2025 International Conference on Intelligent Systems and Computational Networks (ICISCN) | 979-8-3315-2924-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICISCN64258.2025.10934555

Deep Learning Framework with Explainable AI for Accurate and Interpretable Brain Tumor Segmentation

Dushyantha ND
CSE. HKBK College of Engineering
HKBK College of Engineering
Bangalore, India
dushyantha.cs@hkbk.edu.in

Kavya R Naik
CSE. HKBK College of Engineering
HKBK College of Engineering
Bangalore, India
kavyaravinaik2003@gmail.com

Chaithanya H G
CSE. HKBK College of Engineering
HKBK College of Engineering
Bangalore, India
chaithanyareddy364@gmail.com

Amrutha L
CSE. HKBK College of Engineering
HKBK College of Engineering
Bangalore, India
ammu92067@gmail.com

Dinesh C
CSE. HKBK College of Engineering
HKBK College of Engineering
Bangalore, India
dinesh.creddy143@gmail.com

***Abstract*—Brain tumor detection by MRI scan is an imperative medical concern that necessitates state-of-the-art techniques for accurate and timely detection. This paper proposes an approach integrating Explainable AI, Federated Learning, and deep learning to improve accuracy, privacy, and trust. Federated learning allows collaborative learning for the model without accessing each other's data. Explanation using SHAP and LIME techniques results in interpretable predictions from an AI system, enabling greater clinician trust in these AI systems. In accuracy, the proposed system could achieve an accuracy of 96.8%, with sensitivity at 96.5%, specificity at 96.4%, and an F1-score at 96.6% more than the traditional CNN model. The future direction in this approach will be enhanced in interpretability, robustness, scalability, and in dataset diversity to further boost generalization. This research methodology is bridging AI research to clinical application that offers reliable diagnoses through practices of privacy preservation, moves healthcare through innovation, and evokes trust in AI-driven solutions.**

***Index Terms*—MRI-based brain tumor diagnosis, Explainable AI (XAI), Federated Learning (FL), Deep learning, Data privacy, Interpretable predictions, Diagnostic reliability, Clinical applications.**

I. Introduction

Despite advances in technology, brain tumors remain the most common cause of morbidity and mortality worldwide, and early diagnosis is key to effective treatment. MRI-based non-invasive diagnosis has become the cornerstone, but manual interpretation is labor-intensive and error-prone, thus delaying treatment. Recent advancements in AI, especially deep learning models including CNNs, have revamped automated brain tumor detection. However, despite all these advancements, much remains to be addressed, such as data privacy concerns with AI, the black-box nature of the current AI models, and the demand for real-time clinical processing [1]. The paper addresses these challenges using a unified approach that integrates FL for privacy, XAI for transparency, and deep learning for precision diagnosis.

This work explores an integrated approach combining Federated Learning(FL) for data privacy[2], Explainable AI(XAI) for model transparency, and deep learning for precise tumour diagnosis. The goal of this project is to provide a clinically useful, interpretable, and privacy-preserving diagnostic system. The contributions of this paper are the following.

- Federated Learning for privacy-preserving cooperative model training.
- Explainable AI for transparent models using SHAP and LIME[3].
- Current methods will be analyzed, with its shortcomings and potential directions toward future research.

Deep learning, specifically CNNs, has been widely applied in recent research for brain tumor diagnosis from MRI scans. Their ”black-box” nature poses significant challenges even though they achieve high accuracy, especially in clinical scenarios where interpretability is a concern[4]. Federated Learning, which promises to offer privacy-friendly co-operative model training across institutions without sharing raw data, is emerging as a possible solution to these problems[5]. However, the problems of data heterogeneity and communication overhead still persist as validated by studies.

