Applied Data Science and Analysis
Vol.2026, **pp**. 41–50
DOI: https://doi.org/10.58496/ADSA/2026/003; ISSN: 3005-317X
https://mesopotamian.press/journals/index.php/ADSA

Research Article

Interpretative Artificial Intelligence for Brain Tumor Detection and Classification

Karthik Kumar Vaigandla [1],*,

[1] *Associate Professor, Electronics and Communication Engineering, Balaji Institute of Technology and Science, Warangal, Telangana, India*

ARTICLE INFO

Article History
Received 17 Nov 2025
Revised 15 Dec 2025
Accepted 20 Jan 2026
Published 21 Feb 2026

Keywords
Brain Tumor,
Machine Learning,
Deep Learning,
Explainable AI,
AI,
Healthcare.

ABSTRACT

A brain tumor, a predominant source of unregulated cellular proliferation in the central nervous system, poses significant difficulties in medical detection and therapy. Timely and precise identification is crucial for successful intervention. Early malignancies within adult brains are universally lethal. Recent advancements in artificial intelligence (AI) within the realm of computer vision have facilitated the automated characterization and diagnosis of brain tumor lesions. The precise identification and categorization of brain tumors are essential elements of medical diagnosis. Recently, AI methodologies have gained prominence in improving brain tumor detection. Nonetheless, AI models frequently exhibit a deficiency in transparency, which poses significant challenges in critical domains such as healthcare. This study presents an Explainable AI (XAI) system designed for brain tumor identification, providing doctors with clear and interpretable insights into model decisions. This system utilizes advanced XAI techniques to enhance confidence, reliability, and clinical acceptance of AI-based tumor detection technologies. Creating explainable AI methodologies will be crucial for enhancing human-machine interactions and aiding in the identification of appropriate training techniques. Future endeavors will enhance the dataset and implement discoveries in real-time diagnostic equipment, therefore improving the discipline.

1. INTRODUCTION

The human brain has billions of neurons, synapses, and nerve cells that regulate several vital body activities. Similar to other body organs, the brain may experience issues, including the development of growths referred to as brain tumors. These tumors arise from aberrant cell proliferation in different areas of the brain, potentially resulting in severe neurological issues and adversely affecting a patient's quality of life [1]. Brain tumors are recognized as one of the most fatal illnesses worldwide, significantly affecting death rates across all age demographics. Such a disease causes problems with health care assignment and treatment [2]. The tumors are exhibited in various forms such as glioma, meningioma, pituitary tumors and the tumor free cases. There are almost 120 different varieties of tumor and diverse personalities with different sizes, and in addition, the complex structure of the brain works against them. It is important to have precise detection and segmentation in the tumorous area including edema, necrotic center, as well as neoplastic tissue in order to have proper diagnosis and treatment planning of the same [3]. The necessity of an automatic identification and classification of the medical images, particularly the brain tumor diagnosis, is the most important concern. The early diagnosis and classifications of the cancers that affect the brain are vital in early treatment and improved chances of patients. The developments in artificial intelligence (AI) have triggered opportunities in human life in a number of areas such as industry, business, education and healthcare. There are Deep learning (DL) techniques in AI that are useful in offering effective autonomous picture classification in medical applications [4] The traditional modeling strategies such as the linear regression and the decision trees provide a clear association between the inputs and the ensuing choices in the model runs [5]. Such models are also called white-box models; however, they often do not work as effectively as the black-box models, such as convolutional neural network (CNN), complex ensembles, and other DL models. The latest ones are more accurate yet lack explainability. Explainable models determine the importance of one characteristic compared to model predictive output which provides us with interpretable tools to understand DL outcomes [6]. XAI plays an imperative role in mitigating the bias in AI-based decisions. Modelling bias may occur even before training training or testing [7]. The information being used to train the model can be entangled with implicit bias. It follows that the need to identify and deal with potential biases in datasets is thus a critical part of any ethical AI strategy. The training process of AI should aim at building the trustworthy,

*Corresponding author. Email: vkvaigandla@gmail.com clear, and impartial models. This study provides an XAI framework to specifically address brain tumor diagnosis with an aim of addressing the current limitations of interpretability among medical solutions using AI. This is a technology that combines strong machine learning (ML) algorithms with XAI to produce accurate diagnostic predictions accompanied by interpretable explanations. This will increase credibility and collaboration between AI systems and medical practitioners. The critical elements of the project entail the collection of data, its preprocessing, DL model building, and the incorporation of XAI methods. The evaluation of the effectiveness of the proposed system will be conducted on the data on brain tumors in real-life setting in terms of accuracy, interpretability, and usability. Such a project can make a great impact upon the sphere of medical diagnosis, particularly, brain tumor detection. This technology gives doctors understandable knowledge of the AI-made predictions thus enabling well-informed decision-making process, reducing their dependence on subjective representation, and ultimately, patient care and outcomes.

Brain Tumor Detection: Identifying the presence of a tumor in brain images such as MRI or CT (Computed Tomography) scans. DL Models like CNNs are trained on vast datasets of annotated brain images to detect anomalies indicative of tumors. AI models identify spatial and textural features of the brain tissue. AI spots irregularities in brain scans that might signify tumors.

Brain Tumor Classification: Determining the type of tumor (e.g., benign or malignant). For malignant tumors, AI can further classify subtypes like Glioma, Meningioma, or Pituitary tumors. AI models (e.g., CNNs, Transfer Learning using ResNet, VGG, etc.) are trained to classify tumor types based on patterns and texture. Hybrid models combining DL with traditional ML (like SVM or Random Forest) enhance performance. AI learns the distinct features of each tumor class and outputs the probability of tumor types.

