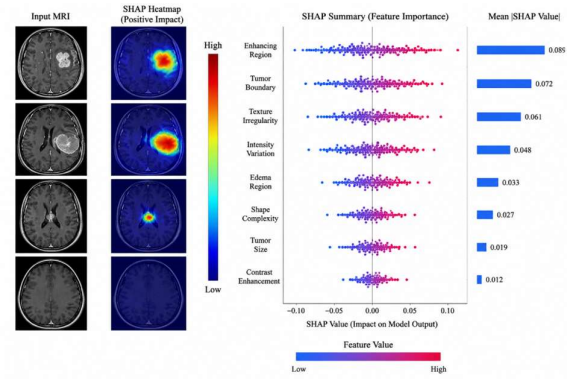

Fig. 6. SHAP-based feature relevance assessment for elucidating brain tumor categorization utilizing MRI images.

With the aid of SHAP further interpretations of features revealed, but also highlighted that the characteristics of the tumor boundaries, texture imperfections and fluctuations in intensity were significant for the results of categorization. SHAP provides feature-level explanation by highlighting the influence of tumor texture, edge and intensity features on prediction as illustrated in Fig. 6.

DISCUSSION

i. Advantages of the Proposed Framework

This proposed approach doesn't face the dilemma of embedding CNN-based spatial representation and Transformer-based sequencing into one framework. The feature fusion with weighted method significantly enhanced the multiclass tumor discriminating ability by better classification. Explainable AI strategies strengthened the clarity of predictions and also promoted self-confidence in the minds of health care specialists taking part in choice making.

ii. Clinical Importance

The design would offer trustworthy support in categorizing the tumors for radiologists and medical practitioners using MRI. The explainability module supports doctors in understanding the prediction behaviour, to validate the localization of the tumour and improve the confidence level to perform a clinical analysis.

iii. Limitations of the Proposed Work

The suggested model attained decent classification accuracy; however, the contextual learning provided by the Transformer resulted in higher computational complexity and GPU memory utilization as compared to CNN stand-alone architectures. Variations in MRI acquisition, imbalanced MRI datasets, etc. may affect

their generalization ability in large-scale clinical settings.

iv. Future Clinical Deployment Challenges

The proposed system will need to be expanded across multiple centers with MRI scans, and real-time integration into health care. It is necessary to optimize further for low-cost, efficient deployment and low computing cost to realize realistic clinical imaging systems.

CONCLUSION

For clinical decision making based on MRIs, the authors suggest a novel Explainable Hybrid CNN–Transformer Framework. The proposed technique utilizes CNN for spatial feature extraction and Transformer for the contextual learning approach to improve the performance in brain tumor multi-class classification. In addition, increased interpretability and transparency of the results of the diagnostic procedure with the help of EAI methods: Grad-CAM analysis, attention heat maps, SHAP interpretation. The experimental results revealed that the proposed framework can classify the image with better accuracy than the widely used model namely CNN, ResNet model, EfficientNet model, and Vision Transformer model.The weighted fusion method performed well for improving the visualization of local and global contextual information of the tumor in the MRI images. The explainability module also assisted with the detection of diagnostically relevant tumor locations which is crucial for radiological validation and improves the confidence of clinicians when making decisions during healthcare. The proposed framework obtains a good classification accuracy but the complexity of the Transformer-based learning model appears to be a challenge to large-scale applicability in clinical use. This involves finding ways of making computing more efficient in the future, such as larger datasets of mult-centre MRI data, as well as developing real time intelligent healthcare systems for real world clinical applications.

REFERENCES

[1] van der Velden BHM, Kuijf HJ, Gilhuijs KGA, Viergever MA. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. Medical Image Analysis. 2022;79:102470. DOI: 10.1016/j.media.2022.102470.

[2] Rao S, Li Y, Ramakrishnan R, Hassaine A, Canoy D, Cleland JGF, et al. An explainable transformer-based deep learning model for the prediction of incident heart failure. IEEE Journal of Biomedical and Health Informatics. 2022;26(7):3362-3372.