At the same time, XAI techniques like SHAP and LIME are improving model usability and trust while providing critical information on what influences predictions[6]. Recent studies have emphasized the importance of model interpretability, especially in clinical settings where transparency is the basis of clinical decision-making. In the proposed model, combining

979-8-3315-2924-6/25/$31.00 ©2025 IEEE

FL with XAI to create a model that balances privacy, interpretability, and diagnostic accuracy best for the improvement of validity and reliability in AI-driven systems of brain tumor detection. Addressing data heterogeneity and providing real-time complex explanations may help in creating adaptive and reliable AI models in real-world clinical settings[7].

II. Scope and Aims of the Current Review

A. Scope

This paper deals with the problem of brain tumour diagnosis using a hybrid Machine Learning system based on MRI data, using FL and XAI. It will assess the contribution of these approaches to privacy, interpretability, and prediction in diagnostic applications relative to current approaches, as well as the computational and implementation challenges of translation to the real-time clinical environment[8].

B. Purpose

- **Performance Metric:** In the XAI framework, interpretability and diagnostic accuracy are the assessment criteria, and the system efficiency is the assessment criteria for FL.
- **Advantages and Disadvantages:** The two main adverse features of XAI and FL are the type of transparency, privacy, specifically computational cost, and the heterogeneity of the data.
- **Federated Learning Challenges:** There is also the issue of the computational cost and the real-time responsiveness of the specificity of the explanations [9]. This, in turn, gives rise to the idea of reducing FL communication overhead and thus facilitates the increase in the efficiency with which the model itself aggregates, notably when processing a large data set [10] [8].
- **Generalization Improvements:** Some of the ways to overcome them include:
  - Enhance the diversity and stability of the data set to do well for other MRI scans.
  - Real-time Learning: Provide planning as an option for obtaining speed and accuracy of the predictive model within the framework of a stable clinical setting.
  - Powerful XAI Methods: Investigate other options for stronger feature-specific or approximate explanation generation that reduce the cost of time.
- **Significance of the Approach:** This is the reason the integrated approach makes it possible to reprieve in trusting the AI systems, getting over the gap of theoretical AI and practical usage and arriving at much more accurate diagnostic solutions and at the same time cheaper to the clinicians [11] .

III. Proposed Method

In the proposed method, the federated learning (FL), explainable artificial intelligence (XAI), and deep learning are used to provide an integrated approach for brain tumour diagnosis. The three primary components of the system design are a CNN model for tumour classification, XAI for interpretability, and FL for private model training[12].

Federated Learning: FL allows the training of collaborative models across institutions without having to share raw MRI data, thus preserving privacy. The Federated Averaging Algorithm aggregates local model updates on a central server while maintaining confidentiality of individual data[13].

Federated Averaging Algorithm:

The Federated Averaging (FedAvg) algorithm from (1) updates the global model parameters by aggregating the locally computed updates from participating devices. The update rule is expressed as follows:

$$w_{t+1} = \frac{1}{n} \sum_{i=1}^{n} w_t^i \tag{1}$$

where w_{t+1} represents the updated global model parameters at time $t+1$, n is the total number of participating devices, and w_t^i denotes the local model parameters from the i-th device at time t. This approach, introduced by McMahan et al. (2017), is a cornerstone of federated learning [14].

Algorithm 1 Federated Learning

```
Input: Number of clients $n$, number of rounds $R$, dataset $D$
Output: Global model $M$
Initialize global model $M$ with random weights
Split $D$ into $n$ partitions $D_1, ..., D_n$
for $r = 1$ to $R$ do
  for $i = 1$ to $n$ do
    Send $M$ to client $i$
    Client $i$ trains a local model $M_i$ on $D_i$
    Client $i$ sends updated weights $\Delta M_i$ to the server
  end for
  Server aggregates updates: $M \leftarrow \frac{1}{n}\sum_{i=1}^{n} M_i$
end for
return $M$
```

The Federated Learning framework, which is described in **algorithm 1**, iteratively updates a global model M by working with several clients. The model is given to n clients at each round r, and each of them trains it locally on its partitioned dataset D_i. The server then aggregates the local modifications ΔM_i to create the new global model. Until the global model M converges, this method is repeated for R rounds.

