An Explainable Machine Learning Model for Early Detection of Brain Tumors: Integrating Multi-Modal Medical Imaging and Intelligent Feature Fusion

S. Rukmani Devi
Department of Computer Science, Saveetha College of Liberal Arts and Sciences, SIMATS Deemed to be University, Saveetha Nagar, Thandalam, Chennai, India
rukmanibaveshnambi@gmail.com

J. Jean Justus
Department of CSE, SRM Institute of Science and Technology, Ramapuram, Chennai, India
jeanjustusj@gmail.com

M. Vanathi
Department of CSE, Sathyabama Institute of Science and Technology, Chennai, India
mercyvanu@gmail.com

Bobba Veeramallu
Department of Computer Science and Engineering, Koneru Laxmaiah Education Foundation, Vijayawada, Andhra Pradesh, India
bvmallu@kluniversity.in

V. Aruna
Department of Management Studies, St.Joseph's Institute of Technology, Chennai, India
arunasivakumar28@gmail.com

T. C. Manjunath
Department of CSE, Rajarajeswari College of Engineering, Bangalore, Karnataka, India
tcmanju@iitbombay.org

Valisher Sapayev Odilbek Uglu
Department of General Professional Subjects, Mamun University, Khiva, Uzbekistan
sapayev_valisher@mamunedu.uz

Banu S.
Department of CSE, KCG College of Technology, Chennai, India
banuphd2023@gmail.com (corresponding author)

Received: 16 April 2025 | Revised: 12 May 2025, 2 June 2025, and 10 June 2025 | Accepted: 14 June 2025

 DOI: https://doi.org/10.48084/etasr.11539

ABSTRACT

The early diagnosis of Brain Tumors (BT) is a critical challenge in medical imaging. This study proposes an explainable machine learning (XAI) framework that integrates multimodal imaging, including Magnetic Resonance Imaging (MRI) and Computed Tomography (CT), for accurate and interpretable BT

detection. A hybrid feature extraction strategy was employed, combining deep learning-based spatial features with handcrafted texture descriptors, including GLCM and LBP. These features are fused using an attention-based mechanism to enhance discriminative performance. The refined features are classified using an ensemble of Random Forest, XGBoost, and Deep Neural Networks. Explainability is incorporated using SHAP and Grad-CAM to visualize the model's decision rationale. Experiments on publicly available datasets demonstrate superior performance, achieving 97.3% accuracy, 96.4% precision, 96.0% recall, and 96.2% F1-score, outperforming existing methods while ensuring clinical interpretability.

Keywords-explainable machine learning; brain tumor; classification; feature fusion; multimodal; accuracy; neural network

I. INTRODUCTION

Brain Tumors (BT) are a type of severe neurological disease that dominates cognitive and motor functions, ultimately causing deadly complications [1]. In BT, the type, location, and growth rate determine the severity, so it is critical to detect early to achieve effective treatment and good patient survival. Magnetic Resonance Imaging (MRI) offers better soft tissue contrast, while Computed Tomography (CT) provides high-resolution images, which are predominantly used in disease prediction. Since manual interpretation is time-consuming and prone to errors, there is a need for automated diagnostic solutions [2]. Radiologists traditionally utilize medical imaging techniques, such as MRI and CT scans, to detect and classify tumors. Since they can provide superior soft tissue contrast, MRI scans are used to identify tumor boundaries and structures. However, CT scans are also good at producing high-resolution images to detect abnormalities in brain tissue. However, the effectiveness of these imaging modalities requires manual interpretation by radiologists, which is tedious and prone to human error. The diagnosis is highly variable and depends greatly on the experience and expertise of the radiologist [3]. Combining MRI and CT imaging provides a more comprehensive representation of brain anatomy, enhancing tumor visibility, especially in regions where either modality alone may be insufficient.

With the advancement of Artificial Intelligence (AI) and Machine Learning (ML), large-scale data can now be used to improve BT detection [4]. Classifying tumors such as gliomas, meningiomas, and pituitary tumors has been performed accurately using DL, especially Convolutional Neural Networks (CNN) and transformer-based architectures. However, DL models are black boxes, and healthcare professionals are concerned about the lack of transparency and justification of model predictions [5]. To address these shortcomings, XAI can be used to enhance interpretability by providing explanations of the decision-making process.

