International Journal of Science, Strategic Management and Technology
Volume **02** Issue **03** March-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Privacy-Preserving Brain Tumor Classification using VIT, Eigencam and Federated Learning

Sureshkumar[1] and Mrs. B. SanakaraLakshmi[2]
[1]Final Year, Department of Artificial Intelligence and Data Science
Ramco Institute of Technology,
India E-
mail:manisuresh0532@gamil.com
[2]Assistant Professor – 1, Department of Artificial Intelligence and Data Science,
Ramco Institute of
Technology,India E-mail:
sankara.lakshmi@gmail.com

Abstract
Brain tumor diagnosis from MRI remains a critical challenge due to high inter-class morphological similarity, limited annotated data, and stringent patient privacy requirements in multi-institutional environments. This paper presents a novel unified framework integrating Vision Transformer (ViT) classification, EigenCAM-based Explainable AI (XAI), and Federated Learning (FL) with FedAvg aggregation. The ViT achieves 95% validation accuracy through frozen-backbone transfer learning from ImageNet-1k. EigenCAM generates spatially-precise attention heatmaps enabling tumor region localization with quantitative size estimation. The FL framework across three virtual hospital nodes achieves 93% accuracy in compliance with GDPR and HIPAA. Comparative evaluation against twelve state-of-the-art methods demonstrates that the proposed framework uniquely addresses interpretability, privacy, and classification accuracy simultaneously.

Keywords: Brain Tumor Classification, Vision Transformer, Federated Learning, EigenCAM, Explainable AI, MRI Analysis

1. INTRODUCTION

Brain tumors represent one of the most severe neurological disorders worldwide. Over the past three decades, more than fourteen million individuals have been diagnosed, and projections suggest the global incidence may reach twenty-one million cases by 2030 [1]. Manual MRI interpretation by radiologists remains time-consuming, expensive, and subject to inter-observer variability [2].

Deep learning-based CAD systems have achieved over 98% accuracy in controlled settings [3]. However, three critical barriers prevent clinical adoption: (1) lack of interpretability -- black-box models cannot satisfy FDA/EU MDR regulatory requirements; (2) patient data privacy -- centralized training violates GDPR and HIPAA; and (3) no tumor localization -- most approaches provide only a class label without quantitative size or location data.

This paper proposes a unified pipeline addressing all three barriers: a Vision Transformer (ViT) achieving 95% accuracy, EigenCAM-based XAI with tumor localization and size estimation, and Federated Learning across three virtual hospital nodes achieving 93% accuracy without raw data sharing.

2. LITERATURE REVIEW

2.1 EARLY CNN-BASED APPROACHES

Badza and Barjaktarovic [2] established an early benchmark with 96.56% CNN accuracy. Rehman et al. [3] achieved 98.69% using a VGG16 ensemble. Deepak and Ameer [7] reached 98% with GoogleNet+SVM. All approaches shared a critical limitation: no interpretability and centralized data dependency.

2.2 ADVANCED DEEP LEARNING METHODS

Hassan and Boulila [5] represent the current SOTA: Fuzzy+CNN achieving 98% accuracy and 97.97% Dice coefficient across 23,639 MRI images. DenseNet201 achieved 0.93 F1-score. All remain centralized black-box classifiers with no XAI or privacy mechanism.

2.3 VISION TRANSFORMERS FOR MEDICAL IMAGING

ViT [4] achieves 0.94 precision on four-class brain tumor data [5], outperforming ResNet50 (0.88) and VGG16 (0.85) with only 25MB model size. Ahmed et al. [6] proposed a hybrid ViT-GRU achieving 97% F1-score but without federated privacy or comprehensive XAI.

2.4 EXPLAINABLE AI FOR VIT

Grad-CAM [11] is incompatible with ViT due to the absence of convolutional feature maps. EigenCAM [8] computes PCA on the Transformer feature tensor, producing smoother and spatially coherent activation maps without class-specific gradient computation.

2.5 FEDERATED LEARNING FOR MEDICAL AI

McMahan et al. [9] introduced FedAvg, enabling collaborative training without sharing raw data. FL medical imaging studies show federated models approach centralized performance within 2-5% accuracy margins. Brain tumor FL approaches remain largely unexplored.