Brain Tumor Segmentation: Segmenting (isolating) the tumor region from healthy brain tissues for precise localization and analysis. This involves pixel-wise classification to distinguish tumor boundaries. AI uses models like U-Net, Mask R-CNN, or DeepLab for pixel-level tumor segmentation. AI divides the brain image into pixels and classifies each pixel as either belonging to the tumor or not. It uses layers of convolution to preserve spatial details and extract fine-grained tumor boundaries. Recent advances in medical AI, particularly DL-based diagnostic systems, have raised significant cybersecurity and trust concerns. Medical AI systems operate on sensitive patient data and are vulnerable to threats such as data poisoning, adversarial attacks, model inversion, and privacy leakage, which can compromise diagnostic integrity and patient safety. Ensuring the robustness, reliability, and security of AI-driven medical decision-making systems has therefore become a critical cybersecurity challenge. This study addresses these concerns by focusing on secure and trustworthy AI-based medical image analysis, contributing to Medical AI assurance through improved model reliability, attack resilience awareness, and secure deployment considerations.

2. LITERATURE REVIEW

Rapid and accurate detection of brain tumors is crucial for effective therapy. MRI scans are crucial to this technique; they may provide interpretative issues. DL has become an effective instrument for brain tumor identification, with recent advancements aimed at automating and improving the precision of brain tumor diagnosis. The suggested an Explainable AI approach towards Epileptic Seizure Detection model in [8] integrates CNN, RNN, and BERT. The dataset is initially evaluated on audio employing CNN and RNN. Decision-level fusion is utilized in a multimodal sentiment analysis model, succeeded by text analysis employing BERT. The integration of CNN and BERT yields superior performance compared to RNN. The authors intend to augment the model to incorporate photos and video. The research presents a XAI methodology to improve the interpretability of DL models in the classification of retinal OCT images [9]. Class activation maps, attention-based methods, saliency maps, etc, are proposed to show how the model focuses on important parts of the image. Evaluation of a public OCT longitudinal dataset shows improved accuracy and interpretability in comparison with the traditional DL approaches. The authors stress the potential of their approach in improving assessments of test of the retinal disorders. In [10] an XAI approach is presented to predict drug sensitivity in cancer cell lines. The authors use a deep neural network (DNN) to predict the results based on genetic characteristics and use Layer-wise Relevance Propagation (LRP) to calculate feature significance ratings to make the interpretation. Evaluation over a publicly available dataset shows superior performance in terms of accuracy and interpretability compared to the traditional ML models. The authors intend to suggest that their XAI framework is capable of helping to discover the key genetic traits and will help develop the personalized cancer treatment. In [11] suggest such a cascased CNN approach to the automated liver and tumors segmentation based on CT and MRI volumes. They use dilated convolution and residual connection in a two-stage structure to promote the accuracy of segmentation. Evaluation of publicly available datasets shows that its methodology has an accuracy and efficiency advantage over the traditional segmentation methods. The authors have emphasized the practical outcome of their approach to perform computer-aided diagnosis and treatment planning of liver cancer. The authors in [12] provide a new methodology that has applied XAI approaches to understand the spatio-temporal dynamics of the variables involved in the risk of COVID-19. They use complex ML algorithms to identify trends and connections between such variables. With their rule-based models and the analysis of feature importance, they provide considerable information and make interpretable explanations. The research presents the XAI4COVID model, which facilitates forecasts of COVID-19 cases and fatalities while emphasizing the significance and influence of all variables. A retinal disease categorization model for OCT images is provided in [13]. The authors want to improve interpretability and transparency by integrating a DNN with XAI approaches. The research attains significant accuracy with CNN models and uses the LIME framework to elucidate misclassifications. Comprehensive assessment, verification across many datasets, and incorporation of domain expertise are emphasized for practical application. The research in [14] seeks to improve the identification of liver tumors in MRI scans. The authors introduce MIMFNet, a multimodal detection framework that integrates local and global data across many scales and modalities to enhance accuracy and localization. The assessment validates the efficacy of the suggested framework. The authors propose future directions, including the assessment of bigger datasets, the integration of supplementary clinical data, the resolution of computing efficiency issues, and the investigation of interpretability and visualization methodologies. The emphasis is on using XAI methodologies in healthcare to improve the transparency and interpretability of AI models. Authors underscore the necessity of integrating XAI methodologies into healthcare AI systems to provide interpretable justifications for forecasts. Identified research needs including established assessment measures, ethical norms, integration of domain knowledge, and user-friendly interfaces. Future endeavors include the development of communication tools, the exploration of ensemble methodologies, and the assessment of XAI's influence on decision-making, patient care, and trust in AI systems [15]. DL techniques have exhibited unparalleled sensitivity and precision in tumor segmentation and malignancy classification [16]. Numerous research has investigated XAI methodologies to illustrate the learning progress at each neuronal layer [17]. The strategies produce saliency maps of identifying what the pixels/features contribute to the learning process. It has been suggested that XAI can increase the training power of the DL models through tracking of intermediate stages at the inter-neuron layers [18]. The clarity of DL predictions and final results may be considerably enhanced by post hoc explainable assessment of AI outcomes, thus increasing their predictability, understandability, and trustworthiness, as well as relations between humans and machines [19]. Further, the post hoc methods can be used to complement the study of clinical neurophysiology and neuro-imaging data such as the position of the brain sub-regions [20].

Brain Tumor MRI Classification Using Deep Learning: DL techniques, particularly CNNs, have been widely adopted for brain tumor classification using MRI. Early studies employed custom CNN architectures to differentiate between benign and malignant tumors, demonstrating improved accuracy over traditional machine learning methods. Recent works leverage deeper architectures and ensemble learning to enhance classification performance across multi-class tumor datasets. Public datasets such as BraTS and Figshare MRI collections have facilitated benchmarking and comparative analysis of classification models.