This study presents an XAI model for tumor detection using Multi-Modal Medical Imaging (MMMI), MRI, and CT scans, to increase the accuracy of BT detection. An intelligent Feature Fusion Mechanism (FFM) is adopted to extract meaningful data and improve model performance. The model is more interpretable using explainability methods such as Shapley Additive Explanations (SHAP) and Gradient-weighted Class Activation Mapping (Grad-CAM). The proposed method bridges the gap between AI-based automation and clinical trust, making ML solutions more viable for real-world medical applications. Table I compares methods such as CNNs, transfer learning, ensemble learning, and ML classifiers.

TABLE I. ANALYSIS OF LITERATURE REVIEW

Ref.	Techniques	Outcome	Inference	Drawbacks
[6]	DNN, MobileNetV2 for FE, M-SVM	97.47% accuracy on BraTS and 98.92% on Figshare	Improves tumor classification	Requires enormous computational resources
[7]	GLCM, FE, RF classifier	Effective BT detection with high sensitivity	Robust classification for medical images	Lacks DL-based FS
[8]	HOG, modified ResNet50	88% accuracy with HOG and ResNet50	HOG improves FS in DL	Moderate accuracy - low compared to CNN
[9]	EfficientNet-B0	98.87% accuracy	Improves classification	High dependency on hyperparameter tuning
[10]	VGG-16 CNN,	98.5% accuracy with VGG-16, 98.14%	Enhances predictive accuracy	Requires extensive computational resources
[11]	CNN, CRF	99.88% accuracy, 99.86% sensitivity, 99.96% specificity	Residual connections improve segmentation	Limited validation across multiple datasets
[12]	CLAHE, K-means, PSO, XGBoost	97% accuracy, 98% precision,	Enhances classifier performance	Relatively lower classification results

II. PROPOSED METHOD

The proposed XAI method for early BT detection unifies multimodal imaging, feature fusion, and explainability to improve classification accuracy and interpretability. Paired MRI and CT images were obtained from The Cancer Imaging Archive (TCIA) [13], a centralized repository that provides public access to curated medical imaging datasets. Specifically, the TCGA-GBM (The Cancer Genome Atlas Glioblastoma Multiforme) [13], Br35H dataset [14], and CPTAC-GBM (Clinical Proteomic Tumor Analysis Consortium GBM) [15] cohorts were used, which offer pre-diagnosed, multi-modal MRI and CT brain scans along with the corresponding clinical and genomic metadata. These datasets are widely used in research for glioblastoma segmentation and classification tasks due to their multimodal completeness and clinical annotations. The Br35H dataset [14] is used for MRI-only classification tasks, as it provides high-resolution, labeled brain tumor images categorized into glioma, meningioma, pituitary tumor, and no tumor. The dataset supports balanced representation across tumor classes and is suitable for training DL models in diagnostic imaging.

The proposed method includes three main components: Feature Extraction (FE), attention-based Feature Selection (FS) and fusion, and ensemble classification. A two-step fusion process is employed, where the handcrafted and deep features are initially concatenated into a unified feature vector. Subsequently, an attention-based weighting mechanism is applied. This mechanism calculates importance scores for each feature using a trainable attention layer, enhancing relevant features while suppressing noise and redundancy, thereby improving classification performance. Given a set of medical images from multiple modalities (e.g., MRI/CT), the input image set is defined as:

$$I = [I_1, I_2 \ldots . I_n] \quad (1)$$

where I_i indicates an individual image corresponding to a particular modality. Handcrafted features, such as Gray-Level Co-occurrence Matrix (GLCM), Local Binary Patterns (LBP), and HOG, are effective in capturing textural nuances, including tissue heterogeneity and texture coarseness, which are diagnostically significant. These are complemented with deep features from ResNet50 and EfficientNet, which capture abstract spatial patterns. Each image uses DL-based FE as f_θ, transforming the raw image into a feature vector:

$$X_i = f_\theta(I_i), \quad X_i \in \mathbb{R}^d \quad (2)$$

where d is the dimension of the FE space. Statistical and texture-based features H_i are extracted using traditional FS methods such as GLCM and LBP:

