# Leveraging Federated Learning & Explainable AI for Robust Brain Tumor Diagnosis


**Submitted By**

**MASRAFE BIN HANNAN SIAM**

221-35-1022

Department of Software Engineering

Daffodil International University

**Supervised by**

**MR. MD. SHOHEL ARMAN**

Assistant Professor

Department of Software Engineering

Daffodil International University


A thesis submitted in partial fulfillment of the requirement for the degree of

Bachelor of Science in Software Engineering

Fall-2025

# Leveraging Federated Learning & Explainable AI for Robust Brain Tumor Diagnosis

MASRAFE BIN HANNAN SIAM

Bachelor of Science

DAFFODIL INTERNATIONAL UNIVERSITY

## APPROVAL

This thesis titled on "**Leveraging Federated Learning & Explainable AI for Robust Brain Tumor Diagnosis**", submitted by **MASRAFE BIN HANNAN SIAM (ID: 221-35-1022)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

### BOARD OF EXAMINERS

**Chairman**

**Dr. Imran Mahmud**
**Professor & Head**
Department of Software Engineering
Faculty of Science and Information Technology Daffodil International University

**Internal Examiner 1**

**Afsana Begum**
**Assistant Professor**
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

**Internal Examiner 2**

**Md. Shohel Arman**
**Assistant Professor**
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

**Internal Examiner 3**

**Nadira Islam**
**Assistant Professor**
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

**External Examiner**

**Md Manowarul Islam**
**Professor**
Department of Computer Science and Engineering
Jagannath University, Bangladesh

**DAFFODIL INTERNATIONAL UNIVERSITY**

**DECLARATION OF THESIS AND COPYRIGHT**

| | | |
|---|---|---|
| Author's Full Name | : | MASRAFE BIN HANNAN SIAM |
| Date of Birth | : | 09 March 2002 |
| Title | : | Leveraging Federated Learning & Explainable AI for Robust Brain Tumor Diagnosis |
| Academic Session | : | Fall 2025 |

I declare that this thesis is classified as:

| | | |
|---|---|---|
| ☑ | CONFIDENTIAL | (Contains confidential information under the Official Secret Act 1997) * |
| ☑ | RESTRICTED | (Contains restricted information as specified by the organization where research was done) * |
| ☐ | OPEN ACCESS | I agree that my thesis to be published as online open access (Full Text) |

I acknowledge that Daffodil International University reserves the following rights:

1. The Thesis is the Property of Daffodil International University.
2. The Library of Daffodil International University has the right to make copies of the thesis for the purpose of research only.
3. The Library of Daffodil International University has the right to make copies of the thesis for academic exchange.

Certified by:

_____
(Student's Signature)

MASRAFE BIN HANNAN SIAM
_____
Student ID :221-35-1022
Date :26/11/2025

_____
(Supervisor's Signature)

MR. MD. SHOHEL ARMAN
_____
MR. MD. SHOHEL ARMAN
Date :26/11/2025

NOTE: * If the thesis is CONFIDENTIAL or RESTRICTED, please attach a thesis declaration letter.

# Supervisor's Declaration

I hereby declare that I have checked this thesis and, in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Bachelor of Science.

_____

(Supervisor's Signature)

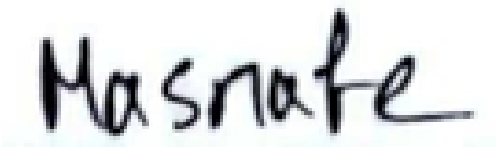
Full Name : MR. MD. SHOHEL ARMAN

Position : Assistant Professor

Date : 26 November 2025

# Student's Declaration

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Daffodil International University or any other institution.



_____

(Student's Signature)

Full Name : MASRAFE BIN HANNAN SIAM

ID Number : 221-35-1022

Date : 26 November 2025

# Leveraging Federated Learning & Explainable AI for Robust Brain Tumor Diagnosis

MASRAFE BIN HANNAN SIAM

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Bachelor of Science

Department of Software Engineering (Major in Data Science)

DAFFODIL INTERNATIONAL UNIVERSITY

November 2025

# Acknowledgement

Since childhood, I've been fascinated by new ideas and by figuring out how things work, and that curiosity is what ultimately led me toward the field of Machine Learning. Getting to work with lung cancer images and design models and federated learning pipelines for early detection has been a genuinely meaningful and rewarding part of my journey. I am deeply grateful to the Almighty for giving me the strength, patience, and opportunity to complete this work.

I owe a profound debt of gratitude to my parents, whose love, support, and encouragement have been constant throughout my life. Their prayers and belief in me have played a major role in helping me reach this stage and complete this thesis.

I would also like to express my sincere thanks to Dr. Imran Mahmud, Head of the Department of Software Engineering, and to all my respected teachers at Daffodil International University. Their guidance, knowledge, and continuous support have been invaluable for both this research and my overall academic growth.

My heartfelt gratitude goes to my supervisor, Mr. Md. Shohel Arman. At every step of this work, his time, knowledge, and helpful criticism have been very important. His consistent supervision, thoughtful suggestions, and encouragement helped me stay focused and push through the challenging phases of this thesis.

Lastly, I am truly thankful to my friends and classmates at DIU. Their cooperation, late-night discussions, and constant motivation made this journey not only easier but also far more enjoyable. I deeply appreciate the support they have given me and the positive energy they brought throughout this experience.

# Abstract

Brain MRI classifiers sometimes have trouble generalizing between hospitals because of privacy rules and differences in protocols. In addition to being accurate, clinicians need clear models that back up predictions with anatomically sound evidence. We develop and evaluate an explainable federated learning (FL) framework that safeguards privacy for the classification of four categories of brain tumors: glioma_tumor, meningioma_tumor, pituitary_tumor, and no_tumor. Our objective is to align model selection with a quantitative notion of explanation faithfulness as well as generalization performance. We train parameter-efficient CNNs (ShuffleNetV2, RegNetY400, MobileNetV3-Large), deeper CNNs (ResNet-50, DenseNet-121), a compact Custom CNN, and (Hybrid Swin-T + DenseNet-121, MLP) under synchronous FedAvg using 10,417 de-identified, single-channel MRI slices distributed across four clients. A harmonized classification head and preprocessing pipeline are shared by all backbones at 224×224. After each local round, clients compute Grad-CAM++ overlays and a lightweight, deletion-style faithfulness score on a fixed validation subset; they never share images or heatmaps with the server, only model weights and scalar summaries (loss, ACC, Macro-F1, and faithfulness mean±std). The server aggregates updates, records round-wise trajectories, uses validation Macro-F1 for checkpoint selection, and applies tie-breaking rules that favor higher and more consistent faithfulness. The best held-out accuracy is achieved by RegNetY400 (Test ACC = 0.9827), followed closely by ShuffleNetV2 and MobileNetV3-Large. The hybrid Swin-T+DenseNet121 exhibits the smallest cross-client dispersion, indicating particularly stable performance across sites. In terms of explanation quality, ShuffleNetV2 attains the strongest combination of top validation F1 (0.9592 at R20) and deletion-style faithfulness (mean 0.38 at R18), with RegNetY400 ranking second (mean 0.41 at R20). When rounds are close to their respective best checkpoints, models with higher faithfulness generally also show higher validation F1, suggesting a positive coupling between generalization and explanation quality. Overall, the proposed FL pipeline turns explainability from a purely post-hoc visualization step into a federated training signal. This design makes the model more accurate without giving away any raw data. By combining strong backbones with a quantitative, privacy-preserving faithfulness metric, the system makes accurate, clear, and auditable models that can be used at many clinical sites.

# Table Of Contents

# List Of Figures

# List Of Tables

# CHAPTER 1
# Introduction

## Introduction

Brain tumors are among the most serious neurological conditions, often leading to lasting disability or death if diagnosis is delayed. In day-to-day clinical practice, magnetic resonance imaging (MRI) is the primary non-invasive tool for locating and characterizing these lesions, helping clinicians distinguish between glioma, meningioma, pituitary tumors, and cases with no visible tumor. In principle, deep learning models can support this process by learning subtle visual patterns directly from MRI data and providing fast, consistent predictions that complement expert judgement [4]. But in practice, it's still hard to turn these promising models into reliable, usable solutions that work with more than one hospital and scanner.

One big problem is that medical image data are spread out across different organizations. Strict privacy regulations, ethical considerations, and organizational policies make it difficult or even impossible to pool raw MRI scans into a single centralized repository. Federated learning (FL) offers a way around this barrier by shifting the focus from "moving data" to "moving models": each hospital trains a shared model locally on its own de-identified images and only sends parameter updates back to a coordinating server for aggregation [1–3]. Early work in medical imaging has shown that this strategy can enable multi-site training without exposing raw data, including for brain tumor MRI segmentation and classification tasks [6, 10]. Even so, federated models must still cope with non-IID data, protocol differences, and variable client participation, all of which can affect how well they generalize in real clinical settings.