Transfer Learning in Brain Tumor Analysis: To address limited annotated medical data, transfer learning has become a dominant approach in brain tumor MRI classification. Pre-trained models such as VGG, ResNet, DenseNet, and EfficientNet have been fine-tuned on brain MRI datasets, achieving high accuracy with reduced training complexity. These approaches benefit from learned low-level features while adapting higher layers to domain-specific tumor characteristics.

Explainable AI in Medical Imaging : Explainable AI (XAI) has gained importance in medical imaging to improve the transparency and trustworthiness of deep learning models. Techniques such as Grad-CAM, LIME, and SHAP have been applied to brain tumor classification models to highlight discriminative regions influencing predictions. XAI methods assist clinicians in validating model decisions, thereby increasing clinical acceptance and supporting diagnostic confidence.

Security and Trust Considerations in Medical AI : Recent studies emphasize that medical AI systems are vulnerable to adversarial manipulation, data poisoning, and privacy attacks, which can compromise diagnostic reliability. Consequently, research has increasingly focused on robust and trustworthy AI models for healthcare applications. Integrating security-aware design and explainability mechanisms is considered essential for the safe deployment of brain tumor classification systems in clinical environments.

TABLE I. LITERATURE REVIEW

Ref.	Application	AI Technique	Dataset / Modality	Key Contribution	Explainability / XAI Method
[8]	Epileptic Seizure Detection	DL	EEG Signals	Proposed an explainable AI framework for seizure detection	SHAP, Feature Attribution
[9]	Retinal Disease Classification	CNN / DL	Retinal OCT Images	Investigated model interpretability for retinal disease classification	Grad-CAM, Saliency Maps
[10]	Drug Sensitivity Prediction	DNN	Cancer Cell Lines	Predicted drug response while providing model transparency	SHAP, Feature Importance
[11]	Liver and Tumor Segmentation	Cascaded Fully Convolutional Neural Networks (FCNs)	CT & MRI Volumes	Automatic segmentation of liver and tumors from volumetric images	Visual inspection of segmentation maps

[12]	COVID-19 Risk Factor Analysis	Spatio-Temporal DL	Epidemiological Data	Interpreted temporal and spatial risk factors for COVID-19	Feature attribution, Attention maps
[13]	Retinal Disease Classification	DNN	OCT Images	Combined CNN with explainability for retinal disease prediction	Grad-CAM
[14]	Liver Tumor Detection	Multi-Scale Multi-Modal CNN	MRI Images	Proposed a fusion network for accurate tumor detection	Saliency / Activation Maps
[15]	General Healthcare AI	Explainable AI Frameworks	Medical Data	Surveyed XAI techniques applied in healthcare systems	Multiple XAI methods reviewed
[16]	Brain Tumor Classification	DL	MRI Images	Surveyed multigrade brain tumor classification techniques	Overview of interpretable methods included
[17]	Interpretable Image Recognition	ProtoPNet	Image Data	ProtoPNet	Prototype-based interpretability
[18]	General Image Classification	CNN	ImageNet / Medical Images	Developed Grad-CAM for visual explanations	Grad-CAM (Gradient-weighted Class Activation Maps)
[19]	Medical Decision Support	XAI Models	Clinical / Medical Data	Developed human-centric explainable AI for decision support	Feature importance, visual explanations
[20]	Neuroscience / Behavioral Studies	XAI Techniques	Neurostimulation Data	Applied explainable AI for understanding brain-behavior relationships	Saliency maps, model interpretation

3. METHODOLOGY

The method applied in this research involves an accurate and systematic approach, illustrated in Figure 1, which offers a detailed account of the research methodology employed in this work. The model was trained using the brain tumor classification dataset [21], a comprehensive compilation of MRI images exclusively intended for the classification of brain tumors. The dataset comprises a diverse collection of brain MRI scans from individuals with various tumor types, including no tumor, glioma, meningioma, and pituitary tumors, amounting to a total of 2200 files. The dataset utilizes a multimodal imaging technique including MRI scans, guaranteeing a uniform and standardized data collecting method. This enhances the facts in this research, rendering it more relevant and credible. To maintain consistency, each image in the collection were downsized to dimensions of 150 × 150 × 3. Throughout training, the sequence of pictures was randomized to accelerate convergence and inhibit the CNN from remembering the training order. Figure 2 displays many samples from each category: DWI, FLAIR, T1, T2-weighted, and T1 Fat-Sat from the dataset.