$$H_i = g(I_i), \quad H_i \in \mathbb{R}^{\mathrm{m}} \quad (3)$$

where m is the dimension of the handcrafted feature vector. The features extracted are concatenated to form an initial feature representation:

$$F_i = [X_i, H_i] \in \mathbb{R}^{(d+m)} \quad (4)$$

An attention mechanism is introduced to optimize FS, which assigns importance weights to features, reducing redundancy and enhancing discriminative capability. The attention weight for each feature dimension j is calculated as:

$$\alpha_j = \frac{Exp(W_j \cdot F_i)}{\sum_{k=1}^{d+m} Exp(W_k \cdot F_i)} \quad (5)$$

where W_j is the trainable attention weight corresponding to the j^{th} feature. Then, the attention-weighted feature vector is calculated as:

$$\widehat{F_i} = \sum_{j=1}^{d+m} \alpha_j \cdot F_{ij} \quad (6)$$

where $\widehat{F_i}$ is the optimized feature vector for classification.

The refined feature vector $\widehat{F_i}$ is fed into an ensemble of classifiers comprising Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Deep Neural Networks (DNN). The prediction outputs from individual classifiers are assumed as:

$$\mathcal{Y}_{RF} = h_{RF}(\widehat{F_i}) \quad (7)$$

$$\mathcal{Y}_{XGBoost} = \mathcal{Y}_{XGBoost}(\widehat{F_i}) \quad (8)$$

$$\mathcal{Y}_{DNN} = h_{DNN}(\widehat{F_i}) \quad (9)$$

where $h_{RF}, \mathcal{Y}_{XGBoost}, h_{DNN}$ are the prediction functions of RF, XGBoost, and DNN. The final classification output Y_{final} is obtained using a majority voting mechanism:

$$Y_{final} = arg\ max_c \sum_k 1(y_k = c) \quad (10)$$

where 1 is the indicator function, ensuring that the most frequent class label is selected.

Shapley Additive Explanations (SHAP) calculate feature importance by distributing the prediction difference among input features:

$$\phi_j = \sum_{S \subseteq F\{j\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f(S \cup \{j\}) - f(S)] \quad (11)$$

where ϕ_j is the SHAP value for the feature j, indicating its contribution to the classification outcome. Grad-CAM produces heatmaps that highlight the spatial areas of the image most relevant to the model's decision, aiding radiologists in understanding and verifying predictions. Grad-CAM generates visual explanations by calculating gradients of the output class score Y_c for the feature maps of the last convolutional layer as:

$$L_c^k = \sum_{i,j} \frac{\partial Y_c}{\partial A_{i,j}^k} \quad (11)$$

where L_c^k is the Grad-CAM heatmap that highlights the most significant regions in the medical image.

The hybrid FE process combines both handcrafted and deep features. Handcrafted features, using GLCM and LBP, capture texture-based patterns crucial for tumor characterization. Additionally, HOG is used to extract edge-based spatial representations. These are complemented by deep features derived from CNN-based encoders, which capture high-level semantic structures in the image. Grad-CAM produces heatmaps that highlight the spatial areas of the image most relevant to the model's decision, aiding radiologists in understanding and verifying predictions.

III. RESULTS AND DISCUSSION

This study utilized a total of 4,500 2D paired MRI-CT brain tumor slices from 41 patients, obtained from TCGA-GBM [13] and CPTAC-GBM [15], ensuring that the time interval between scans did not exceed six months. These images underwent intensity normalization, registration in the MNI152 template, and extraction of 2D slices from 3D volumes. Additionally, 7,023 MRI-only images, spanning four classes (glioma, meningioma, pituitary tumor, and no tumor), were extracted from the Br35H dataset [13]. All images were preprocessed using histogram equalization, skull stripping, and standard augmentation techniques. The CPTAC-GBM dataset [15] was split into 70:15:15 for training, validation, and testing. Handcrafted and DL-based FE was used on this dataset to train, validate, and test ML models for automatic BT detection and classification [16-17]. MRI-CT pairings were filtered to ensure a clinical interval of less than six months. All MRI volumes were registered on the MNI152 template using affine transformations. Intensity normalization and skull stripping were applied to minimize inter-scan variability. Two-dimensional slices were selected to maximize tumor visibility, based on available segmentation masks.