At the same time, there is growing agreement that accuracy alone is not enough for clinical adoption. Radiologists and neurosurgeons need to know whether a model is "looking" at the right structures and basing its decisions on plausible neuroanatomical evidence rather than artifacts or dataset shortcuts. Explainable AI (XAI) methods such as Grad-CAM create class-specific saliency maps that highlight which regions of an MRI slice were most influential for a given prediction, making model behavior more transparent and easier to critique [8]. However, most existing studies still treat these explanations as a qualitative afterthought useful for illustrative figures, but not tightly integrated into how models are trained, selected, or monitored over time.

Recent work on perturbation-based evaluation has started to close this gap by providing quantitative measures of explanation "faithfulness." Techniques such as deletion curves track how quickly a model's confidence drops as the most salient pixels are gradually removed, offering a way to score how strongly a heatmap is tied to the model's actual decision process [5, 9]. Despite their promise, these metrics have rarely been incorporated into federated learning pipelines, where privacy constraints, heterogeneous data, and limited communication budgets already make training more complex.

This thesis is motivated by these scientific and practical gaps. Building on the ideas above, it develops a federated learning framework for four-class brain tumor classification from MRI that treats explainability as a first-class objective rather than an optional extra. Multiple CNN and hybrid backbones are trained collaboratively across simulated hospital clients using FL [10], while each client also computes Grad-CAM maps and a deletion-style faithfulness score on a fixed validation subset [11]. Only model weights and compact scalar summaries are shared

with the server, so privacy is preserved. By jointly tracking performance metrics (such as accuracy and macro-F1) and explanation faithfulness across rounds, the thesis aims to produce models that are not only accurate, but also more transparent and trustworthy for multi-site brain tumor diagnosis.

# Background

Brain tumors are a major cause of neurological illness and cancer-related death worldwide, and their outcome depends heavily on how early and accurately they are detected. In most hospitals, magnetic resonance imaging (MRI) is the standard, non-invasive way to visualize brain structure and identify suspicious lesions. These scans help clinicians distinguish between common tumor types such as glioma, meningioma, and pituitary adenoma and cases where no tumor is present. At the same time, MRI protocols, scanner vendors, and acquisition settings can differ significantly from one institution to another, which makes it difficult to build a single, robust model that works equally well everywhere.

Over the last decade, deep learning especially convolutional neural networks (CNNs) and hybrid CNN transformer architectures has shown strong performance in classifying and segmenting brain tumors from MRI, often matching or exceeding traditional machine-learning pipelines that rely on hand-crafted features [12]. These models can capture complex visual patterns directly from images and provide fast, automated predictions that are useful for triage, second opinions, and decision support. However, many of the best-performing systems are trained on centralized datasets that combine images from multiple sites into a single server. In practice, such aggregation is often blocked or discouraged by privacy regulations, institutional policies, and the understandable reluctance to share raw medical data outside local control [2,3].

Federated learning (FL) was proposed as a way to move beyond this “centralized data” assumption while still benefiting from diverse, multi-site training [1]. Instead of uploading images to a common repository, each participating hospital (or “client”) keeps its MRI data on site, trains a shared model locally, and only sends model updates such as weights or gradients back to a coordinating server for aggregation [1–3]. Early studies in medical imaging have demonstrated that FL can preserve data privacy while still achieving strong performance on segmentation and classification tasks, including brain tumor MRI [14]. Nonetheless, federated models must cope with additional challenges such as non-identically distributed (non-IID) data across clients, different class balances, and variable participation, all of which can affect convergence and generalization.

At the same time, there is growing recognition that high accuracy alone is not sufficient for clinical use. Clinicians need tools that are not only correct on average, but also transparent in how they arrive at individual predictions. Explainable AI (XAI) methods such as Grad-CAM produce class-specific heatmaps that highlight which regions of an input image were most influential for the model’s decision, making it easier to judge whether the model is focusing on plausible tumor regions or being distracted by artifacts and background structures [13]. More recent work on perturbation-based evaluation, including deletion and insertion curves, provides quantitative ways to assess how “faithful” these saliency maps are to the underlying decision process [5,9]. However, most existing applications of XAI in brain tumor imaging and medical FL still use explanations as a separate, qualitative step rather than integrating them as measurable signals during training and model selection [10–14]. This gap motivates the present thesis, which explores how federated learning and explainable AI can be combined to produce

brain tumor classifiers that are both privacy-preserving and more trustworthy in their decision-making.

# Problem Statement

Despite the clinical importance of early and reliable brain tumor diagnosis, there is still not widely adopted, automated pipeline that can transform multi-centre brain MRI data into accurate and trustworthy predictions while respecting strict privacy constraints. In routine practice, radiologists must visually inspect MRI scans and integrate multiple sequences, which is time-consuming and prone to variability between readers, especially under heavy workload. Deep learning models have shown that they can classify and localize brain tumors with high accuracy on research datasets, but these models are usually trained on centrally collected data and often overlook real-world constraints such as hospital-specific protocols, heterogeneous scanners, and local data-governance rules [10–13]. As a result, many promising models remain confined to single-centre studies and do not translate smoothly into everyday clinical workflows.

Federated learning (FL) has been proposed to train models collaboratively without moving raw data off-site, by keeping patient images within each hospital and only exchanging model updates with a central server. Initial applications in medical imaging suggest that FL can approach the performance of centralized training on tasks such as segmentation and classification, including brain tumor MRI, while offering stronger privacy guarantees [2]. However, most of these studies focus primarily on classification or segmentation accuracy and convergence speed. They pay less attention to how well the resulting global model behaves across clients with non-identically distributed (non-IID) data, and they rarely provide mechanisms for systematically monitoring whether model decisions remain clinically meaningful as training progresses. This is a critical gap when the goal is not just to "get high accuracy," but to build tools that clinicians can trust across diverse sites.

At the same time, explainable AI (XAI) methods such as Grad-CAM have become popular for visualizing which regions of an MRI slice influence the model's prediction, and perturbation-based metrics like deletion curves offer a way to quantify the "faithfulness" of these explanations to the underlying decision process [4,5]. Yet in most existing work, these explanations are generated only after training and are used qualitatively for example, to produce a few illustrative heatmaps rather than being integrated into the training loop or used as part of model selection in federated settings [14]. There is currently no established framework that combines multi-client FL for brain tumor classification with client-side explanation generation and quantitative faithfulness scoring, while keeping raw data private and providing the server with compact, comparable signals about both performance and trustworthiness.

The core problem addressed in this thesis, therefore, is the absence of a practical, privacy-preserving federated learning framework for four-class brain tumor MRI classification that treats explainability as a first-class, measurable objective. Specifically, there is a need for an approach that can

(i) Train and evaluate multiple backbone architectures across simulated hospital clients using FL [6,10].
(ii) Preserve data privacy by sharing only model parameters and summary statistics.
(iii) Quantify and track explanation faithfulness alongside conventional metrics such as accuracy and macro-F1 [7–9].

This work aims to fill that gap by designing and evaluating such a framework.

# Research Gaps

Although there has been substantial progress in automated brain tumor diagnosis using deep learning and, more recently, federated learning, several important gaps remain between current research and what is needed for a practical, trustworthy clinical system. First, most high-performing brain tumor classifiers are still developed under a centralized training paradigm, where MRI scans from multiple sources are pooled into a single dataset [7]. This setting does not reflect the reality of hospital data governance, where privacy regulations and institutional policies often prevent large-scale data sharing [3]. Existing FL studies in medical imaging have shown that collaborative training without raw data exchange is feasible [6,14], but they often focus on binary or limited-class tasks, single backbone architectures, or simplified client distributions, leaving open questions about how well FL can support more realistic, four-class brain tumor classification setups with diverse model families.

Second, the behavior of federated models under non-IID conditions is still not sufficiently characterized for this application. Many works demonstrate that FL can approach the performance of centralized training in aggregate [3] but provide limited analysis of how individual clients with imbalanced or skewed class distributions affect global convergence, stability, and fairness. There is relatively little work that compares multiple CNN and hybrid backbones such as DenseNet, ResNet, MobileNet-style and transformer-enhanced models under the *same* federated configuration for brain MRI, while systematically examining cross-client variability and robustness. This makes it difficult to answer practical design questions, such as which architectures offer the best trade-off between performance, communication cost, and stability across heterogeneous sites.