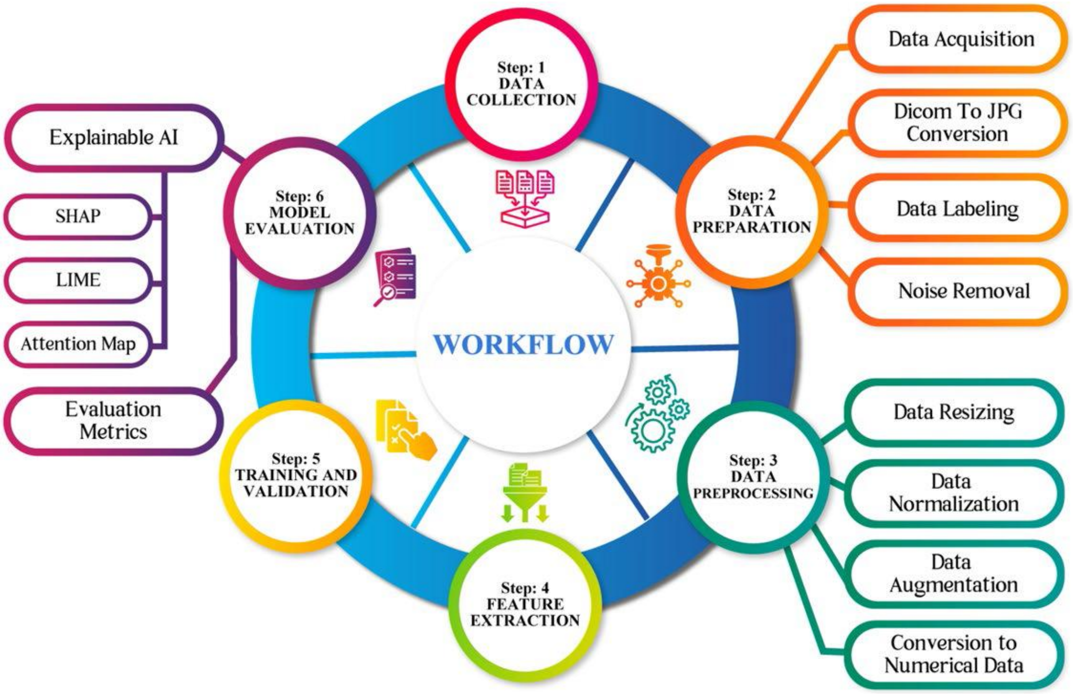

Fig. 1. Suggested approach delineating the specific phases encompassed in the study

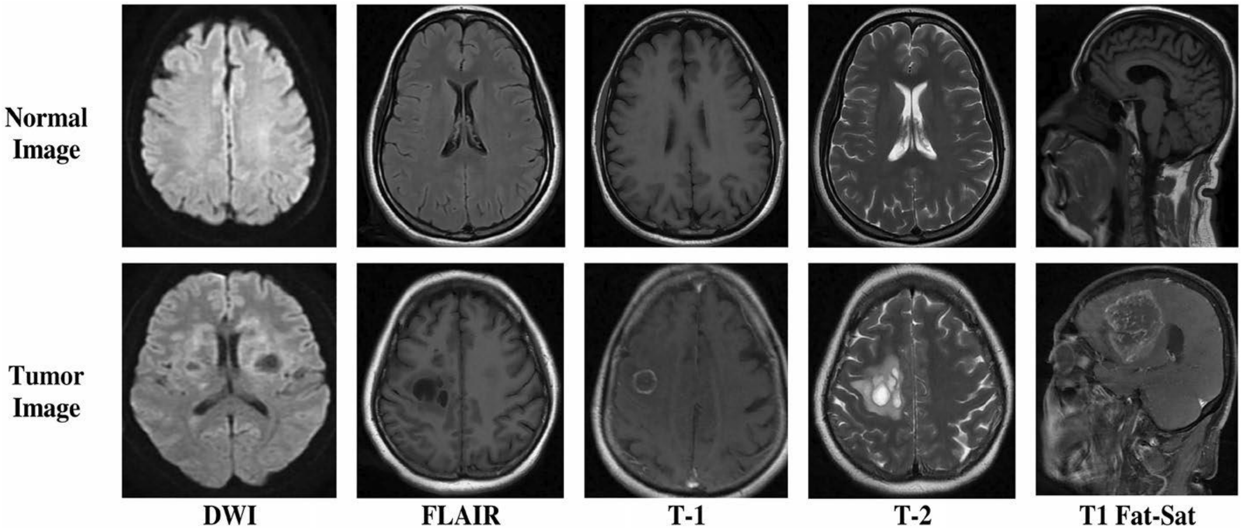

Fig. 2. Sample images

Preprocessing MRI pictures enhances their quality for DL analysis while preserving their integrity. Here utilized motion correction, data resizing, normalization, augmentation, and conversion to numerical format. Motion artifacts were addressed Employing retrospective and future motion correction techniques. Dimensionality decrease improves computational efficiency and model performance [22]. Normalization of pixel values to a range between 0 and 1 enhanced the accuracy of feature extraction. To rectify the data imbalance resulting from the disproportionate ratio of pathological to normal MRI pictures, utilized data augmentation techniques. This increased our dataset to 3264 MRI pictures, alleviating the deficiency and enhancing the learning and efficacy of our suggested model. The conversion of grayscale images to RGB format enabled the utilization of pre-trained models, hence improving performance. These preprocessing techniques guarantee optimum outcomes in subsequent picture classification challenges [23]. The suggested system, seen in Figure 3, has many stages: feature extraction, a CNN model, statistical performance indicators, and explanation production frameworks. To enhance precision, the CNN model was trained on two iterations of the dataset, with an output layer designed with a 1×4 configuration. The Adam optimizer, configured with default parameters, was employed for optimization, while the ReLU and softmax functions were utilized as activation functions. The final CNN model was utilized for several tasks, including assessing statistical correctness and producing explanations via LIME and SHAP methodologies. A gradient-based method was employed for SHAP explanations, whereas LIME explanations were obtained by perturbation calculations. A dual-input CNN network is a classification-based explainable model in the system. ReLU activation is used in all the hidden layers, represented by its ease of calculating and getting the values of the derivatives, that are either 0 or 1, based on the positivity of the inputs. The Adam optimizer was used with the default values whereas the loss function type was sparse categorical cross-entropy. In order to extract localities in input data, the convolutional layers have been trained using the kernel size of 3x3.