Python 3.9, TensorFlow 2.8, PyTorch 1.12, and OpenCV were used to implement the model. MONAI and SimpleITK handle medical imaging tasks. ResNet50, EfficientNet, GLCM, LBP, and HOG were used for FE. The training and testing setup used a multi-modal learning model based on MRI-CT paired slices and standalone MRI images. The BT dataset (4,500 MRI CT slices) was split into 70% training, 15% validation, and 15% testing for robust generalization. The 7,023 MRI-only dataset was split into 80% training, 10% validation, and 10% testing for classification into glioma, meningioma, pituitary, and no tumor. DL (CNNs, transformers) and ensemble models (RF, XGBoost) were used in training. Figure 1 shows a preprocessed and segmented image.

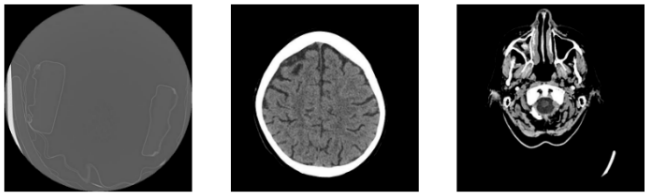

Fig. 1. Brain CT image, preprocessed and segmented.

This study also used other models to compare the proposed one. ResNet50, which is configured as a 50-layer residual network pre-trained on ImageNet, was fine-tuned using the MRI-CT multimodal dataset. Training was carried out with a batch size of 32, a learning rate of 0.0001, and Adam optimizer for 1500 epochs. EfficientNet-B0, a compound-scaled CNN, was trained from scratch using RMSProp with a dropout rate of 0.2 and L2 regularization. This model showed faster convergence and lower overfitting compared to other CNNs. Similarly, VGG-16 was used as a baseline DL architecture. The initial layers were frozen, and the last three convolutional blocks were trained along with a new classifier head. All models were implemented using NVIDIA RTX 3090 and A100 GPUs. Four key metrics were used to evaluate the performance of the existing and proposed models: accuracy, precision, recall, and F1-score [18]. Accuracy measures how well the model works by dividing the number of correctly classified cases by all predictions, as shown in Figure 2. Since accuracy alone may be misleading in imbalanced datasets, other performance metrics were also utilized. Precision defines the reliability of positive predictions, i.e., the proportion of correctly predicted positive cases among all positive cases. Sensitivity measures the model's capacity to correctly identify all actual positive cases with few False Negatives (FN). F1-score gives a harmonic mean of precision and recall and is a balanced metric when there is a trade-off between False Positives (FP) and FN. Together, these metrics determine the effectiveness of the model, especially in medical imaging, where minimizing misclassification errors is critical to ensure a reliable diagnosis. Table II shows the results of the models tested. The accuracy across models improves with epochs. The proposed XAI model consistently outperformed all other methods, achieving 97.3% accuracy at 1500 epochs, followed by Efficient-Net (94.2%) and ResNet50 (92.9%). Traditional ML models, such as SVM (83.2%) and RF (86.3%), showed comparatively lower performance. CNN-based models, such as VGG16 (93.8%) and CNN (91.5%), improve significantly with training but still fall short of the proposed XAI model. As shown in Figure 3, precision follows a similar trend. The proposed XAI model achieves 96.4% precision at 1500 epochs, showing its superior classification capability. Efficient-Net (93.1%) and VGG16 (92.7%) closely follow. SVM and RF achieved lower precision values of 81.7% and 84.2%, respectively. Among DL models, CNN reached 90.3%, but the proposed XAI model still outperformed it.