Third, most existing studies treat explainable AI as a separate, post-hoc step rather than an integral component of the training and evaluation pipeline. Grad-CAM and related saliency methods are widely used to generate visual explanations for deep models in brain tumor imaging [8], but these heatmaps are typically shown only for a handful of examples and are interpreted qualitatively. Perturbation-based metrics such as deletion and insertion curves provide a way to quantify how faithfully these explanations reflect the model's internal decision process [5,9], yet they are rarely reported alongside standard metrics like accuracy and F1, and almost never used to guide model selection or monitor training dynamics in federated settings.

Finally, there is a notable lack of frameworks that combine federated learning with quantitative explainability in a privacy-preserving way. To the best of current knowledge, prior work has not systematically explored a setup where each client generates both Grad-CAM explanations and a numerical faithfulness score on local validation data, then shares only compact summary statistics with the server. As a result, servers in most FL studies have no direct signal about how trustworthy the model's reasoning is across sites, and cannot, for example, choose between two models with similar accuracy but different explanation quality. This thesis addresses these gaps by proposing and evaluating a framework that jointly considers classification performance, federated training behavior, and explanation faithfulness for four-class brain tumor MRI classification.

# Objectives

The overall objective of this thesis is to design and evaluate a federated learning framework for four-class brain tumor classification from MRI that is not only accurate and privacy-preserving, but also explicitly incorporates explainability as a measurable property of the

model's behavior. Instead of relying on centralized data, the framework is built around multi-client training with local explanation generation, so that the global model can be assessed in terms of both predictive performance and the faithfulness of its visual explanations.
More specifically, the thesis pursues the following objectives:

1. Develop a multi-client FL pipeline for brain tumor MRI classification:
   Implement a synchronous federated learning setup in which multiple simulated hospital clients collaboratively train a shared model for four-way classification (glioma, meningioma, pituitary, no_tumor) using de-identified MRI slices, while keeping all raw images within each client and exchanging only model parameters and summary statistics with the central server.

2. Benchmark diverse backbone architectures under a unified FL configuration:
   Fine-tune and compare several convolutional and hybrid backbones such as lightweight CNNs and deeper or transformer-enhanced models under the same federated protocol, using a harmonized classification head, fixed data splits, and shared hyperparameters. The goal is to find which architectures give the optimum trade-off between accuracy, macro-F1, and practical issues like model size and stability among clients.

3. Integrate explainable AI into the federated pipeline:
   Equip each client with an explainability module that computes Grad-CAM heatmaps on local validation images, so that class-specific regions of interest can be visualized for the federated models without exposing raw MRI data to the server. This allows qualitative inspection of whether the models focus on plausible tumor regions across different sites.

4. Quantify explanation faithfulness using perturbation-based metrics:
   Move beyond purely qualitative heatmaps by computing a deletion-style faithfulness metric (Deletion AUC) on a fixed validation subset at each client. This provides a numerical score for how strongly the highlighted regions influence the model's predictions, enabling objective comparison of explanation quality across rounds and architectures [5,9].

5. Jointly analyze performance and faithfulness across federated rounds:
   We continuously track standard performance metrics loss, accuracy, and macro-F1 alongside Deletion~AUC from all clients to see how they evolve over the federated rounds. In particular, we examine whether models with higher validation F1 also achieve higher and more stable faithfulness scores, and how this relationship differs across backbones.

6. The aim is to use these XAI signals not just for visualization, but to guide how we select, compare, and interpret models in a more informed way:
   Investigate how quantitative explainability metrics can be used in conjunction with accuracy for example, as a tie-breaker between models with similar validation performance to support more informed selection of global checkpoints that are both accurate and trustworthy. The aim is to outline practical guidelines for making explainability a first-class signal in federated medical imaging workflows.

Together, these objectives are intended to bridge the gap between purely performance-driven federated learning studies and the clinical need for transparent, privacy-aware brain tumor diagnosis systems that can be trusted across multiple institutions.

## Motivation

The motivation for this thesis comes from the gap between what is technically possible with modern deep learning and what is actually needed in day-to-day neuro-oncology practice. Brain tumors remain life-threatening conditions where small delays or misinterpretations in imaging can have serious consequences for treatment planning and patient outcomes. MRI has become the workhorse modality for detecting and characterizing these lesions but interpreting multi-sequence scans for large numbers of patients is demanding, time-consuming work that depends on sustained expert attention. In busy clinical environments, it is unrealistic to expect that every case will receive unlimited time from multiple specialists, yet the stakes of each decision remain high [10].

At the same time, research over the last decade has shown that deep learning models can recognize subtle patterns in brain MRI and achieve strong performance in tumor classification and segmentation, often approaching or surpassing traditional pipelines based on hand-crafted features [9]. This creates a tempting vision: AI systems that help radiologists triage cases, highlight suspicious regions, and provide a consistent second opinion. However, many of these models are trained on centrally aggregated datasets that are difficult to reproduce in real hospitals because of privacy regulations, institutional policies, and patient trust concerns [2,3]. In practice, most medical centres still hold their own data locally and are reluctant to share raw scans beyond their firewall.

Federated learning offers a promising compromise by allowing institutions to "collaborate without sharing data": each client trains the model on its own de-identified images, and only updates to the model parameters are sent to a central server for aggregation. Early studies indicate that such federated setups can come close to centralized performance on medical imaging tasks, including brain tumor MRI [6]. For someone interested in practical deployment, this is highly motivating FL provides a realistic path to multi-centre brain tumor classifiers that respect privacy laws and local governance rules. Yet existing work often stops at demonstrating that accuracy is "good enough," without fully addressing how these models behave under non-IID client distributions or how their decisions can be trusted in a setting where no single institution ever sees the full dataset [14].

Trust is not just a matter of numerical accuracy. Clinicians need to know whether a model is basing its decisions on plausible tumor regions or on spurious correlations and artifacts. Explainable AI methods like Grad-CAM have become popular because they can visualize which parts of an MRI slice contributed most to a prediction, giving radiologists a way to cross-check the model's focus with their own expectations. However, in many studies these explanations are used only for a few illustrative examples and are not treated as something that can be measured, compared, or monitored systematically. Perturbation-based metrics, such as deletion curves and Deletion AUC, offer a way to quantify how tightly a saliency map is linked to the model's internal decision process [5], but they are rarely integrated into federated learning pipelines.

This thesis is motivated by the belief that a practical AI assistant for brain tumor diagnosis must be both privacy-preserving and meaningfully explainable. Relying solely on accuracy,

especially in a multi-site federated setting, leaves important questions unanswered: does the global model behave consistently across clients with different class distributions? Do high-performing checkpoints actually rely on clinically relevant image regions? And can we detect when a model's reasoning becomes less trustworthy, even if its accuracy remains stable? By designing a framework in which each client not only trains the model but also computes Grad-CAM explanations and quantitative faithfulness scores on local validation data, this work aims to move explainability from a cosmetic add-on to a core part of how federated models are trained, selected, and evaluated.

## Summary

This chapter has set the stage for the rest of the thesis by outlining why automated, trustworthy brain tumor diagnosis from MRI is both important and challenging. We began by describing the clinical context in which radiologists must interpret complex MRI scans for conditions that are time-critical and potentially life-threatening, and we highlighted how recent advances in deep learning have produced powerful tumor classifiers that, in principle, could support this work. At the same time, we noted that most existing models assume access to centralized multi-centre datasets an assumption that clashes with real-world constraints around data privacy, institutional policies, and local governance.

We then discussed how federated learning offers a way for multiple institutions to collaborate without sharing raw images, and how early studies in medical imaging, including brain tumor MRI, suggest that FL can approach centralized performance while keeping data on-site. However, we also pointed out that accuracy alone is not sufficient for clinical adoption, especially in heterogeneous, non-IID settings. Explainable AI methods such as Grad-CAM and perturbation-based metrics like deletion curves provide tools to visualize and quantify how models make decisions, but they are still rarely integrated as core signals in federated pipelines.

From this discussion, we formulated the central problem of this thesis: the lack of a practical, privacy-preserving framework for four-class brain tumor classification that jointly considers classification performance and explanation faithfulness in a federated setting. In response, we defined a set of objectives that include building a multi-client FL pipeline, benchmarking diverse backbones under a unified configuration, and embedding both qualitative (Grad-CAM) and quantitative (Deletion AUC) explainability into the training and evaluation process. The overarching motivation is to move from purely performance-driven studies toward models that are both accurate and meaningfully interpretable across sites.

The next chapter builds on this foundation by reviewing related work in more detail. It surveys existing research on deep learning for brain tumor MRI, federated learning in medical imaging, and explainable AI for clinical decision support, and identifies more precisely where the proposed framework fits within and extends the current literature.