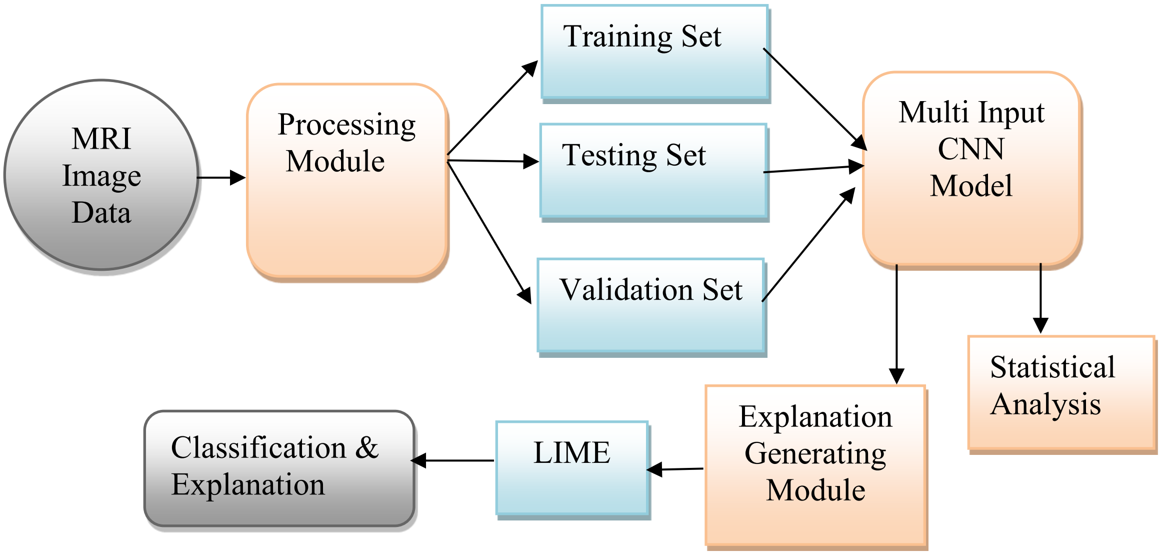

Fig. 3. Explainable AI for Brain Tumor Detection

Convolutional Neural Network (CNN) Architecture:

The proposed model employs a Convolutional Neural Network (CNN) designed to effectively extract discriminative spatial features from MRI images while maintaining a balance between representational capacity and computational efficiency [24-27]. The network consists of a sequence of convolutional, pooling, regularization, and fully connected layers. The CNN Architecture is shown in figure 4.

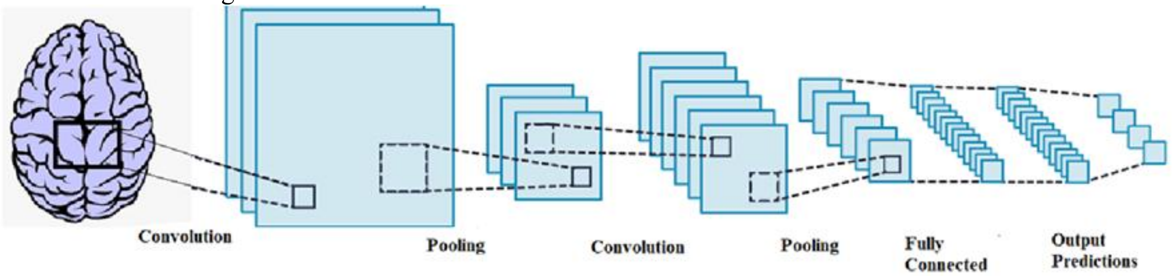

Fig. 4. CNN Architecture

Convolutional Layers : The CNN comprises three convolutional blocks. Each block includes a convolutional layer followed by a nonlinear activation function. The first convolutional layer uses 32 filters of size 3 × 3 with a stride of 1 and same padding to preserve spatial resolution. This layer captures low-level features such as edges, textures, and intensity variations. The second and third convolutional layers employ 64 and 128 filters, respectively, also with 3 × 3 kernels, enabling the extraction of more complex and abstract features such as tissue patterns and structural variations in brain MRI scans. Rectified Linear Unit (ReLU) activation is applied after each convolution to introduce non-linearity and accelerate convergence.

Pooling Layers : Each convolutional block is followed by a max-pooling layer with a pool size of 2 × 2. Max pooling progressively reduces the spatial dimensions of the feature maps, thereby lowering computational complexity, reducing the number of parameters, and providing translation invariance to small spatial shifts in the input images.

Dropout Regularization: To mitigate overfitting, dropout layers are introduced after the pooling operations and before the fully connected layers. Dropout rates of 0.25 in the convolutional blocks and 0.5 in the fully connected layers are used. During training, dropout randomly deactivates a fraction of neurons, forcing the network to learn more robust and generalized feature representations.

Fully Connected Layers: The flattened feature maps are passed to a fully connected (dense) layer consisting of 256 neurons, which integrates the learned spatial features into a compact representation suitable for classification. ReLU activation is used in this layer to maintain non-linearity. The final output layer is a dense layer with softmax activation, corresponding to the number of target classes, and produces normalized class probabilities.

Parameter Count and Model Complexity: The total number of trainable parameters in the CNN is determined by the number of filters, kernel sizes, and neurons in the dense layers. Convolutional layers contribute parameters in the form of filter weights and biases, while fully connected layers account for the majority of parameters due to their dense connectivity. The proposed architecture maintains a moderate parameter count, ensuring efficient training and inference while avoiding excessive model complexity that could lead to overfitting.

Algorithm:
Input: Brain MRI scans (T1, T2, FLAIR) Output: Tumor detection and classification with interpretable explanations
Step 1: Data Acquisition • Collect a dataset of brain MRI scans from public repositories or hospital databases. • Ensure the dataset includes labeled tumor types
Step 2: Data Preprocessing • Resize all MRI images to a fixed resolution • Normalize pixel intensity values to range [0,1]. • Apply noise reduction techniques • Perform data augmentation to increase diversity: rotation, flipping, scaling.
Step 3: Feature Extraction • Extract relevant features using CNNs (Use pre-trained CNNs or custom CNN layers) • Extract deep features from intermediate layers for interpretation. • Optionally, extract radiomic features: texture, shape, intensity, and histogram-based features.
Step 4: Tumor Detection • Apply CNN-based segmentation to identify tumor regions. • Generate binary mask of tumor area (0 → non-tumor, 1 → tumor)
Step 5: Tumor Classification • Input segmented tumor region into a classification network. • Classify tumor into categories. • Compute prediction probabilities for each class.
Step 6: Model Interpretation (XAI) • Apply interpretability techniques to explain predictions: • Overlay interpretability maps on original MRI for visual explanation.
Step 7: Evaluation • Evaluate model performance using metrics: • Verify interpretability maps with expert radiologists for clinical validation.
Step 8: Deployment • Deploy model in a clinical decision support system (CDSS). • Provide predictions with confidence scores and interpretability maps for physician review.