TABLE II. PERFORMANCE COMPARISON

Method	Epoch	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
SVM (handcrafted features)	300	78.6	76.2	75.9	76
	600	80.1	78.4	77.9	78.1
	900	81.5	79.7	79.3	79.5
	1200	82.4	80.9	80.4	80.6
	1500	83.2	81.7	81.3	81.5
RF	300	81.3	79.1	78.5	78.8
	600	82.7	80.5	79.8	80.1
	900	83.9	81.6	81.1	81.3
	1200	85.1	83	82.5	82.7
	1500	86.3	84.2	83.7	83.9
ResNet50	300	88.9	87.3	86.5	86.9
	600	90.1	88.7	87.9	88.3
	900	91.2	89.9	89.3	89.5
	1200	92.1	91	90.5	90.7
	1500	92.9	91.7	91.2	91.4
Efficient-Net	300	91.2	90	89.7	89.9
	600	92.3	91.1	90.8	90.9
	900	93	91.9	91.5	91.7
	1200	93.7	92.6	92.2	92.4
	1500	94.2	93.1	92.8	92.9
VGG16	300	90.8	89.5	89.1	89.3
	600	91.9	90.7	90.3	90.5
	900	92.5	91.4	91	91.2
	1200	93.2	92.1	91.7	91.9
	1500	93.8	92.7	92.3	92.5
CNN	300	87.1	85.6	84.9	85.2
	600	88.4	87	86.3	86.6
	900	89.5	88.2	87.6	87.9
	1200	90.6	89.4	88.8	89
	1500	91.5	90.3	89.8	90
XGBoost	300	85.2	83.7	82.9	83.3
	600	86.5	85	84.2	84.6
	900	87.7	86.3	85.6	85.9
	1200	88.9	87.6	87	87.3
	1500	90	88.7	88.1	88.4
Proposed XAI method	300	94.8	93.6	93.1	93.3
	600	95.5	94.4	94	94.2
	900	96.1	95.2	94.8	95
	1200	96.7	95.8	95.4	95.6
	1500	97.3	96.4	96	96.2

F1-score follows a similar trend with DL models outperforming traditional ones. The proposed XAI model achieved a 96.2% F1-score, significantly ahead of EfficientNet (92.9%) and VGG16 (92.5%). CNN and XGBoost achieved competitive scores of 90.0% and 88.4%, respectively, while SVM was the weakest at 81.5%. Recall also increased with more training epochs. The proposed XAI model achieved 96.0% recall at 1500 epochs, outperforming EfficientNet (92.8%) and ResNet50 (91.2%). CNN showed competitive recall at 89.8%, but SVM and RF remained lower at 81.3% and 83.7%. In general, the proposed XAI model outperformed the VGG16 and EfficientNet models.

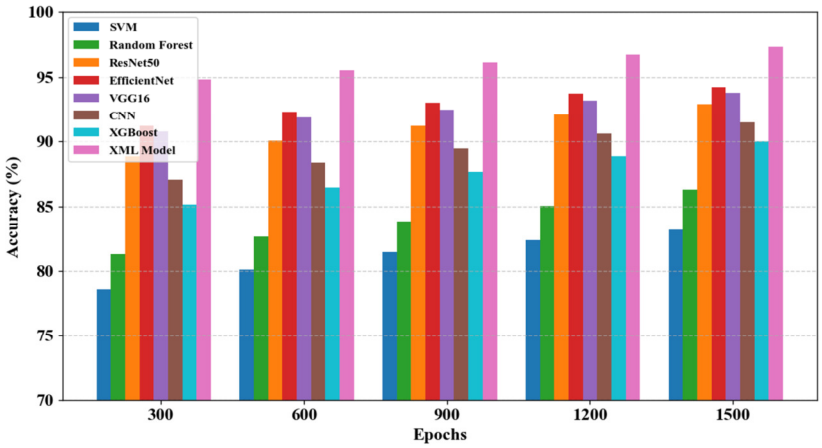

Fig. 2. Comparison of accuracy.

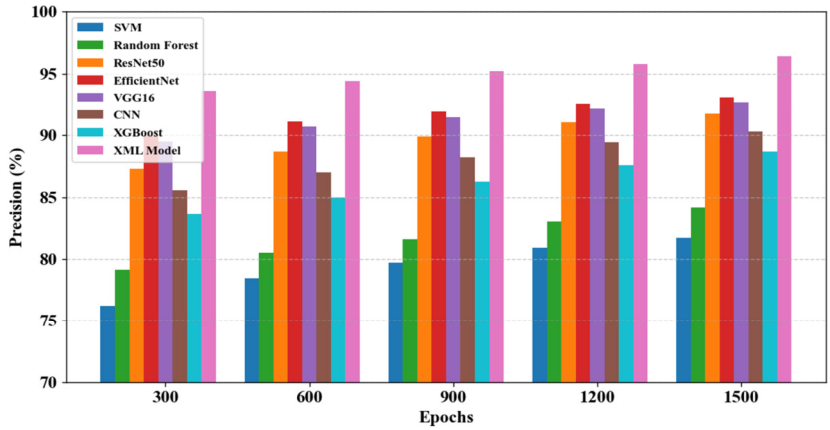

Fig. 3. Comparison of precision.