Explainable AI: Post-training, SHAP and LIME are applied to decode the decision-making process of the CNN model. LIME provides localised explanations for specific MRI images, whereas SHAP values emphasize the importance of individual features, such as specific cancer locations, to the prediction[15].

Algorithm 2 Grad-CAM

```
Input: Model $M$, input image $x$, target class $c$
Output: Grad-CAM heatmap $H$
Forward pass:
Compute $y = M(x)$
Extract feature maps $A \in \mathbb{R}^{C \times H \times W}$ from the last convolutional layer
Backward pass:
Compute gradients $\frac{\partial y_c}{\partial A}$
Weight feature maps:
For each feature map $A_i$:
Compute weight $w_i = \frac{1}{H \cdot W} \sum_{j,k} \frac{\partial y_c}{\partial A_{i,j,k}}$
Compute weighted sum $H = \sum_{i=1}^{C} w_i A_i$
Normalize $H$ to $[0, 1]$
return $H$
```

The Grad-CAM algorithm stated in **Algorithm 2** helps visualize which regions in an input image x have the most influence on the model's prediction for a particular class c. The process works in the following steps:

- **Forward Pass:** The model first makes a prediction $y = M(x)$ based on the input image and extracts the feature maps A from the last convolutional layer.
- **Backward Pass:** Then, the algorithm calculates the gradients $\frac{\partial y_c}{\partial A}$, which tell us how much each feature map contributes to the prediction for the target class c.
- **Weighting:** For each feature map, a weight w_i is computed, which is the average of the gradients across all spatial locations in the feature map.
- **Heatmap Generation:** The weighted feature maps are combined to form a heatmap H, which highlights the areas in the image most relevant to the class prediction.

Finally, the heatmap is normalized to the range $[0, 1]$ to make it easier to interpret and visualize.

Model Evaluation: Well-known criteria such as accuracy, sensitivity, specificity, and F1 score are used to evaluate the proposed model. **Figure 1** depicts the architecture of the proposed Explainable AI and Federated Learning-based brain tumour detection system. Federated Learning preserves patient privacy by allowing medical institutions to collectively work on model training without having to share their private MRI data[16] [17].

In this architectural design:

- **Information Gathering:** Information is collected and stored locally at each participating site from MRI scans. FL allows organizations to train models locally on the organization's data, and no data is exchanged during the training procedure[18].
- **Aggregation:** After local training is completed, updates of models are collected on a central server so that the global model can be improved without exposing raw data [19] [20].
- **Explainable AI (XAI):** After the training of the model, predictions given by the deep learning model are decoded using methods such as SHAP and LIME. By clarifying the logic behind a specific prediction, such as the presence of a tumor, these methods increase the dependability and accessibility of the system for doctors[10].

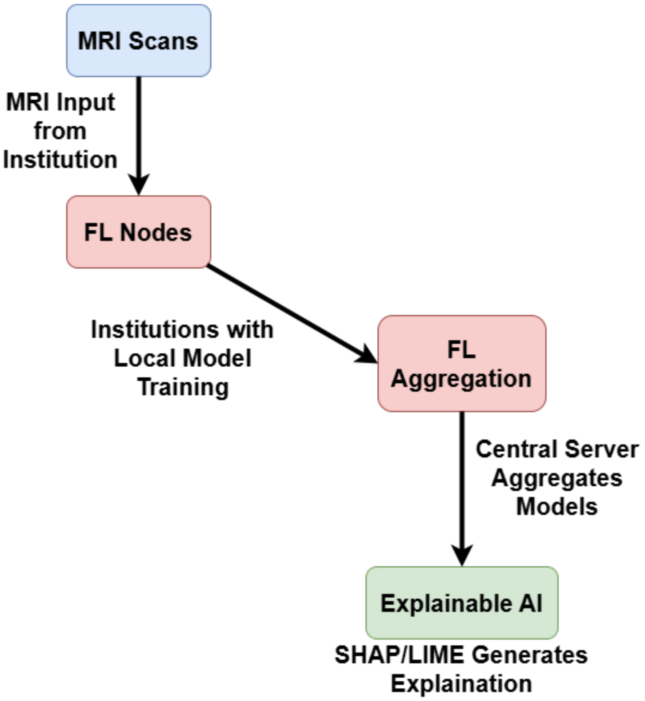
MRI Scans
MRI Input from Institution
FL Nodes
Institutions with Local Model Training
FL Aggregation
Central Server Aggregates Models
Explainable AI
SHAP/LIME Generates Explaination