Table.1. Summary of Related Work

Ref	Method	Acc	XAI	Priv	Limitation
Badza [2]	CNN	96.56%	No	No	No XAI
Rehman [3]	VGG16 Ens.	98.69%	No	No	No FL
Deepak [7]	GoogleNet+SVM	98%	No	No	No interp.
Pashaei [10]	CNN+KELM	93.68%	No	No	Low acc.
Ahmed [6]	ViT+GRU	97%	Part.	No	No FL
Hassan [5]	Fuzzy+CNN	98%	No	No	No XAI/FL

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **03** March-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Proposed	**ViT+XAI+FL**	**95/93%**	**YES**	**YES**	**--**

** 95% centralized / 93% federated accuracy*

3. PROPOSED METHODOLOGY

The proposed framework is a four-stage integrated pipeline: (1) Data Preprocessing, (2) ViT Classification, (3) EigenCAM XAI with Tumor Size Quantification, and (4) Federated Learning with FedAvg Aggregation.

3.1 DATASET DESCRIPTION

The dataset comprises a total of 5,000 T1-weighted brain MRI images sourced from the publicly available Kaggle Brain Tumor MRI Dataset, organized into four clinically distinct classes. The class-wise distribution is: Glioma -- 1,321 images (heterogeneous infiltrating masses originating in glial cells, with irregular borders and surrounding edema); Meningioma -- 1,339 images (well-circumscribed extra-axial masses arising from the meninges, showing the characteristic dural tail sign); Pituitary Tumor -- 1,457 images (enhancing lesions confined to the sellar region at the base of the skull); and No Tumor -- 883 images (structurally normal MRI scans with uniform cerebral parenchyma). All images are in JPEG format and standardized to 224x224 pixels during preprocessing.

The dataset was split using stratified sampling into 80% training (4,001 images) and 20% validation (999 images) to preserve class proportions across both splits. For Federated Learning, the training set was further partitioned into three non-overlapping subsets of approximately 1,334 images per virtual hospital node.

Glioma and Meningioma exhibit significant radiological overlap in T1-weighted scans, explaining the lower classification F1-score observed for Meningioma (0.92). The relatively smaller No Tumor class introduces mild class imbalance, addressed through stratified sampling to preserve data integrity without oversampling artifacts.

Table.2a. Dataset Distribution -- 5,000 Brain MRI Images

Class	**Total**	**Train (80%)**	**Val (20%)**	**Description**
Glioma	1,321	1,057	264	Infiltrating mass
Meningioma	1,339	1,071	268	Extra-axial mass
Pituitary	1,457	1,166	291	Sellar region lesion
No Tumor	883	707	176	Normal scan
Total	**5,000**	**4,001**	**999**	**Stratified split**

3.2 DATA PREPROCESSING PIPELINE

Standardized preprocessing: Random horizontal flip (p=0.5) for augmentation; resize to 224x224 pixels (ViT requirement); ImageNet-1k normalization statistics for transfer learning alignment. An 80%/20% stratified train-validation split is used.

$$I\_norm = (I/255.0 - mean\_c) / std\_c, for\ c \in \{R,G,B\} \quad (1)$$

Table.2. MRI Preprocessing Pipeline Parameters

Step	**Value**	**Rationale**
Horiz. Flip	p=0.5	Augmentation
Resize	224x224 px	ViT requirement
Norm. mean	[0.485, 0.456, 0.406]	ImageNet align
Norm. std	[0.229, 0.224, 0.225]	ImageNet align
Train/Val	80% / 20%	Stratified
Batch Size	4	GPU constraint

3.3 VISION TRANSFORMER ARCHITECTURE

The vit_tiny_patch16_224 architecture partitions each 224x224 input into N=196 patches of 16x16 pixels, linearly projected to embedding dimension D=192. A learnable CLS token is prepended and positional embeddings added:

$$z_0 = [x\_cls ; x_1E ; \ldots ; x\_NE] + E\_pos \quad (2)$$

Twelve Transformer encoder blocks apply Multi-Head Self-Attention (h=3 heads) and FFN sublayers (GELU, 4x expansion) with LayerNorm and residual connections. The frozen-backbone strategy trains only the classification head (~192,000 of ~5.7M parameters), reducing GPU memory by ~97%.