4. RESULTS

A wide range of measures is employed to quantify the performance of the proposed model, and they include accuracy, precision, recall, and the F1-score. These scores showing how the model is improving are calculated based on the data provided regarding confusion matrix i.e True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN), etc.

$$\text{Precision} = TP / (TP+FP)$$

$$\text{Recall} = TP / (TP+FN)$$

F1 - Score = 2 X Precision X Recall/ (Precision + Recall)
Accuracy = (TP + TN)/ (TP + TN+FP+FN)

XAI is important in AI systems that can be observed and understood and healthcare, where the choices of AI have an impact on the patient outcome. Since he was one of the first people to come up with medical picture classification, XAI makes human beings understand complex AI decisions. The interpretability of DL based models can be improved by using techniques such as LIME, SHAP, and Attention Maps to detect brain tumors. As a part of the current work, we resorted to the XAI techniques to enhance the explainability of the proposed model and provide substantial information about the way it makes a decision. The integration can help healthcare practitioners to know how the model produces its forecasts and this develops a sense of confidence and ease in making a better diagnostic and treatment decisions. In investigation, the model proposed was able to recognize positive situations maximally with an average accuracy estimation of 99.01% when the average precision score was 97.4%. The 97.2% average recall level justified the notion that the model can define the true positives appropriately without getting false negatives. The F1-score depicted the worthy performance of the model and earned a phenomenal average value of 97.1%, be conditioning the scores of accuracy and recall. To explore the training and performance power of the pre-trained transfer learning models, I have utilized them. The dataset was divided into training, validation, and testing subsets. Model training and hyperparameter tuning were performed exclusively on the training and validation sets. Final performance evaluation was conducted only on the held-out test set, and the reported results correspond to this final test evaluation. To ensure clarity and consistency, a single final test accuracy and associated metrics are reported throughout the manuscript. As Table 2 demonstrates the findings received our model performed extremely well. Figure 4 shows the performance analysis at various nodes and Training/Test Loss at Various Nodes is represented in Figure 5. To prevent data leakage and ensure unbiased performance evaluation, strict data separation protocols were followed. The dataset was split into training (70%), validation (15%), and testing (15%) subsets prior to any data augmentation. Data augmentation techniques were applied exclusively to the training set, and no augmented samples were included in the validation or test sets. The dataset used in this study provides image-level labels without explicit patient identifiers. To mitigate potential leakage risks, duplicate and near-duplicate images were checked and removed prior to dataset splitting. Furthermore, each image appears in only one subset, ensuring complete separation between training, validation, and testing data. All model hyperparameters were tuned using the training and validation sets only. Final performance metrics were computed solely on the held-out test set, which remained untouched during training and model selection. These measures collectively minimize data leakage and support reliable performance assessment.

TABLE II. FINDINGS OF 10-FOLD CROSS-VALIDATION

No. of Fold	Precision %	Recall %	F-1 Score %	Training Loss (ms)	Test Loss (ms)	Training Accuracy %	Test Accuracy %
1	98	98	98	0.0043	0.0107	100	100
2	96	96	96	0.0137	0.0581	99.8	98.3
3	98	97	97	0.0079	0.0317	99.9	99.2
4	99	99	99	0.0037	0.0096	99.7	100
5	100	100	100	0.003	0.0678	100	97.5
6	96	96	96	0.0041	0.048	100	98.3
7	97	97	97	0.0097	0.0381	99.6	99.1
8	98	97	97	0.0024	0.0149	100	100
9	96	96	95	0.0037	0.0565	100	98.4
10	96	96	96	0.0044	0.0371	99.8	99.3
Average	97.4	97.2	97.1	0.00569	0.03725	99.88	99.01