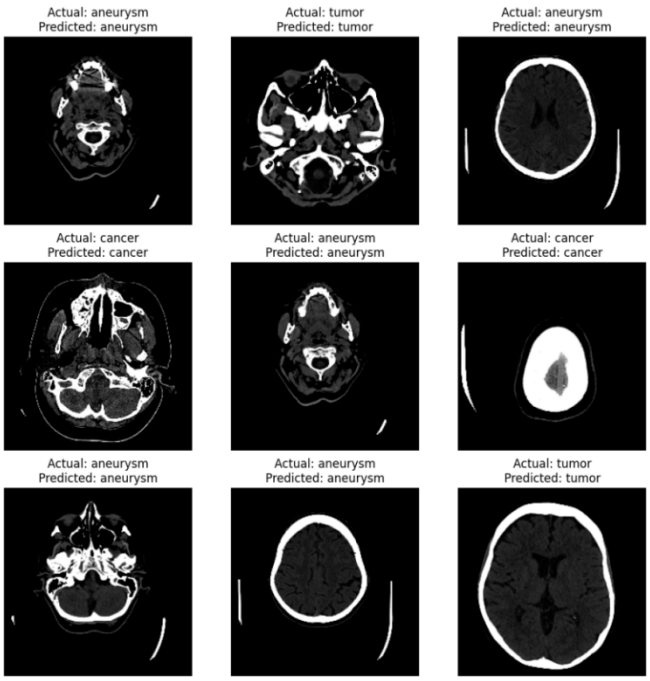

Fig. 4. Brain CT classification images.

CNN and XGBoost improved steadily with epochs, while SVM and RF made moderate gains. From the results shown in Figure 4, it is evident that the proposed XAI model is superior to the other models in handling multimodal BT classification.

IV. CONCLUSION

This study presented a comprehensive performance evaluation of different ML and DL models for multimodal BT classification using MRI and CT images. The proposed XAI model, using multimodal feature fusion with DL optimization, achieved the best performance with 97.3% accuracy, 96.4% precision, 96.0% recall, and 96.2% F1-score after 1500 epochs. The results of the comparison indicate that DL methods were better than traditional ones, such as SVM (83.2%) and RF (86.3%), as models such as VGG16 (93.8%) offered improved performance compared to conventional options. The performance of the proposed XAI model was superior to the models that incorporate only a single modality of data, and such integration of multimodal data leads to robust tumor detection and classification.

Despite promising results, this study has certain limitations. First, the model operates primarily on 2D slices, potentially losing contextual information inherent in complete 3D MRI and CT volumes. Second, the computational cost of

multimodal processing and ensemble classification can be high, making real-time deployment challenging in resource-constrained settings. Third, although SHAP and Grad-CAM provide interpretability, more studies with clinicians are needed to evaluate the trust and usability of these explanations in a real-world clinical workflow. Future research can further examine the use of attention-based transformer models for improved feature learning, incorporate 3D volumetric analysis to expand the database, and, among other things, utilize XAI to enhance interpretability in clinical applications. Optimizing lightweight models can also aid in real-time tumor detection in resource-constrained environments.

REFERENCES

[1] A. A. Asiri, T. A. Soomro, A. A. Shah, G. Pogrebna, M. Irfan, and S. Alqahtani, "Optimized Brain Tumor Detection: A Dual-Module Approach for MRI Image Enhancement and Tumor Classification," *IEEE Access*, vol. 12, pp. 42868–42887, 2024, https://doi.org/10.1109/ACCESS.2024.3379136.

[2] R. Damaševičius, S. K. Jagatheesaperumal, R. N. V. P. S. Kandala, S. Hussain, R. Alizadehsani, and J. M. Gorriz, "Deep learning for personalized health monitoring and prediction: A review," *Computational Intelligence*, vol. 40, no. 3, 2024, Art. no. e12682, https://doi.org/10.1111/coin.12682.