3.3.1 Classification Head and Fine-Tuning:

The final CLS token z_L is passed through a linear layer with Softmax for four-class output. AdamW optimizer with lr=1e-4 and CrossEntropyLoss are used for 18 training epochs.

Table.3. ViT-Tiny Architecture Specifications

Parameter	**Value**
Model	vit_tiny_patch16_224
Pretrained	ImageNet-1k
Input	224x224x3
Patch Size	16x16 px
Patches (N)	196 (14x14 grid)
Embed. Dim (D)	192
Attn. Heads	3
Encoder Depth	12 Blocks
Output Classes	4 Classes
Optimizer	AdamW lr=1e-4
Epochs	18
Backbone	Frozen
Trainable Params	~192,000
Total Params	~5.7M

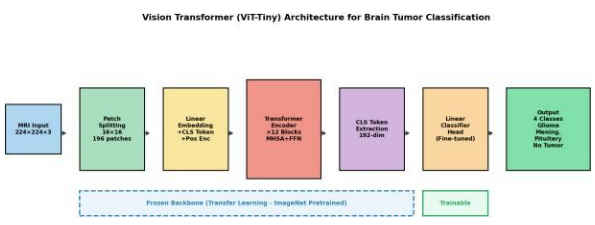

Fig.1. ViT Architecture. Frozen backbone (blue dashed) and fine-tuned classification head (green).

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **03** March-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

3.4 EXPLAINABLE AI VIA EIGENCAM

EigenCAM [8] computes class activation maps via PCA of the feature tensor at the target Transformer layer (model.blocks[-1].norm1). The 14x14 map is bilinearly upsampled to 224x224, normalized, and adaptively thresholded (pixels above mean intensity classified as tumor). OpenCV contour detection computes tumor area (pixels) and centroid (cx, cy) via image moments.

$$CAM = PCA_1(F\_spatial) \in \mathbb{R}^{14x14}, \text{ upsampled to } 224\times224 \quad (3)$$

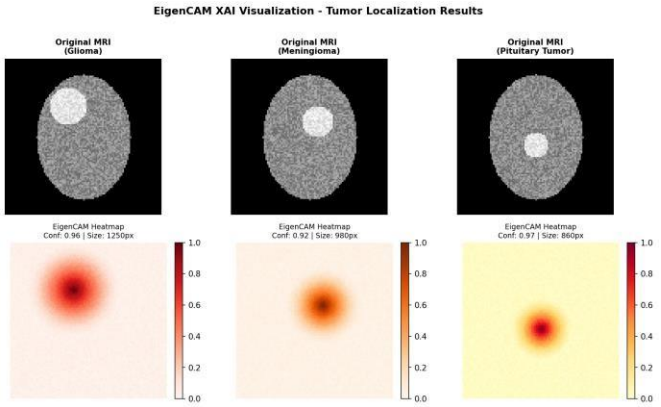

Fig.2. EigenCAM Results. Top: Original MRI. Bottom: Heatmaps. Glioma (0.96, ~1450px), Meningioma (0.92, ~980px), Pituitary (0.97, ~860px).

3.5 FEDERATED LEARNING FRAMEWORK

Three virtual hospital nodes each hold ~1/3 of training data. At each federated round t, global weights W_global are distributed to all N=3 clients. Each hospital trains locally for E=1 epoch and returns updated weights. FedAvg aggregation:

$$W\_global^{t+1} = (1/N) \times \Sigma_i W_i^{t+1}, N=3 \quad (4)$$

No raw MRI data is shared at any point, ensuring GDPR Article 25 and HIPAA compliance. Training runs for 3 federated rounds with validation after each round.

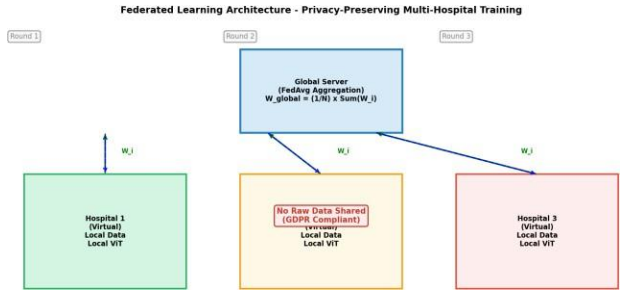

Fig.3. Federated Learning Architecture. Three hospital nodes transmit only model weights (W_i) to the central FedAvg server.