# CHAPTER 2
# Literature Review

## Introduction

This chapter examines the current literature concerning automated brain tumor diagnosis via MRI, emphasizing deep learning, federated learning, and explainable AI. The aim is to understand how current methods are designed, what levels of performance they achieve, and where they fall short when considered from the perspective of real-world deployment. In particular, we look at three interconnected threads: (i) centralized deep learning approaches for brain tumor classification and segmentation, (ii) federated learning in medical imaging as a privacy-preserving alternative to centralized training, and (iii) explainable AI techniques, including both visual heatmaps and quantitative faithfulness metrics, that seek to make model decisions more transparent [1].

By organizing the literature along these axes, we can better position the proposed framework within current research. The discussion highlights how prior work has demonstrated strong performance on benchmark datasets, but often under assumptions such as centralized data access or purely qualitative explanations that are difficult to satisfy in routine multi-centre practice [2]. The chapter concludes by summarizing the key gaps that motivate this thesis and by clarifying how the proposed approach extends existing studies on federated learning and explainable AI for brain tumor MRI [4, 5].

## Previous Literature

### Deep Learning for Brain Tumor MRI

Early computer-aided diagnosis systems for brain tumor analysis mostly relied on hand-crafted features extracted from MRI slices such as intensity statistics, texture patterns, and shape descriptors which were then fed into traditional classifiers like support vector machines or random forests. These approaches were helpful as early baselines, but because they depended on manually designed features, they often struggled to generalize across different scanners, acquisition protocols, and clinical sites. With the rise of deep learning, especially convolutional neural networks (CNNs), the field shifted toward end-to-end models that automatically learn hierarchical feature representations directly from raw or minimally processed images [9].

Several studies have demonstrated that CNN-based models can achieve substantial outcomes on public brain tumor datasets, particularly in tasks such as glioma grading, tumor subregion segmentation, and multiclass tumor classification. Architectures such as ResNet, DenseNet, lightweight MobileNet-style networks, and newer hybrid CNN transformer models are now widely used as backbone feature extractors, since they can capture both fine-grained local texture and broader anatomical context [13]. In most cases, these networks are first pretrained on large natural image datasets and then fine-tuned on brain MRI, often with medical imaging–specific data augmentation to improve robustness and generalization. Reported performance on benchmark datasets is frequently high, with macro-F1 and accuracy scores that suggest practical utility for triage, second opinions, or quality assurance in radiology workflows.

Despite these advances, most of the literature still assumes access to centralized training data. Images from many institutions are integrated into a single repository, and the model is

optimized as if all data were held in one place. This setting simplifies experimentation but stands in tension with real-world privacy regulations, institutional policies, and patient expectations [2,3]. Moreover, many articles focus primarily on global performance measures, with minimal research of how models behave on specific subpopulations, during distribution shifts, or in the context of scanner and procedure heterogeneity. As a result, there remains a gap between benchmark performance and robust, multi-centre deployment.

## Federated Learning in Medical Imaging

Federated learning (FL) was introduced as a way to train models collaboratively without requiring centralized access to raw data, by sending a shared global model to each client, training it locally, and aggregating updates at a central server [3]. In its basic form, often referred to as Federated Averaging (FedAvg), the server initializes the model, distributes it to participating clients, and then repeatedly averages their locally updated parameters to produce a new global model [1]. This paradigm aligns naturally with hospital data governance, where institutions are willing to participate in collaborative training but are reluctant or legally unable to share raw medical images.

In medical imaging, several works have applied FL to tasks such as segmentation of brain tumors, cardiac structures, or organs-at-risk, as well as classification of lesions in brain, lung, and other organs [14]. These studies generally report that federated models can approach the performance of centrally trained models, while offering stronger privacy guarantees because patient-level images never leave the local site. Some works also explore variants of the basic FedAvg algorithm, such as methods designed to cope better with non-identically distributed (non-IID) data, communication constraints, or heterogeneous hardware across clients [6].

However, several limitations remain. Many federated imaging studies consider relatively simple class structures (e.g., binary classification or limited multi-class setups) and focus on a single backbone architecture. There is generally insufficient research of how client-specific factors such as class imbalance, local sample size, or acquisition technique differences affect convergence and global performance. Furthermore, the primary emphasis tends to be on standard metrics like accuracy or Dice coefficient, with less attention to how the model's decision-making process can be inspected or trusted across institutions [15]. This leaves open the question of how FL can be combined with systematic explainability to support more reliable multi-site brain tumor diagnosis.

## Explainable Ai for Brain Tumor Diagnosis

Explainable AI (XAI) has gained significance as physicians and regulators seek more openness from deep learning algorithms employed in high-stakes applications. In the area of brain tumor imaging, one of the most extensively used approaches is Grad-CAM, which produces class-specific heatmaps by backpropagating gradients from the output layer to convolutional feature maps [4 ; 12]. These heatmaps highlight image regions that contributed most strongly to a model's prediction and can be overlaid on MRI slices to help radiologists see whether the network is focusing on plausible tumor regions, peritumoral edema, or irrelevant background structures.

Many studies now include Grad-CAM or similar saliency maps as part of their evaluation, often presenting a small set of representative examples where the model "looks at the right place." This qualitative use of XAI can be helpful for building intuition and spotting obvious failure modes, but it does not easily scale or support systematic comparison between models. To

address this, perturbation-based evaluation techniques have been proposed, in which the most salient pixels are progressively removed or masked and the resulting drop in model confidence is tracked over time [5,9]. Metrics such as Deletion AUC summarize how quickly the model's output deteriorates when its highlighted evidence is removed, providing a quantitative notion of explanation faithfulness.

In brain tumor imaging, these quantitative XAI metrics are still relatively rare compared to visual heatmaps. Moreover, in most works both centralized and federated explanations are generated after training and used mainly for illustrative figures rather than being integrated into the training loop, model selection, or monitoring pipeline. In federated settings, there is the added complication that raw images cannot be shared with the server, making it difficult to centralize explanation analysis. This creates a need for frameworks in which explainability is computed locally at each client, summarized through compact statistics, and treated as a core signal alongside accuracy and loss [7; 11].

## Summary

The literature reviewed in this chapter shows clear progress in three areas: deep learning for brain tumor MRI, federated learning for privacy-preserving medical imaging, and explainable AI for visualizing and assessing model decisions. CNN-based and hybrid architectures have achieved strong performance on benchmark datasets; FL has demonstrated that multi-centre collaboration is possible without pooling raw data; and XAI methods such as Grad-CAM and perturbation-based metrics offer tools to probe how models make decisions.

At the same time, important gaps remain. Most high-performing brain tumor classifiers are trained under centralized data assumptions that do not align with real hospital constraints. Federated studies in medical imaging tend to prioritize accuracy and convergence while giving less attention to systematic, quantitative explainability across clients. Existing XAI applications in this domain are largely qualitative, limited to visual heatmaps and small case studies, with quantitative faithfulness metrics rarely being integrated into training or model selection.

These gaps motivate the framework proposed in this thesis, which explicitly combines multi-client federated learning for four-class brain tumor classification with client-side generation of Grad-CAM explanations and deletion-based faithfulness metrics. The method seeks to protect privacy by just disclosing model parameters and short summary statistics. This lets the server see how both the quality of the explanations and the performance change over time. The next chapter goes into a lot of information regarding the method used in this paper, including the dataset, federated setup, model architectures, and explainability procedure.

# CHAPTER 3
# Methodology

## Proposed Framework

We provide privacy-preserving and interpretable training of brain MRI classifiers through a synchronous FedAvg pipeline (Fig. 1). In our system, four client hospitals fine-tune pretrained backbones using de-identified, single-channel slices that have been resampled to 224 × 224 pixels. No raw images, patient identifiers, or explanation heatmaps are ever shared between sites. In this setup, we evaluate a compact custom CNN, a hybrid Swin-T + DenseNet-121 model, deeper CNNs such as ResNet-50 and DenseNet-121, and more parameter-efficient architectures like ShuffleNetV2, RegNetY400, and MobileNetV3-Large. For fairness and comparability, we replace each backbone's original classifier with a common prediction head, so that all models produce outputs in a consistent format.

GAP/Flatten → Linear → ReLU → Dropout → Linear → Softmax,

This illustrates the chances of each class for four labels: glioma_tumor, meningioma_tumor, pituitary_tumor, and no_tumor.

At round *t*, the server sends out the current global weights $w^t$ plus a brief fit-configuration that has the learning rate, the number of local epochs *E*, and the loss-function flags. Then, each client trains for E local epochs on its own patient-wise 80/10/10 split (train/validation/test), records {loss, accuracy, macro-F1}, and executes a lightweight client-side XAI probe (Grad-CAM on the last convolutional block + a deletion-style faithfulness test). Clients just send back their modified weights and a few scalar metrics.