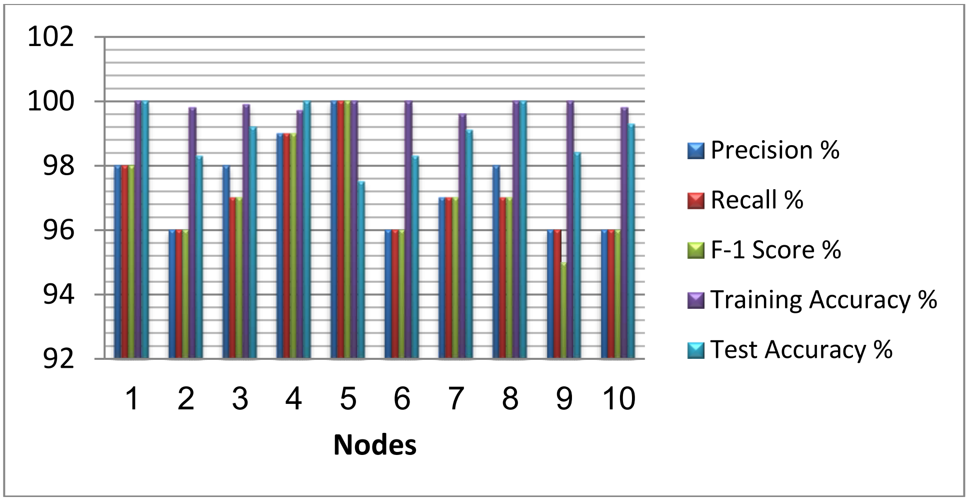

Fig. 5. Performance Analysis at Various Nodes

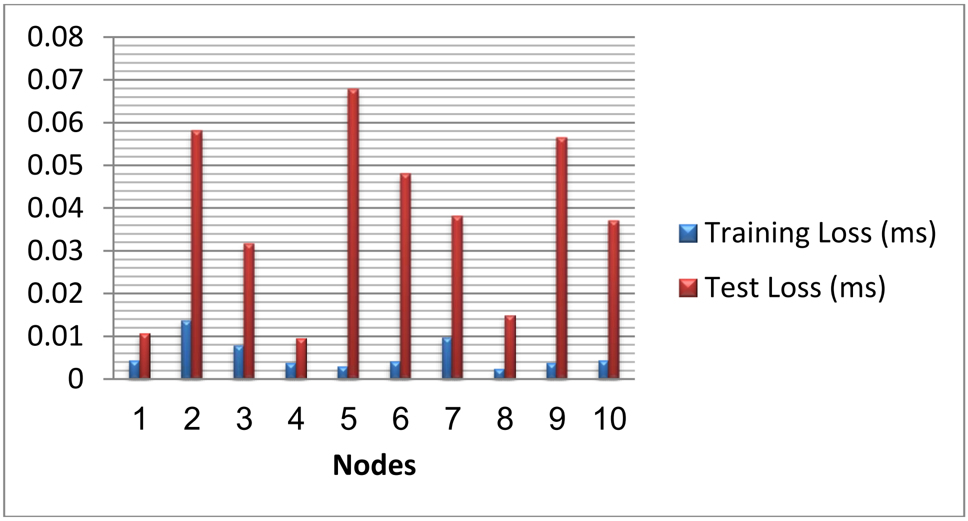

Fig. 6. Training/Test Loss at Various Nodes

The model was trained for 50 epochs, using a monitoring mechanism utilizing min mode and a patience level of three to prevent overfitting during CNN callbacks. Following training, the model produced 26 erroneous predictions, resulting in a training loss of 0.062. The model achieved 99.88% training accuracy and 99.01% final test accuracy, as shown in Figure 7.

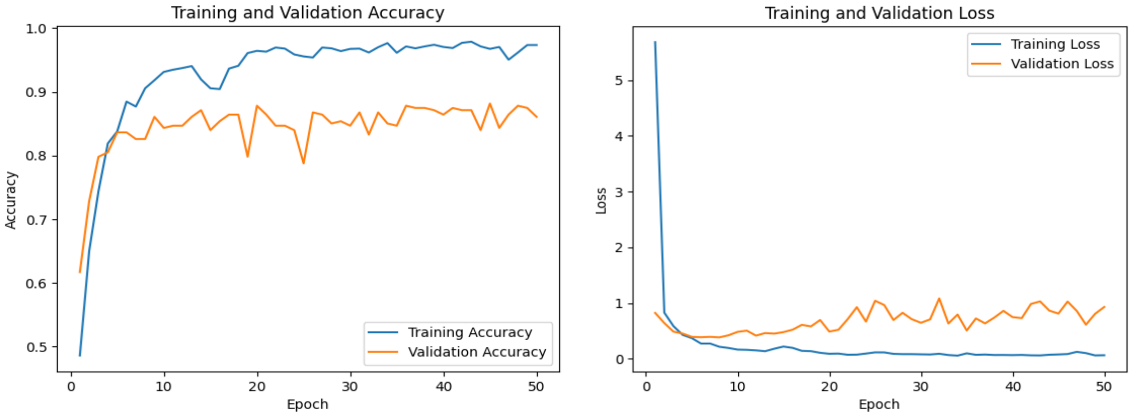

Fig 7. Training and Validation Accuracy and Loss

5. CONCLUSION

In conclusion, this work introduces and evaluates the performance of our newly created model for brain tumor detection and classification. I evaluated the model's performance by comparing it to numerous pre-trained models using transfer learning. The model attained an accuracy of 99.01% on the Brain Tumor Classification (MRI) dataset, indicating its capacity to reliably detect brain tumors and aid clinicians in emphasizing key regions, possibly saving time and lives. Future work entails increasing the dataset to improve accuracy and cover a wider spectrum of tumor types. Furthermore, I hope to increase the system's adaptability by allowing it to forecast cancers from a variety of medical imaging sources. I also intend to provide a user-friendly interface with thorough explanations so that non-specialists may comprehend and profit from the system. Ultimately, objective is to make this technology broadly available and useful in medical settings. In conclusion, I employed a well-known explainable AI algorithm to compare the performance of DL approaches for localizing tumor tissues in the brain. Incorporating explainable AI in DL workflows improves human-machine communication and can help identify the best training scheme for clinical issues and AI learning progress.

Conflicts of Interest
The authors declare no conflicts of interest.

Funding
No funding.

Acknowledgment
None.