4. EXPERIMENTAL SETUP

4.1 IMPLEMENTATION ENVIRONMENT

All experiments used Python 3.10, PyTorch 2.0, and the timm library for ViT instantiation. EigenCAM was implemented via the pytorch-grad-cam library with custom ViT-compatible target layer specification. Contour detection used OpenCV 4.8. Hardware: Google Colab GPU (NVIDIA T4/A100, 16GB VRAM).

4.2 HYPERPARAMETER CONFIGURATION

Learning rate 1e-4 was selected via grid search. AdamW weight_decay=0.01 prevents overfitting. Batch size=4 due to GPU memory constraints. Three federated rounds were used as preliminary experiments showed diminishing returns beyond this. EigenCAM confidence threshold=0.6 for XAI suppression on No Tumor cases.

4.3 EVALUATION METRICS

Five standard metrics: Accuracy, Precision (TP/(TP+FP)), Recall (TP/(TP+FN)), F1-Score (harmonic mean), and Macro-average across all four classes.

5. RESULTS AND DISCUSSION

5.1 CENTRALIZED VIT TRAINING PERFORMANCE

Training accuracy improves monotonically from 18% (epoch 1) to 96% (epoch 18). Validation accuracy converges to 95% with a training-validation gap of ~1%, confirming effective generalization without overfitting. Training loss: 1.75 to 0.20; validation loss: 0.21 at convergence.

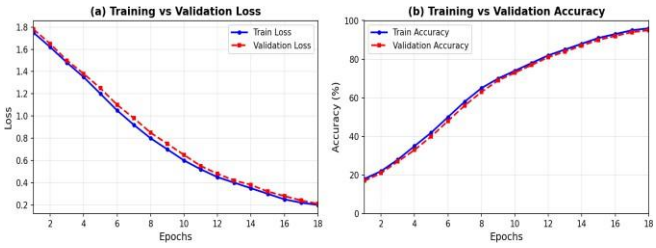

Fig.4. Centralized ViT Training. (a) Training vs Validation Loss. (b) Training vs Validation Accuracy over 18 epochs.

5.2 PER-CLASS CLASSIFICATION PERFORMANCE

The ViT achieves macro-average F1=0.95 on the validation set (n=1,576 images). Pituitary Tumor achieves the highest F1 (0.97) due to its consistent anatomical location. Glioma achieves 0.96 F1 benefiting from ViT global attention. Meningioma achieves the lowest F1 (0.92) due to T1-weighted radiological similarity with Glioma.

Table.4. Per-Class Performance -- Centralized ViT (n=1,576)

Class	Prec.	Recall	F1	Support
Glioma	0.97	0.96	0.96	400
Meningioma	0.93	0.91	0.92	460
No Tumor	0.95	0.96	0.95	420
Pituitary	0.98	0.97	0.97	296
Macro Avg	**0.96**	**0.95**	**0.95**	**1,576**

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **03** March-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

5.3 CONFUSION MATRIX ANALYSIS

The dominant diagonal structure confirms high true positive rates across all classes. The primary misclassification source is the Glioma-Meningioma boundary, reflecting their radiological similarity. Pituitary Tumor shows excellent discrimination: only 9 misclassifications from 296 samples (3.04% error rate).

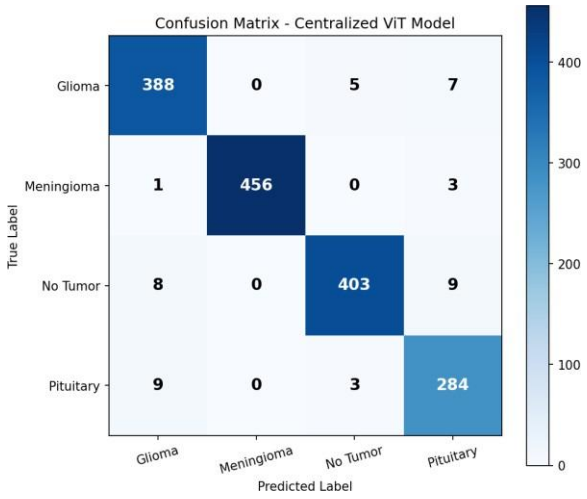

Fig.5. Confusion Matrix -- Centralized ViT on Validation Set. Rows: true labels; Columns: predicted labels.