The server calculates sample-size-based aggregation weights and updates the global model depending on the set of responding clients $S^t$ and their local training sizes $\{n_k\}$. We keep track of round-level metrics including loss, accuracy, and macro-F1, as well as federated faithfulness summaries. We use validation macro-F1 to find the best global checkpoint, and if there are ties, we choose the one with the higher and most consistent faithfulness scores.

We use a deletion curve to measure faithfulness. We gradually hide the most important pixels and keep track of the anticipated class probability $p_c(m)$ as a function of the deletion fraction $m$. We use the trapezoid rule to find the area under this curve, and then we flip it into a monotonic score so that larger numbers mean more faithfulness. The server only gets the mean and standard deviation of this score for each client.

This setup gives us clear, round-by-round learning curves and XAI trends that can be checked and audited, all without ever sharing raw data. In turn, it helps us select models that not only achieve strong classification performance but also provide reliable, consistent reasoning across all participating sites.

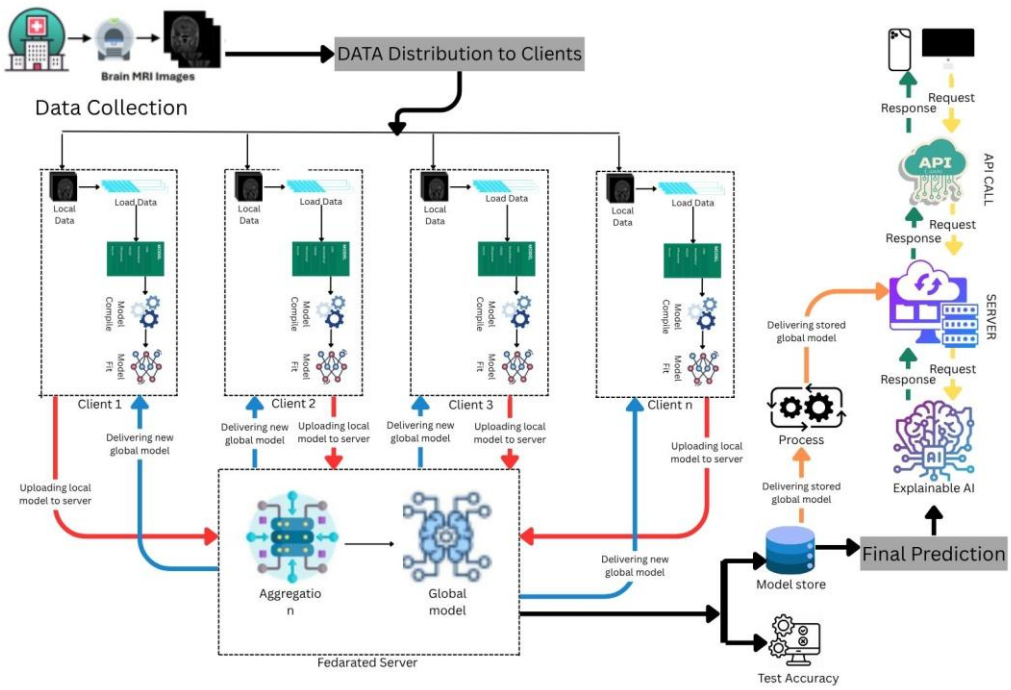



***Figure 1: The applied federated learning workflow for classifying brain MRIs into four classes, using client-side training and XAI probes along with server-side FedAvg aggregation and round-wise monitoring.***

# Federated Learning Process

We use a synchronous Federated Averaging (FedAvg) loop to classify brain MRIs with client-side explainability. In this system, the central server runs the training rounds, and each client trains on its own and sends back only model weights and a few scalar telemetry variables (loss, accuracy, macro-F1, and faithfulness). Images, patient identities, and saliency maps are never shared.

Let $S^t$ be the group of clients that answer correctly at round $t$ , $n_k$ be the size of client k's local training set, and $w^t$ be the current global parameters. FedAvg first figures out adjusted weights for sample sizes

$$\tilde{P}_k = \frac{n_k}{\sum_{J \in S^{(t)}} n_j},$$

and then modifies the global model to

$$\mathrm{w}^{(t+1)} = \sum_{\mathrm{k} \in S^{(t)}} \tilde{\mathrm{P}}_{\mathrm{k}} \mathrm{w}_{\mathrm{k}}^{(t)} = \mathrm{w}^{(t)} + \sum_{\mathrm{k} \in S^{(t)}} \tilde{\mathrm{P}}_{\mathrm{k}} \left( \mathrm{w}_{\mathrm{k}}^{(t)} - \mathrm{w}^{(t)} \right)$$

If some clients are delayed or skip a round, the server simply aggregates the updates from the clients that did respond and renormalizes the participation weights $pk$ accordingly. For this study, the federated loop is kept "crypto-free," meaning that secure aggregation and differential privacy mechanisms are not enabled. However, the architecture can work with both of these features if they are implemented in the future.

The logic for the round is as follows:

1. Broadcast: The server communicates the current global weights $w^{(t)}$ and a compact

fit configuration (learning rate, number of local epochs E, loss function, and scheduler parameters) to all clients.

2. Local training: Each client trains on its own 80/10/10 train/validation/test split using the preprocessing and augmentations that have been set up.

3. For the XAI probe, after training finishes each client runs Grad-CAM on the last convolutional layer and then computes a deletion-based AUC score to measure faithfulness, reporting the mean and standard deviation of this metric [17].

4. At the end of each round, the clients send back their updated model weights along with scalar metrics like loss, accuracy, macro-F1, and a summary faithfulness score.

5. The server then uses FedAvg to combine these weights and keeps track of the performance metrics from round to round to see how the global model changes over time. Validation macro-F1 picks the best global checkpoint. If there is a tie, it picks the model that has better and more consistent faithfulness over rounds.

## Local Training on Client *i*

Each client $i$ receives $w^t$ and a fit config, then trains on its private $D_i^{tr/val/te}$ (80/10/10, patient wise). Let $|D_i^{tr}| = n_i$, batch size $B_i$, local epochs $E$, and LR $\eta_t$. We use *AdamW* (weight decay $10^{-4}$) [18] with optional Reduce on Plateau.

Cost sensitive objective. For $K$=4 classes with client local counts $\{n_{i,k}\}$ and $\in$>0,

$$\omega_k = \frac{\frac{1}{n_{i,k}+\epsilon}}{\frac{1}{k}\sum_{j=1}^{k}\frac{1}{n_{i,k}+\epsilon}}, \qquad P = softmax\big(f_w(x)\big)$$

We train with weighted cross entropy $\mathcal{L}_{CE}(x,y) = -w_y \, log \, Py$; optionally switch to Focal Loss ($\gamma$=2) [19]:

$$\mathcal{L}_{FOCAL}(x,y) = -w_y\big(1-P_y\big)^{\gamma} \, log \, P_y$$

Label smoothing ($\varepsilon$=0.1) is supported by mixing the one hot target with the uniform prior. Loop. Initialize optimizer/scheduler; train $E$ epochs (forward → CE/Focal → backprop → AdamW step); validate (loss/ACC/Macro F1), keep the best local checkpoint by Macro F1; step the scheduler if enabled. Run the Grad CAM probe on xai_samples and compute Deletion AUC mean/std. Return

$$\omega_i^{(t)}, train\_loss, val\_loss, val\_acc, val\_macro\_F1, xai\_del\_auc\_mean, xai\_del\_auc\_std$$

## Model Aggregation at The Central Server

The server aggregates only model weights and scalars never raw data. It logs round wise accuracy and faithfulness curves, and tracks two checkpoints: Best F1 and (within a small F1 tolerance) a tie break favoring higher, steadier Deletion AUC. Optional sanity checks clip extreme update.

## Explainable Ai for Model Interpretation

We pair Grad-CAM maps with a deletion-style faithfulness score to align accuracy with plausible evidence. On the last conv block (DenseNet for the hybrid),

$$\alpha_k^c = \frac{1}{HW}\sum_{i,j}\frac{\partial s^c}{\partial A_{ij}^k}, \qquad \mathrm{CAM}^c = \mathrm{ReLU}\left(\sum_k \alpha_k^c A^k\right)\ (\rightarrow [0,1]\ \text{overlay})$$

yielding class-focused heatmaps [17]. Mask top-salient pixels; track $p_c(m)$ vs. masked fraction $m$. Deletion AUC

$$\mathrm{AUC}_{\mathrm{del}} \approx \sum_t \frac{p_c(m_{t-1}) + p_c(m_t)}{2}(m_t - m_{t-1})$$

the server logs accuracy and faithfulness and uses the latter to break near-ties.