[3] Borys K, Schmitt Y, Nauta M, Seifert C. Explainable AI in medical imaging: An overview for clinical practitioners. European Journal of Radiology. 2023;162:110787.

[4] Mienye ID, Sun Y. A survey of explainable artificial intelligence in healthcare. Smart Health. 2024;32:100441.

[5] Sadeghi Z, Alizadehsani R, Cifci MA, Kausar S, Rehman R, Mahanta P, et al. A review of explainable artificial intelligence in healthcare. Computers in Biology and Medicine. 2024;178:108247.

[6] Muhammad D, Hossain MS, Alhamid MF. A systematic review of explainable artificial intelligence in medical image analysis. Ain Shams Engineering Journal. 2024;15(9):103091.

[7] Bharati S, Mondal MRH, Podder P. A review on explainable artificial intelligence for healthcare: Why, how and when? IEEE Transactions on Artificial Intelligence. 2024;5(4):1600-1618.

[8] Ghnemat R, Alkasassbeh M, Al-Hawari F. Explainable artificial intelligence for deep learning-based medical image classification. Journal of Imaging. 2023;9(9):177.

[9] Dorfner FJ, et al. A review of deep learning for brain tumor analysis in MRI. NPJ Precision Oncology. 2025;9:45.

[10] Hosny KM, Mohammed MA. Explainable AI and vision transformers for detection and classification of brain tumors using MRI images: A survey. Artificial Intelligence Review. 2025;58:112.

[11] Srinivas VR, et al. Explainable AI-driven MRI-based brain tumor classification. Frontiers in Artificial Intelligence. 2025;8:1700214.

[12] Iftikhar S, et al. Explainable CNN for brain tumor detection and classification using MRI images. Scientific Reports. 2025;15:12044100.

[13] Vamsidhar D, et al. Hybrid model integration with explainable AI for brain tumor detection using MRI images. Scientific Reports. 2025;15:6455.

[14] Gundogan E. A novel hybrid deep learning model enhanced with explainable AI for brain tumor multi-classification from MRI images. Applied Sciences. 2025;15(10):5412.

[15] Talaat FM. An efficient explainable AI model for accurate brain tumor detection using MRI images. Methods. 2025;245:105432.

[16] Abraham LA, et al. Exploring the potential of explainable AI in brain tumor detection and classification using MRI images. Artificial Intelligence Review. 2025;58:410.

[17] Mahadevan A, et al. Efficient and explainable MRI brain tumor classification via deep learning. Results in Engineering. 2025;26:104102.
[18] Rahman A, et al. Enhanced MRI brain tumor detection using transfer learning and pretrained deep CNN models. Scientific Reports. 2025;15:14901.
[19] Patel D, Roy S, et al. Attention-based transformer architectures for healthcare analytics and disease prediction. Artificial Intelligence in Medicine. 2023;142:102615.
[20] Chen X, Li H, Wang Y. Deep learning-based clinical decision support systems for smart healthcare environments. IEEE Access. 2023;11:45678-45692.
[21] Zhou T, Wang X, et al. Explainable multimodal healthcare intelligence using deep neural learning. IEEE Journal of Biomedical and Health Informatics. 2024;28(9):5120-5132.
[22] Gupta J, Seeja KR. A comparative study and systematic analysis of XAI models and their applications in healthcare. Archives of Computational Methods in Engineering. 2024;31(7):3977-4002.
[23] Satushe V, et al. AI in MRI brain tumor diagnosis: A systematic review of machine learning and deep learning approaches. Computer Methods and Programs in Biomedicine Update. 2025;6:100201.
[24] Aksoy S, et al. A web-deployed explainable AI system for brain tumor classification and segmentation using MRI scans. Hemato. 2025;17(8):121.
[25] Gomes EF, et al. Deep learning approaches for brain tumor classification using MRI imaging. Applied Sciences. 2026;16(2):831.