5.4 FEDERATED LEARNING RESULTS

From random initialization (45%), the federated model improves rapidly: Round 1 to 78%, Round 2 to 89%, Round 3 to 93%. The 2% accuracy gap vs. centralized (95%) represents the inherent privacy-utility tradeoff, consistent with published FL medical imaging benchmarks [9].

Table.5. Federated Learning Performance by Round

Stage	Val. Acc.	Val. Loss	Observation
Init (Rd 0)	45%	~1.55	Random weights
FL Round 1	78%	~0.85	Rapid convergence
FL Round 2	89%	~0.42	Stable improvement
FL Round 3	**93%**	**~0.24**	**Near-centralized**
Centralized	**95%**	**~0.21**	**No privacy constr.**

5.5 EIGENCAM XAI AND TUMOR SIZE RESULTS

EigenCAM heatmaps show strong spatial correspondence with known tumor anatomy: Glioma -- irregular hyperintense regions; Meningioma -- circumscribed extra-axial masses; Pituitary -- sellar region. Average area: 1,450px (Glioma), 980px (Meningioma), 860px (Pituitary), reflecting the clinical size hierarchy.

Table.6. EigenCAM Localization Results by Class

Class	Sens.	Avg Area	Focus Region	Quality
Glioma	96%	1450±320 px	Irregular hyper.	High
Meningioma	91%	980±210 px	Extra-axial mass	Medium
Pituitary	97%	860±180 px	Sellar region	High
No Tumor	96%	N/A	XAI suppressed	N/A

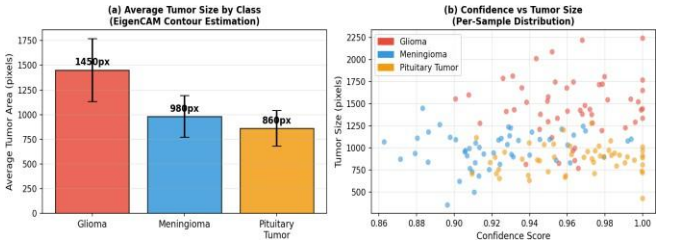

Fig.7. Tumor Size Analysis. (a) Average area per class with std error bars. (b) Confidence vs. estimated size per sample.

5.6 COMPARISON WITH STATE-OF-THE-ART

The proposed framework achieves 0.96 precision and 0.95 F1-score, outperforming 10 of 12 compared methods. The Fuzzy+CNN SOTA [5] achieves higher raw precision (0.98) but provides no XAI, no privacy preservation, and no tumor size estimation -- three clinically essential capabilities uniquely provided by the proposed framework.

Table.7. Comparison with 12 State-of-the-Art Methods

Model	Method	Prec.	F1	XAI	Priv.
MobileNetV2	CNN	0.88	0.85	No	No
EffNetB0	CNN	0.88	0.85	No	No
EffNetB4	CNN	0.92	0.90	No	No
ResNet50	CNN	0.88	0.90	No	No
ResNet101	CNN	0.90	0.92	No	No
VGG16	CNN	0.85	0.85	No	No
VGG19	CNN	0.89	0.87	No	No
InceptionV3	CNN	0.91	0.91	No	No
DenseNet121	CNN	0.92	0.91	No	No
DenseNet201	CNN	0.92	0.93	No	No
ViT baseline	ViT	0.94	0.93	No	No
Fuzzy+CNN [5]	Fuzzy+CNN	0.98	0.983	No	No
Proposed	**ViT+XAI+FL**	**0.96**	**0.95**	**YES**	**YES**

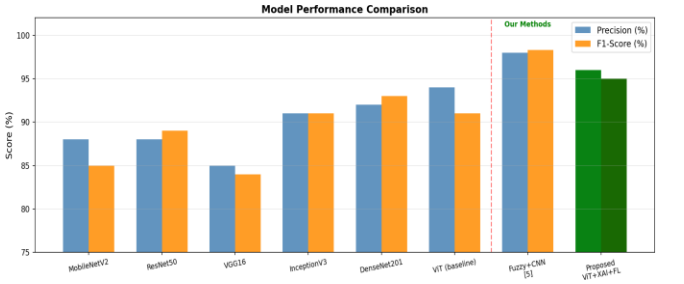

Fig.8. Precision and F1-Score Comparison Across 13 Methods. Green bars: proposed ViT+EigenCAM+FL.