[3] S. Singh, B. Chaudhary, R. Kumar, A. Upadhyay, and S. Kumar, "A Numerical Analysis of Rectangular Open Channel Embedded TiO2-Au-MXene Employed PCF Biosensor for Brain Tumor Diagnosis," *IEEE Sensors Journal*, vol. 24, no. 10, pp. 16047–16054, Feb. 2024, https://doi.org/10.1109/JSEN.2024.3386395.

[4] S. K. Jagatheesaperumal, Q. V. Pham, R. Ruby, Z. Yang, C. Xu, and Z. Zhang, "Explainable AI Over the Internet of Things (IoT): Overview, State-of-the-Art and Future Directions," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 2106–2136, 2022, https://doi.org/10.1109/OJCOMS.2022.3215676.

[5] Y. Ge, L. Xu, X. Wang, Y. Que, and Md. J. Piran, "A Novel Framework for Multimodal Brain Tumor Detection with Scarce Labels," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–14, 2024, https://doi.org/10.1109/JBHI.2024.3467343.

[6] S. Maqsood, R. Damaševičius, and R. Maskeliūnas, "Multi-Modal Brain Tumor Detection Using Deep Neural Network and Multiclass SVM," *Medicina*, vol. 58, no. 8, Aug. 2022, Art. no. 1090, https://doi.org/10.3390/medicina58081090.

[7] R. Anitha and D. Siva Sundhara Raja, "Development of computer-aided approach for brain tumor detection using random forest classifier," *International Journal of Imaging Systems and Technology*, vol. 28, no. 1, pp. 48–53, 2018, https://doi.org/10.1002/ima.22255.

[8] A. K. Sharma *et al.*, "HOG transformation based feature extraction framework in modified Resnet50 model for brain tumor detection," *Biomedical Signal Processing and Control*, vol. 84, Jul. 2023, Art. no. 104737, https://doi.org/10.1016/j.bspc.2023.104737.

[9] T. Zhao, B. L. Faustino, S. K. Jagatheesaperumal, F. de P. S. Rolim, and V. H. C. de Albuquerque, "Reinforcement learning-based adaptive deep brain stimulation computational model for the treatment of tremor in Parkinson's disease," *Expert Systems with Applications*, vol. 267, Apr. 2025, Art. no. 126154, https://doi.org/10.1016/j.eswa.2024.126154.

[10] K. Kamala Devi and J. Raja Sekar, "Optimizing feature selection and parameter tuning for breast cancer detection using hybrid GAHBA-DNN framework," *Journal of Intelligent & Fuzzy Systems*, vol. 46, no. 4, pp. 8037–8048, Apr. 2024, https://doi.org/10.3233/JIFS-236577.

[11] A. Guennich, M. Othmani, and H. Ltifi, "Advanced Brain Tumor Segmentation With a Multiscale CNN and Conditional Random Fields," *IEEE Access*, vol. 13, pp. 34925–34935, 2025, https://doi.org/10.1109/ACCESS.2025.3541183.

[12] C. J. Tseng and C. Tang, "An optimized XGBoost technique for accurate brain tumor detection using feature selection and image segmentation," *Healthcare Analytics*, vol. 4, Dec. 2023, Art. no. 100217, https://doi.org/10.1016/j.health.2023.100217.

[13] N. Pedano *et al.*, "The Cancer Genome Atlas Low Grade Glioma Collection (TCGA-LGG)." The Cancer Imaging Archive, 2016, https://doi.org/10.7937/K9/TCIA.2016.L4LTD3TK.

[14] A. H. Hamada, "Br35H :: Brain Tumor Detection 2020." IEEE DataPort, https://doi.org/10.21227/TBKK-Q937.

[15] National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC), "The Clinical Proteomic Tumor Analysis Consortium Glioblastoma Multiforme Collection (CPTAC-GBM)." The Cancer Imaging Archive, May 10, 2023, https://doi.org/10.7937/K9/TCIA.2018.3RJE41Q1.

[16] "CPTAC Imaging and Annotation Data." https://kaggle.com/code/justinkirby/cptac-imaging-and-annotation-data.

[17] G. Naidu, T. Zuva, and E. M. Sibanda, "A Review of Evaluation Metrics in Machine Learning Algorithms," in *Artificial Intelligence Application in Networks and Systems*, 2023, pp. 15–25, https://doi.org/10.1007/978-3-031-35314-7_2.