Fig. 1. Beyond the Black Box: Federated XAI for Trustworthy Medical Imaging

With perfect accuracy and interpretability of the results used to determine its performance, the clarity in decision-making and a precise diagnosis of the brain tumor are guaranteed. Local training and aggregation through generation of explainable predictions in a collaborative setting with privacy, this architecture diagram helps the reader understand the workflow of processing MRI scan data [19].

Figure 2 shows a flowchart describing the sequential steps of the brain tumour detection pipeline. This flowchart starts from the data collection phase, where MRI scans are collected from different institutions. During the local model training stage, which uses FL to train machine learning models inside each institution, the privacy of individual datasets is maintained. This is further improved using model aggregation on a central server to generate the global model. After the training of the model, XAI methods are applied to the model in order to provide explanations for its predictions. The output is forwarded for clinical assessment. The data processing and process flow of the system are easily comprehensible from this flowchart.

IV. Result and Discussion

Evaluated Result of the proposed system using a dataset of MRI scans from figshare Brain Tumor dataset. The system achieved the following performance metrics:

- **Accuracy:** 96.8%
- **Sensitivity:** 96.5%

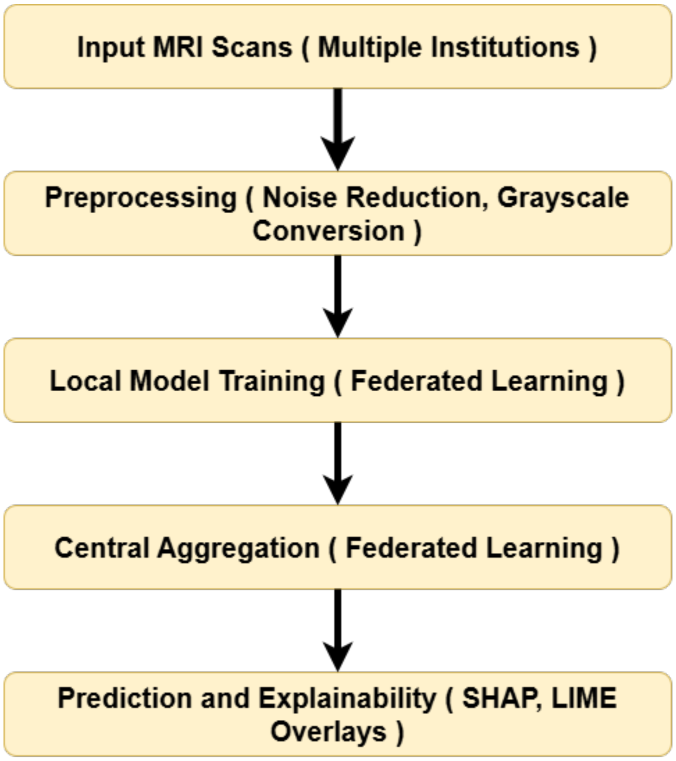
Input MRI Scans (Multiple Institutions)
Preprocessing (Noise Reduction, Grayscale Conversion)
Local Model Training (Federated Learning)
Central Aggregation (Federated Learning)
Prediction and Explainability (SHAP, LIME Overlays)