References

[1] Celik, M. & Inik, O. Development of hybrid models based on deep learning and optimized machine learning algorithms for brain tumor multi-classification. Expert Syst. Appl. 238, 122159 (2024).
[2] Sharif, M., Amin, J., Raza, M., Yasmin, M. & Satapathy, S. C. An integrated design of particle swarm optimization (pso) with fusion of features for detection of brain tumor. Pattern Recogn. Lett. 129, 150–157 (2020).
[3] Sajid, S., Hussain, S. & Sarwar, A. Brain tumor detection and segmentation in mr images using deep learning. Arab. J. Sci. Eng. 44, 9249–9261 (2019).
[4] Battineni, G.; Sagaro, G.G.; Chinatalapudi, N.; Amenta, F. Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis. J. Pers. Med. 2020, 10, 21.
[5] Topol, E.J. High-performance medicine: The convergence of human and artificial intelligence. Nat. Med. 2019, 25, 44–56.
[6] Banapuram, C., Naik, A. C., Vanteru, M. K., Kumar, V. S., & Vaigandla, K. K. (2024). A Comprehensive Survey of Machine Learning in Healthcare: Predicting Heart and Liver Disease, Tuberculosis Detection in Chest X-Ray Images. SSRG International Journal of Electronics and Communication Engineering, 11(5), 155-169.
[7] Antoniadi, A.; Du, Y.; Guendouz, Y.;Wei, L.; Mazo, C.; Becker, B.; Mooney, C. Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review. Appl. Sci. 2021, 11, 5088.
[8] N. Chapatwala, C. N. Paunwala and P. Dalal, ‘An Explainable AI approach towards Epileptic Seizure Detection,’ 2022 IEEE 19th India Council International Conference (INDICON), Kochi, India, 2022, pp. 1-6, doi: 10.1109/INDICON56171.2022.10039982.
[9] T. S. Apon, M. M. Hasan, A. Islam and M. G. R. Alam, ‘Demystifying Deep Learning Models for Retinal OCT Disease Classification using Explainable AI,’2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), Brisbane, Australia, 2021, pp. 1-6, doi: 10.1109/CSDE53843.2021.9718400.
[10] I. S. Gillani, M. Shahzad, A. Mobin, M. R. Munawar, M. U. Awan and M. Asif, ‘Explainable AI in Drug Sensitivity Prediction on Cancer Cell Lines,’ 2022 International Conference on Emerging Trends in Smart Technologies (ICETST), Karachi, Pakistan, 2022, pp. 1-5, doi: 10.1109/ICETST55735.2022.9922931.
[11] Patrick Ferdinand Christ, ‘Automatic Liver and Tumor Segmentation of CT and MRI Volumes Using Cascaded Fully Convolutional Neural Networks,’2017 arXiv:1702.05970v2 [cs.CV]
[12] A. Temenos, M. Kaselimi, I. Tzortzis, I. Rallis, A. Doulamis and N. Doulamis, ‘Spatio-Temporal Interpretation of The Covid-19 Risk Factors Using Explainable Ai,’ IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia,
2022, pp. 7705-7708, doi: 10.1109/IGARSS46834.2022.9884922

[13] M. T. Reza, F. Ahmed, S. Sharar and A. A. Rasel, ‘Interpretable Retinal Disease Classification from OCT Images Using Deep Neural Network and Explainable AI,’ 2021 International Conference on Electronics, Communications and Information Technology (ICECIT), Khulna, Bangladesh, 2021, pp. 1-4, doi: 10.1109/ICECIT54077.2021.9641066.

[14] C. Pan et al., ‘Liver Tumor Detection Via A Multi-Scale Intermediate Multi-Modal Fusion Network on MRI Images,’2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 2021, pp. 299-303, doi: 10.1109/ICIP42928.2021.9506237.

[15] U. Pawar, D. O’Shea, S. Rea and R. O’Reilly, ‘Explainable AI in Healthcare," 2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA), Dublin, Ireland, 2020, pp. 1-2, doi: 10.1109/CyberSA49311.2020.9139655.

[16] Muhammad, K.; Khan, S.; Del Ser, J.; de Albuquerque, V.H.C. Deep Learning for Multigrade Brain Tumor Classification in Smart Healthcare Systems: A Prospective Survey. IEEE Trans. Neural Netw. Learn. Syst. 2020, 32, 507–522.

[17] Chen, C.; Li, O.; Tao, C.; Barnett, A.J.; Su, J.; Rudin, C. This Looks Like That: Deep Learning for Interpretable Image Recognition. arXiv 2018, arXiv:1806.10574.

[18] Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. Int. J. Comput. Vis. 2019, 128, 336–359.

[19] Knapiˇc, S.; Malhi, A.; Saluja, R.; Främling, K. Explainable Artificial Intelligence for Human Decision Support System in the Medical Domain. Mach. Learn. Knowl. Extr. 2021, 3, 740–770.

[20] Fellous, J.-M.; Sapiro, G.; Rossi, A.; Mayberg, H.; Ferrante, M. Explainable Artificial Intelligence for Neuroscience: Behavioral Neurostimulation. Front. Neurosci. 2019, 13, 1346.

[21] Sartaj Bhuvaji, Ankita Kadam, Prajakta Bhumkar, Sameer Dedge, amp; Swati Kanchan. (2020).Brain Tumor Classification (MRI) [Data set]. Kaggle.

[22] Spieker, V. et al. Deep learning for retrospective motion correction in MRI: A comprehensive review. IEEE Trans. Med. Imaging (2023).

[23] Ali, M. S. et al. Alzheimer’s disease detection using m-random forest algorithm with optimum features extraction. In 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA) (ed. Ali, M. S.) 1–6 (IEEE, 2021).

[24] Selvan, P., Kavitha, A., Kavitha, V., & Vaigandla, K. K. (2025, May). Hybrid CNN-BI-LSTM Network for the Accurate Brain Tumour Prediction in MRI Image. In *2025 Global Conference in Emerging Technology (GINOTECH)* (pp. 1-7). IEEE.

[25] Vaigandla, K. K., Padakanti, K. K., Pothkanuri, L., & Nanda, D. (2025, November). A systematic review on cancer and lung cancer detection using machine learning. In *AIP Conference Proceedings* (Vol. 3394, No. 1, p. 020003). AIP Publishing LLC.

[26] Vaigandla, K. K. (2025). Role of IoT and ML in Healthcare. *Babylonian Journal of Artificial Intelligence*, *2025*, 23-36.

[27] Banapuram, C., Naik, A. C., Vanteru, M. K., Kumar, V. S., & Vaigandla, K. K. (2024). A comprehensive survey of machine learning in healthcare: Predicting heart and liver disease, tuberculosis detection in chest X-ray images. *SSRG International Journal of Electronics and Communication Engineering*, *11*(5), 155-169.