# Implementation Details

## Data Preprocessing & Distribution

We used a brain magnetic resonance imaging collection of 10,417 images organized into four classes: *glioma_tumor*, *meningioma_tumor*, *pituitary_tumor* and *no_tumor*. Data are federated across four clients with per-class counts shown in Table 1 (column totals: 2547 / 2712 / 2658 / 2500, summing to 10,417). The size of the dataset for each client is approximately 2,600 images, with the residual 17 images assigned to client 4 in the 80/10/10 splits of Table 2 (8333 / 1042 / 1042 for train/val/test).

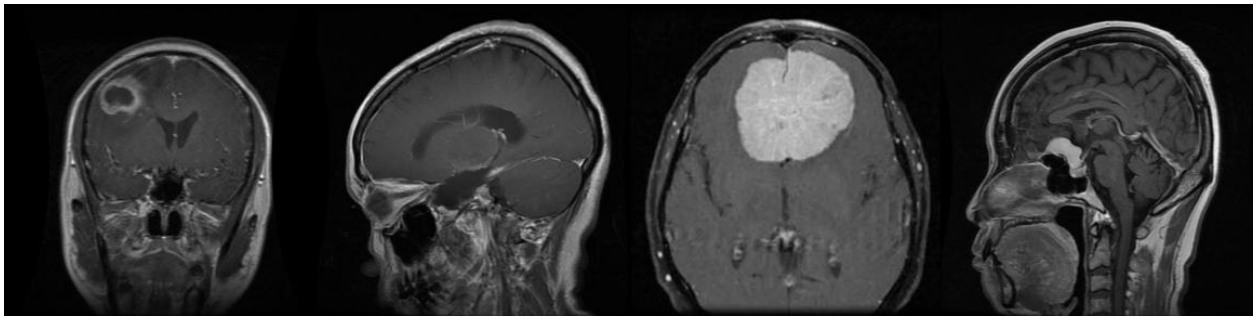
Each image is loaded in grayscale (one channel) and resized to 224 × 224. We apply light, label-preserving cleanup and normalization:

$$\tilde{x} = \frac{x - \mu_i mg}{\sigma_i mg}, \qquad \hat{x} = \frac{\tilde{x} - \min(\tilde{x})}{\max(\tilde{x}) - \min(\tilde{x})} \in [0,1]$$

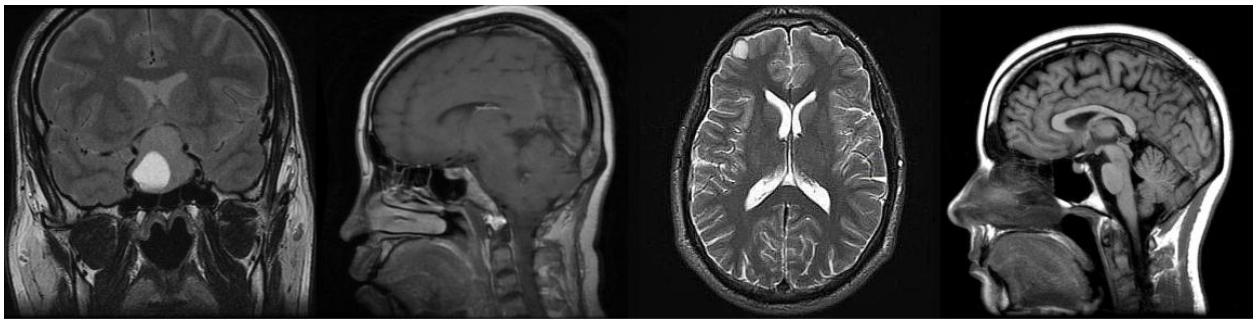
Where $\mu_{img}$, $\sigma_{img}$ are per-image statistics. Validation and test transforms are deterministic (resize → normalize → tensor). Training uses mild augmentations that reflect plausible MRI variability while preserving pathology: small rotations (±5°), translations (≤ 5%), optional horizontal flip, and gentle intensity/contrast jitter. Pipelines are implemented with Albumentations and finalized with tensor conversion.

Each client splits its own local dataset into 80/10/10 train/val/test, patient-wise, to avoid leakage (Table 2). Where feasible, splits are stratified by class to mitigate imbalance (see class distributions in Table 1). Only per-split counts and scalar metrics (loss, ACC, macro-F1) are shared with the server; no images or identifiers are transmitted.

Each split is wrapped in a DataLoader (shuffle for train; deterministic for val/test) with hardware-aware batch size and number of workers. Class weights are computed from the training split to support cost-sensitive losses when needed. Optional visual checks (mini-batch grids) verify that preprocessing, labels, and augmentations behave as expected before federated training. Representative examples of the four classes are shown in (Fig. 2).



(a)Glioma_tumor (b)Glioma_tumor (c)Meningioma_tumor (d)Meningioma_tumor



(e) Pituitary_tumor (f) Pituitary_tumor (g) No_tumor (h) No_tumor

***Figure 2: Class exemplars for brain MRI classification: glioma, meningioma, pituitary, and no_tumor. Slices are preprocessed to grayscale 224×224 for consistency across clients.***

***Table 1: Per-client class distribution for the brain MRI dataset used in federated training. Counts are reported for glioma, meningioma, pituitary, and no_tumor across four clients (total N=10,417 images).***

| Client | Glioma | Meningioma | Pituitary | No_tumor |
|---|---|---|---|---|
| Client-1 | 636 | 677 | 663 | 624 |
| Client-2 | 636 | 677 | 663 | 624 |
| Client-3 | 636 | 677 | 663 | 624 |
| Client-4 | 639 | 681 | 669 | 628 |
| **Totals** | **2547** | **2712** | **2658** | **2500** |

***Table 2: Per-client split of the brain MRI corpus into 80% training, 10% validation, and 10% test sets (patient-wise, stratified where feasible). Totals are balanced to ensure comparable evaluation.***

| Client | Train | Validation | Test |
|---|---|---|---|
| Client-1 | 2080 | 260 | 260 |
| Client-2 | 2080 | 260 | 260 |
| Client-3 | 2080 | 260 | 260 |
| Client-4 | 2093 | 262 | 262 |
| **Totals** | **8333** | **1042** | **1042** |

# Training Configuration

This work fine-tunes convolutional and hybrid backbones on brain MRI slices within a synchronous federated learning (FL) loop. Unless stated otherwise, all clients share a common hyperparameter template; the server can push round-wise overrides (learning rate, local epochs, and loss) to maintain cross-site stability.

On each client, the selected model- mobilenetv3, resnet50, customcnn, shufflenetv2, regnety400, hybrid (Swin-T+DenseNet121), or densenet121 is instantiated and its classifier replaced with a 4-class head (glioma, meningioma, pituitary, no_tumor). All models operate on single-channel (grayscale) MRI inputs at 224 × 224, with a harmonized preprocessing pipeline (resizing, center/zero-padding as needed, per-image normalization) for reproducibility.

We use AdamW with weight decay $1 \times 10^{-4}$ . Unless superseded by the server, the base learning rate is $1 \times 10^{-3}$.
Each round, the server broadcasts a compact fit-config:

- Rounds 1–20: LR = $1 \times 10^{-3}$; local epochs E as configured for the run (default 6).
- Rounds 21–31: switch the loss to Focal Loss (γ = 2) to emphasize hard examples.
- Rounds 32–60: LR → $5 \times 10^{-4}$; local epochs → 4.
- Rounds 61–80: LR → $2\times 10^{-4}$; local epochs 3.
- Rounds > 80: LR → $1 \times 10^{-4}$; local epochs 2.

If no override is sent, clients revert to their local defaults (e.g., LR = $1 \times 10^{-3}$, E = 8). The default criterion is class-weighted Cross-Entropy, which uses weights from each client's training split to reduce class imbalance across {glioma, meningioma, pituitary, no_tumor}. The server can alternatively enable Focal Loss (γ = 2) or Label Smoothing (ε = 0.1). A Reduce on Plateau scheduler (use_scheduler = true) with patience of 10 validation checks complements the server's coarse, round-wise LR annealing.

The default batch size is 16 per client (configurable per run). Training loaders shuffle and drop the last incomplete batch; validation/test loaders are deterministic. Worker counts and pinned memory settings are tuned to host capacity. Training prefers GPU when available; otherwise, CPU with conservative thread limits to avoid oversubscription.

Clients keep their best local weights for recovery but transmit only model parameters and scalar metrics (train/val loss, accuracy, macro F1) to the server. The server records round-wise curves, preserves the best global checkpoint by validation F1, and saves the final global model at the end of training.