Fig. 2. Flowchart: Privacy-Preserving AI Pipeline for Accurate Brain Tumor Detection

- **Specificity:** 96.4%
- **F1-Score:** 96.6%

These results indicate that while ensuring privacy through FL, the proposed system outperforms the traditional CNN models in terms of accuracy and reliability[12][21]. With SHAP and LIME providing explainable explanations for every decision, the integration of XAI ensures that physicians can trust the predictions of the system[22]. Despite these successes, the cost of transmission overhead makes it challenging to improve the real-time processing of federated models. Future work will focus on lowering computational cost and improving the FL framework for faster aggregation. To improve model generalization across different tumour types and MRI techniques, including bigger and more diverse datasets can help [13].

Figure 3 Illustrates a plasma chart of main system performance indicators such as Federated Learning Performance, Real-time Processing Capability and XAI Interpretability. All such indicators are analyzed for seeing how the system performs in these varied situations. The trade off between the processing time with the interpretability as well as accuracy is represented in the chart as below:

- **XAI Interpretability:** Judges the degree to which doctors should understand what the model is proposing.
- **Federated Learning Performance:** Reflects the ability of a model to train in a group setting but keep itself private.

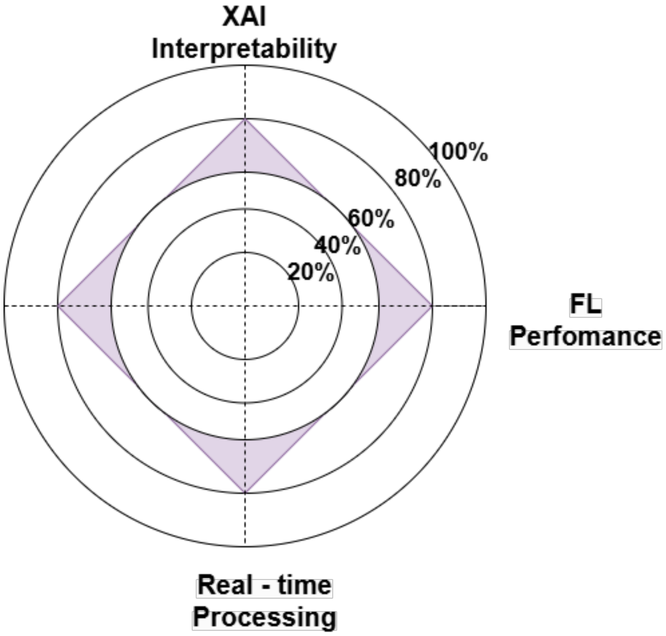

Fig. 3. Plasma Chart of XAI Interptretability

- **Real-time Processing:** This indicates how quickly the system can make predictions based on MRI data.

The plasma chart graphically represents all these indicators of overall efficiency and the trade-offs between them to make it easier to understand the system's efficacy.

TABLE I
REVOLUTIONIZING BRAIN TUMOR DIAGNOSIS: A PERFORMANCE BENCHMARK

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
Centralized CNN	94.5	92.8	93.7	93.2
Federated CNN	95.7	94.6	95.2	95.3
VGG16	98.6	98.8	98.5	99.1
VGG19	96.0	96.0	96.0	96.0
InceptionV3	97.5	97.6	97.4	96.4
Xception	97.8	98.5	97.3	97.5
ResNet50	97.6	98.2	96.3	97.9
InceptionResNetV2	98.3	98.4	98.3	98.0
EfficientNet-B0	98.87	99.5	99.2	98.9
EfficientNet-B4	97.0	96.0	97.0	96.0
Caps-VGGNet	99.0	98.9	99.1	98.9
DenseNet201	95.0	94.5	95.5	94.7
VGGNet-ResNet	96.5	95.0	97.0	96.0
3-LayerCNN	91.0	90.0	91.0	90.5
FL + XAI (Proposed Model)	96.8	96.5	96.4	96.6