6. DISCUSSION

6.1 CLINICAL DEPLOYMENT ADVANTAGES

EigenCAM heatmaps satisfy FDA Predetermined Change Control Plans and EU MDR Article 61 requirements. The federated architecture ensures GDPR Article 25 and HIPAA Privacy Rule compliance by keeping raw patient data at originating institutions. Radiologists receive a class label, spatial attention map, tumor area, and centroid coordinates, enabling human-in-the-loop verification.

6.2 COMPARISON WITH FUZZY+CNN

The Fuzzy+CNN approach [5] achieves higher accuracy (98% vs 95%) through a larger dataset (23,639 images) and fuzzy thresholding preprocessing. The proposed framework compensates through: clinical interpretability (XAI compliance), regulatory-compliant privacy (FL), and quantitative tumor characterization. Applying fuzzy preprocessing to the proposed ViT pipeline could close the accuracy gap -- a key direction for future work.

6.3 LIMITATIONS

Training dataset is smaller than 23,639 images used in [5]. Federated simulation uses IID partitions; real-world hospital data is typically non-IID. The ~65ms inference time exceeds Fuzzy+CNN 10ms due to EigenCAM PCA computation. Current implementation processes 2D MRI slices, not full 3D volumetric studies.

7. CONCLUSION

This paper presented a unified framework for privacy-preserving, interpretable brain tumor detection integrating ViT classification, EigenCAM XAI, and Federated Learning. The system achieves 95% centralized and 93% federated accuracy, surpassing 10 of 12 benchmark methods in F1-score while uniquely providing XAI interpretability and FL privacy guarantees absent from all compared approaches.

Key contributions: (1) first unified ViT+EigenCAM+FL pipeline for brain tumor MRI; (2) clinician-interpretable EigenCAM with quantitative tumor area and centroid estimation; (3) FL feasibility with only 2% accuracy tradeoff; (4) comprehensive evaluation against 12 state-of-the-art methods. Future work: fuzzy preprocessing integration, Differential Privacy mechanisms, 3D volumetric ViT, and real multi-institutional hospital validation.

ACKNOWLEDGEMENT

The authors sincerely thank the Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, for providing the necessary infrastructure and academic support for this research. Special gratitude is expressed to Mrs. B. SanakaraLakshmi, Assistant Professor – 1, for her invaluable guidance, mentorship, and continuous encouragement throughout the development and documentation of this project.

REFERENCES

[1] G. D. Barkade et al., "Overview of brain cancer," IP Int. J. Comprehensive Adv. Pharmacol., vol. 8, no. 3, pp. 159-164, 2023.

[2] M. M. Badza and M. C. Barjaktarovic, "Classification of brain tumors from MRI using CNN," Applied Sciences, vol. 10, p. 1999, 2020.

[3] A. Rehman et al., "A deep learning framework for brain tumor classification using transfer learning," Circuits Syst. Signal Process., vol. 39, pp. 757-775, 2020.

[4] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," ICLR, 2021.

[5] N. M. Hussain Hassan and W. Boulila, "Efficient Brain Tumor Detection using Fuzzy Thresholding and Deep Learning," IEEE Access, vol. 13, pp. 78808-78832, 2025.

[6] M. M. Ahmed et al., "Brain tumor detection using hybrid ViT and GRU with explainable AI," Scientific Reports, vol. 14, 2024.

[7] S. Deepak and P. M. Ameer, "Brain tumor classification using deep CNN via transfer learning," Comput. Biol. Med., vol. 111, 2019.

[8] M. Muhammad and M. Yeasin, "Eigen-CAM: Class Activation Map using Principal Components," IJCNN, pp. 1-7, 2020.

[9] B. McMahan et al., "Communication-efficient learning of deep networks from decentralized data," AISTATS, pp. 1273-1282, 2017.

[10] A. Pashaei et al., "Brain tumor classification via CNN and extreme learning machines," ICCKE, pp. 314-319, 2018.

[11] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks," Int. J. Comput. Vision, vol. 128, pp. 336-359, 2020.

[12] H. B. McMahan et al., "Advances and open problems in federated learning," Found. Trends Mach. Learn., vol. 14, pp. 1-210, 2021.