# Models Used in Client-Side Training

## Custom CNN

Architecture. A compact four-block convolutional encoder with BatchNorm and ReLU after each 3×3 conv, followed by 2×2 max pooling. Channel widths grow 32→64→128→256. For an input $x \in R^{1\times H\times W}$ (grayscale), spatial size halves at each pool (H,W → $H/2, W/2$) four times, then an adaptive average pool collapses to 1×1, yielding a vector $z \in R^{256}$. Convs use Kaiming initialization suitable for ReLU [20]; BatchNorm stabilizes and speeds training [21].
Grayscale/input adaptation. The network is natively single channel: the first conv is Conv2d(1,32,3×3,pad=1), so CT/MRI slices are consumed without RGB replication.
Training head. A small MLP projects z to K classes:

$$o = W_2 \cdot \delta\big(\text{Dropout}(W_1 z + b_1)\big) + b_2, \qquad p = \text{softmax}(o)$$

with layers 256→128→K and dropout 0.4. We minimize cross-entropy on p (optimizer as in the training config).

Explainability. Computed externally: Grad-CAM on the last conv block to localize evidence [17], and Deletion AUC to quantify faithfulness by measuring the confidence drop as top salient pixels are masked.

## ResNet-50

Architecture. Standard ResNet-50 backbone with the final fully connected layer removed, so the network outputs a pooled 2048-D embedding z from layer4 via global average pooling [22].
Grayscale adaptation. The stem convolution is replaced to accept a single input channel (kernel 3×3, stride 1, padding 3), enabling direct ingestion of CT/MRI slices without RGB replication.
Training head. A regularized MLP projects z to K classes:

$$o = W_3\,\delta\big(\mathrm{BN}\,W_2\,\delta(\mathrm{BN}\,W_1 z)\big) + b, \qquad p = \mathrm{softmax}(o)$$

with layer sizes 2048→512→256→K, BatchNorm and Dropout between linear layers.
Explainability. Grad-CAM on layer4 [17; 23] and Deletion-AUC are computed by external utilities using the backbone's final feature maps.

## Swin-T + DenseNet121

Architecture. A two-branch hybrid that fuses a hierarchical transformer (Swin-T) with a convolutional encoder (DenseNet-121). The Swin branch produces a global, windowed self-attention embedding $z_s win \in \mathbb{R}^{d_s}$[24], while the DenseNet branch yields a convolutional embedding $z_d n \in \mathbb{R}^{d_d}$ via dense connectivity [25]. We concatenate and pass through an MLP with layer sizes $(d_s + d_d) \to 512 \to 128 \to$ K, matching the code's Linear → ReLU → Dropout → Linear → ReLU → Linear stack.

$$z = [z_{\text{swin}}; z_{\text{dn}}], \qquad h_1 = \delta(W_1 z + b_1),$$

$$h_2 = \mathrm{Dropout}\big(\delta(W_2 h_1 + b_2)\big), \qquad o = W_3 h_2 + b_3, \qquad p = \mathrm{softmax}(o)$$

Grayscale/input adaptation. Inputs are grayscale CT/MRI. For the Swin branch, the single channel is replicated to RGB before tokenization. For the DenseNet branch, the first 7×7 conv is converted to 1 input channel and initialized by RGB→gray averaging to preserve pretraining.

$$W_{\text{gray}} = \frac{1}{3}(W_R + W_G + W_B)$$

These steps are implemented directly in the model's constructor and forward (replicate-to-3 for Swin; 1-channel conv with averaged weights for DenseNet).
Training head. The fused vector z feeds the MLP described above; we train with cross-entropy on p. The backbones are optionally frozen via a configuration flag.
Explainability. Computed externally. We use Grad-CAM on the last DenseNet block to localize convolutional evidence [17], and attention rollout on Swin windows to visualize transformer token mixing [26]. Faithfulness is summarized with Deletion AUC by masking top salient pixels and tracking the confidence drop.

## DenseNet-121

Architecture. A densely connected CNN where each layer consumes all prior feature maps via concatenation [25]:

$$x_l = H_l([x\_0, x\_1, \ldots, x_l - 1]$$

The composite functions $H_l$ follow the pattern (BN → ReLU → 1×1 conv → BN → ReLU → 3×3) conv, with transition layers consisting of a 1×1 convolution followed by 2×2 average pooling. Global average pooling produces the embedding z ∈ $\mathbb{R}^C$.
Grayscale adaptation. The stem convolution is converted to one input channel and initialized by RGB→gray averaging to preserve the pretrained structure:

$$W_{\text{gray}} = \frac{1}{3}(W_R + W_G + W_B)$$

Training head. The stock classifier is replaced by a small MLP:

$$o = W_3\,\delta(W_2\,\delta(W_1 z + b_1) + b_2) + b_3, \qquad p = \text{softmax}(o)$$

with layer sizes C→512→128→K and dropout 0.5 after the first hidden layer. We train with cross-entropy on p.
Explainability. Post-hoc Grad-CAM on the last dense block (e.g., denseblock4) is used to highlight pathology-driven regions.

## MobileNetV3-Large

Architecture. Efficient backbone built from inverted residual blocks with depthwise separable 3×3 convolutions and squeeze excitation (SE) in selected blocks [27]; nonlinearity is the hard swish from:

$$h\text{-}swish(x) = x \cdot \frac{\text{ReLU6}(x+3)}{6}$$

We retain the pretrained feature stack and remove the stock classifier, so the backbone outputs a pooled feature vector z.
Grayscale/input adaptation. The stem is converted to single-channel input with a 3×3 conv, stride 2, padding 1, allowing CT/MRI slices to be fed directly without RGB replication.
Training head. A regularized MLP projects z to K classes:

$$o = W_3\,\delta\big(\text{BN}\,W_2\,\delta(\text{BN}\,W_1 z)\big), \qquad p = \text{softmax}(o)$$

with layer sizes in→512→256→K, BatchNorm and Dropout (stronger in early layers) between linear layers. We train with cross-entropy on p.
Explainability. Post-hoc attribution is computed externally: Grad-CAM on the last convolutional stage to localize evidence, and Deletion AUC to quantify faithfulness by measuring the class confidence drop as top salient pixels are masked.

## ShufflenetV2

Architecture. An efficient backbone built from inverted residual units that use depthwise 3×3 convolutions, pointwise 1×1 projections, channel split, and channel shuffle to keep MACs and memory traffic low [28]. Channel shuffle permutes feature channels to mix information across groups, writing the channel vector as $x \in \mathbb{R}^C$, a fixed permutation $P$ yields $\tilde{x} = P\,x$. We keep the pretrained feature stack and set the classifier to identity, so the backbone outputs a pooled feature vector z.

Grayscale/input adaptation. The stem conv is converted to one input channel. To preserve pretraining, its weights are initialized by RGB averaging:

$$W_{\text{gray}} = \frac{1}{3}(W_R + W_G + W_B)$$

so single-channel CT/MRI slices can be used directly.
Training head. A small MLP maps z to K classes:

$$o = W_2\,\delta(\text{BN}\,W_1 z), \qquad p = \text{softmax}(o)$$

with layers C→256→K, BatchNorm before each linear, and Dropout (0.4 before $W_1$ , 0.2 before $W_2$). We train with cross-entropy on p.
Explainability. Computed externally: Grad-CAM on the last ShuffleNetV2 stage to localize evidence, and Deletion AUC to quantify faithfulness by tracking the class confidence drop as top salient pixels are masked.

## RegNetY400

Architecture. RegNetY with grouped bottleneck blocks and squeeze excitation (SE) in every block. Stage widths follow the linear rule [29]:

$$w_i = Q_q(w_0 + w_a i)$$

and global average pooling provides the feature vector z.
Grayscale adaptation. The stem convolution is converted to one input channel and initialized by RGB→gray weight averaging:

$$W_{\text{gray}} = \frac{1}{3}(W_R + W_G + W_B)$$

preserving pretrained structure for single-channel CT/MRI.
Training head. A regularized MLP maps z to K classes with BatchNorm and Dropout:

$$o = W_2\,\delta(\text{BN}\,W_1 z), \qquad p = \text{softmax}(o)$$

using layer sizes C→256→K (where C is the pooled channel width). We optimize cross-entropy with AdamW and mild weight decay.
Explainability. Computed externally: Grad-CAM on the last RegNet stage, and Deletion AUC to quantify the faithfulness of the saliency mask.

# Grad-Cam Implementation

After each local train/eval, every client runs a small Grad-CAM probe on a subset of its validation images (xai_samples, default 16). The probe is entirely client-side: it never shares images or heatmaps, only two scalars (mean and standard deviation of Deletion-AUC) with the server for round-level aggregation.