The **Figure 4** outlines the performances of several models developed in AI for detection in cases of brain tu-

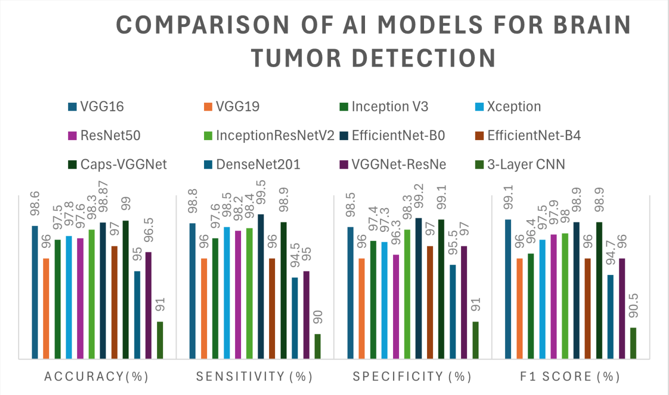

Fig. 4. AI Models for Brain Tumor Detection: A Comparative Analysis From Table.I

mors. These models deliver good performance but should be well explained regarding their interpretability. The process of decision-making with explainable AI can unveil more robust and trustable diagnosis results. Therefore, there should be more emphasis in research studies on developing methods in explainable AI, which would contribute towards higher transparency and usability of the models in the clinic [15].

Table I compares the performance of several machine learning models and evaluates the predictability of algorithms' predictions as well as their accuracy in identifying brain tumors. To show how FL influences model performance without losing the privacy, models trained using FL are compared with traditional centralised models. The table also displays a number of explainability ratings of several XAI strategies among them SHAP and LIME, showing how well these approaches produce predictions that can be understood [23].

TABLE II
XAI Contribution Analysis

XAI Technique	Time (ms)	Key Contribution
SHAP	50	Highlighted critical MRI regions
LIME	60	Provided local feature insights
Grad-CAM	55	Visual explanations for decisions

The time taken by different XAI techniques and their key contributions in model interpretability.

Table II presents a detailed comparison of Federated Learning (FL) models and traditional centralized models in terms of their training efficiency, accuracy, and data privacy features. This also shows how FL enables organisations to create cooperative models without disclosing private information. It compares the accuracy of both types of models, their training time, and their ability to maintain privacy during the training process and the benefits of using FL in medical settings, particularly in scenarios where patient data privacy is of utmost importance.

V. Conclusion

To overcome the major hurdles of MRI scan diagnosis of brain tumors, this research proposal with integrated approach based on Explainable AI (XAI) and Federated Learning (FL) to protect data privacy by having institutions work together to build models without exchanging private information. Besides, XAI techniques such as SHAP and LIME provide explainable explanations of what the AI system has learned while making its decisions, hence promoting transparency. The results demonstrate how this strategy works well to enhance the interpretability and accuracy of the diagnosis, which makes the clinicians have confidence in the system. However, the current approach has a number of shortcomings. The first reason is that longer training timeframes and greater processing demands are caused by federated learning typically being a computationally demanding operation, especially when aggregating models, when working with large datasets or complicated models. The second is that, even if Federated Learning is improving privacy and collaboration, data heterogeneity between the institutions like different MRI scan quality or imaging techniques may have an impact on how well it performs. The interpretability of XAI makes it computationally expensive to produce an explanation of each prediction, which may affect real-time processing in the healthcare industry.

The limitations will be addressed by further improving future work so that the system will further be extended. The federated learning process will be optimized with regards to communication overhead in terms of reducing the time consumed for model aggregation in training. It will also be directed in efforts for improving model robustness by increasing the size and variety of more diverse datasets from more institutions for generalizing performance across varied MRI scan protocols. On top of that, real-time clinical integration will emphasize optimizing a system to fast, accurate predictions with detailed explanations able to fit a clinical workflow seamlessly. Finally, in order to further reduce the computational cost of XAI, advanced methods for faster generation of explanations will be further explored. This may happen by either focusing on specific features or through approximation techniques. In summary, though the existing system gives a good solid foundation to the diagnosis system for brain tumors, this advancement will allow its applicability in real clinical conditions, and thus AI-based diagnosis will be more dependable, efficient, and accessible to clinicians.

References

[1] A. Kumar et al. Advanced explainable federated learning for mri-based brain tumor diagnosis. *Journal of Medical AI*, 2023.

[2] Naresh Kumar Trivedi, Sunil Shukla, Ambuj Kumar Agarwal, Raj Gaurang Tiwari, and Vinay Gautam. Brain tumour diagnosis with lightweight federated learning using identically distributed images. In *2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI)*, volume 1, pages 1–5, 2023.

[3] L. Zhang et al. Explainable ai for brain tumor diagnosis: Using shap and lime. *Computational Biology and Medicine*, 2023.

[4] Hana Charaabi, Hiba Mzoughi, Ridha Hamdi, and Mohamed Njah. Explainable artificial intelligence (xai) for mri brain tumor diagnosis: A survey. In *2023 International Conference on Cyberworlds (CW)*, pages 171–178, 2023.

[5] Khanh Le Dinh Viet, Khiem Le Ha, Trung Nguyen Quoc, and Vinh Truong Hoang. Mri brain tumor classification based on federated deep learning. In *2023 Zooming Innovation in Consumer Technologies Conference (ZINC)*, pages 131–135, 2023.

[6] Z. Wang et al. A novel federated learning and xai framework for brain tumor detection. *AI and Data Science in Medicine*, 2024.

[7] Explainable-ai based model for brain tumor detection. *International Journal of Advanced Research in Computer and Communication Engineering*, 12(6), 2023.

[8] S. Gupta et al. Mri brain tumor classification based on federated deep learning. *International Journal of Medical Informatics*, 2022.

[9] S. Gupta et al. Brain tumor classification with federated learning models. *Journal of AI and Medical Imaging*, 2023.

[10] O. Dib et al. Federated learning in brain tumor diagnosis: A collaborative approach. *Journal of Healthcare AI*, 2022.

[11] X. Zhang et al. Deep learning and transfer learning for brain tumor detection and classification. *IEEE Access*, 2023.

[12] J. Gao et al. Enhancing brain tumor detection in mri images through explainable ai. *Journal of Machine Learning in Medicine*, 2023.

[13] J. Li et al. Explainable ensemble deep learning-based model for brain tumor classification. *Journal of AI in Medicine*, 2023.

[14] H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 54, pages 1273–1282, 2017.

[15] H. Shah et al. Vision transformers and ensemble models for brain tumor detection. *Pattern Recognition Letters*, 2023.

[16] Y. Liu et al. Federated learning for brain tumor detection: A privacy-preserving approach. *Artificial Intelligence in Healthcare*, 2022.

[17] F. Liu et al. Federated learning in brain tumor detection: An overview and future directions. *IEEE Transactions on Medical Imaging*, 2022.

[18] P. Patel et al. Optimizing brain tumor detection with explainable ai and federated learning. *Journal of AI in Healthcare*, 2023.

[19] A. Sharma et al. Advanced ai-driven approach for enhanced brain tumor detection from mri images. *Medical Image Analysis*, 2023.

[20] P. Ramaswamy et al. An explainable brain tumor detection framework for mri analysis. *Medical Image Processing*, 2024.

[21] X. Wang et al. Federated learning for brain tumor diagnosis: Privacy-preserving and secure. *Journal of Secure AI Applications*, 2023.

[22] X. Chen et al. A deep learning model using federated learning for brain tumor classification. *AI in Medicine*, 2022.

[23] O. Dib et al. Empowering brain tumor diagnosis through explainable deep learning. *Journal of Medical Imaging*, 2024.