Target layer & hooks. We pick the last convolutional layer of the current backbone and attach forward/backward hooks to capture activations $A \in \mathbb{R}^{C\times H\times W}$ and their gradients $\partial s^c/\partial A$ with respect to the predicted class logit $s^c$. For class c, Grad-CAM weights are the spatially averaged

gradients [30]:

The class map is the ReLU-weighted sum, followed by min–max normalization to [0,1] for visualization and scoring.

$$\alpha_k^c = \frac{1}{Z}\sum_{i,j}\frac{\partial s^c}{\partial A_{ij}^k}, \quad Z = H \cdot W, \qquad \mathrm{CAM}^c(i,j) = \mathrm{ReLU}\left(\sum_k \alpha_k^c A_{ij}^k\right)$$

Faithfulness metric (Deletion-AUC). To quantify how well the heatmap pinpoints evidence, we follow the deletion protocol: iteratively mask the top-salient pixels and record the target-class probability (or logit) along a schedule $m_t \in [0,1]$ (masked fraction). The Deletion-AUC is the area under the confidence-vs-masking curve, computed via the trapezoid rule:

$$\mathrm{AUC}_{\mathrm{del}} \approx \sum_{t=1}^{T} \frac{p_c(m_{t-1}) + p_c(m_t)}{2} \backslash big(m_t - m_{t-1} \backslash big)$$

where a larger drop in confidence (smaller AUC) indicates a more faithful saliency mask. We report the mean and standard deviation of AUC_del over the probed images.
What the server receives & aggregates. Each client returns xai_del_auc_mean and xai_del_auc_std. The server logs these per round and computes a weighted average across clients (by client sample size), storing round-wise traces for dashboards and final reports.

Outputs. For a few samples ($save\_k$), Grad-CAM overlays are rendered on the original grayscale slice using a heat colormap; these artifacts remain local to the client for qualitative inspection.

# CHAPTER 4

# Experimental Results & Discussion

## Evaluation Metrics

To rigorously evaluate model performance in our 4-class brain MRI setting (glioma, meningioma, pituitary, no_tumor) trained under a synchronous federated learning (FL) setup, we adopt standard multiclass metrics with class-aware aggregation. For multiclass settings with *K* classes, let $(TP_k, FP_k, TN_k, FN_k)$ denote class-wise counts and $n_k$ the number of true samples of class k (so that $N = \sum_{k=1}^{k} n_k$) Per-class precision, recall, and F1 are:

$$P_k = \frac{TP_k}{TP_k + FP_k}, \qquad R_k = \frac{TP_k}{TP_k + FN_k}, \qquad F1_k = \frac{2P_k R_k}{P_k + R_k}.$$

We summarize across classes using:

$$\text{Macro-F1} = \frac{1}{K}\sum_{k=1}^{K} F1_k, \qquad \text{Weighted-F1} = \sum_{k=1}^{K} \frac{n_k}{N} F1_k.$$

## Federated Model Performance

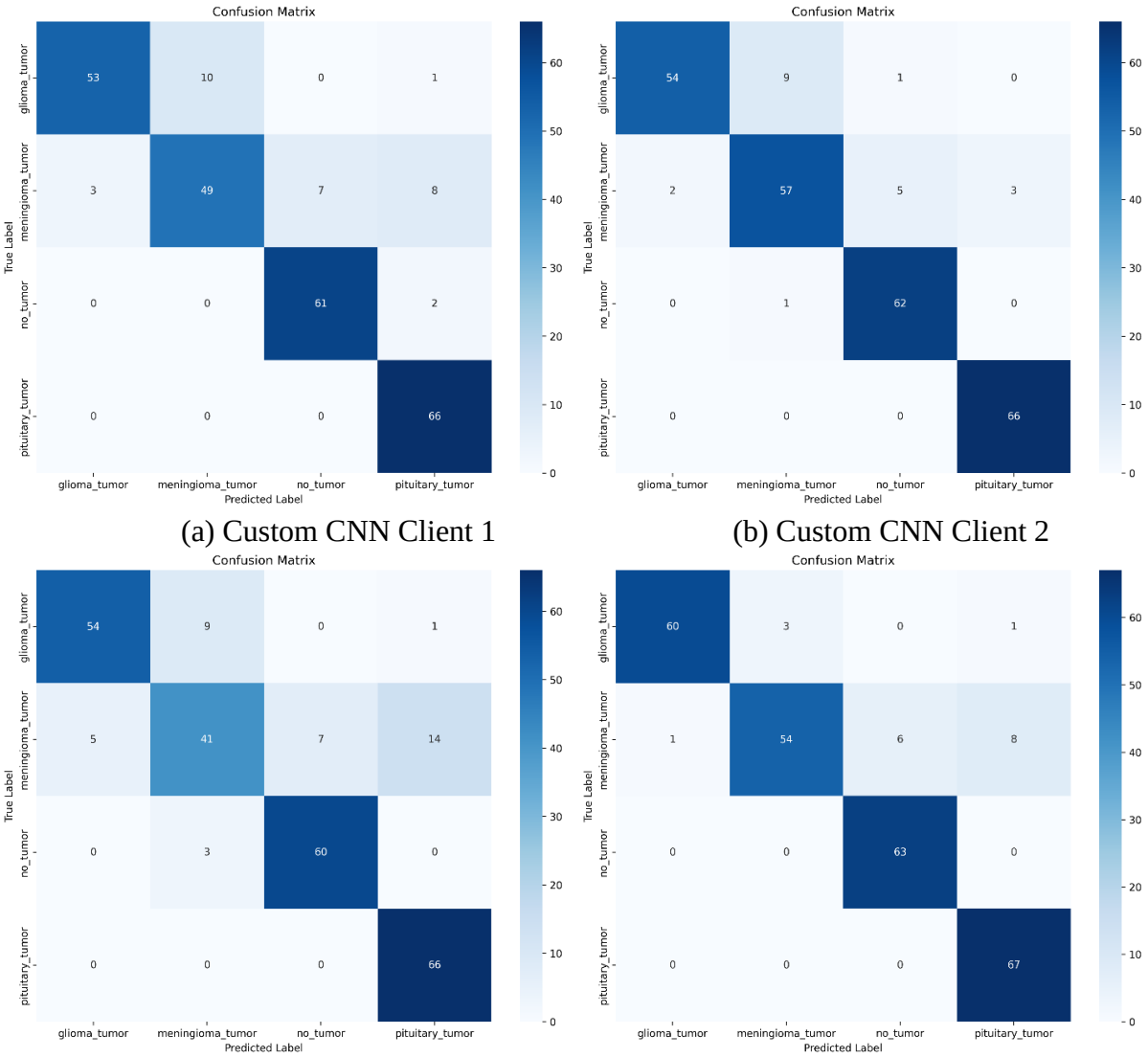
### Custom CNN

In the federated setting, the baseline Custom CNN was trained at the same time on four clients using FedAvg (20 rounds; 2 epochs per round). We chose the optimization and regularization approaches such that the model would always converge, even when the data wasn't IID. As the model trained, the loss kept going down in a smooth, steady pace. At the same time, both accuracy and F1 slowly went up and then leveled off. The validation curves were extremely similar to the training curves, which meant that the model was learning useful patterns from the data instead of just memorizing it or overfitting (Fig. 3, Fig. 4). The model attained a mean client accuracy of 0.8945 ± 0.0360, along with a macro-F1 score of 0.8926 ± 0.0378 and a weighted-F1 score of 0.8915 ± 0.0381 (mean ± SD across clients). Results per site reveal substantial variability: Client 3 marks the lower bound (ACC = 0.8500), while Client 4 achieves the top bound (ACC = 0.9278), a gap of about 7.78 percentage points (Table 3). The transformer CNN hybrid generally converged successfully and functioned effectively across locations, exhibiting minimal sensitivity to variations in client distribution.

***Table 3: Client-wise performance of the Custom CNN in the synchronous FedAvg setup (20 rounds, 2 local epochs/round, 4 clients).***

| Client | ACC | Macro F1 | Weighted F1 |
|---|---|---|---|
| 1 | 0.8808 | 0.8789 | 0.8778 |
| 2 | 0.9192 | 0.9188 | 0.9183 |
| 3 | 0.8500 | 0.8456 | 0.8438 |
| 4 | 0.9278 | 0.9272 | 0.9259 |


***Figure 3: Federated training dynamics of the Custom CNN: loss and accuracy curves, macro-F1, server-side aggregation time, and per-client dataset profiles across 20 FedAvg rounds (2 epochs/round)***




(a) Custom CNN Client 1 (b) Custom CNN Client 2

(c) Custom CNN Client 3 (d) Custom CNN Client 4

***Figure 4: Per-client confusion matrix results for the Custom CNN, illustrating class-level performance on held-out data in the federated brain MRI setting.***

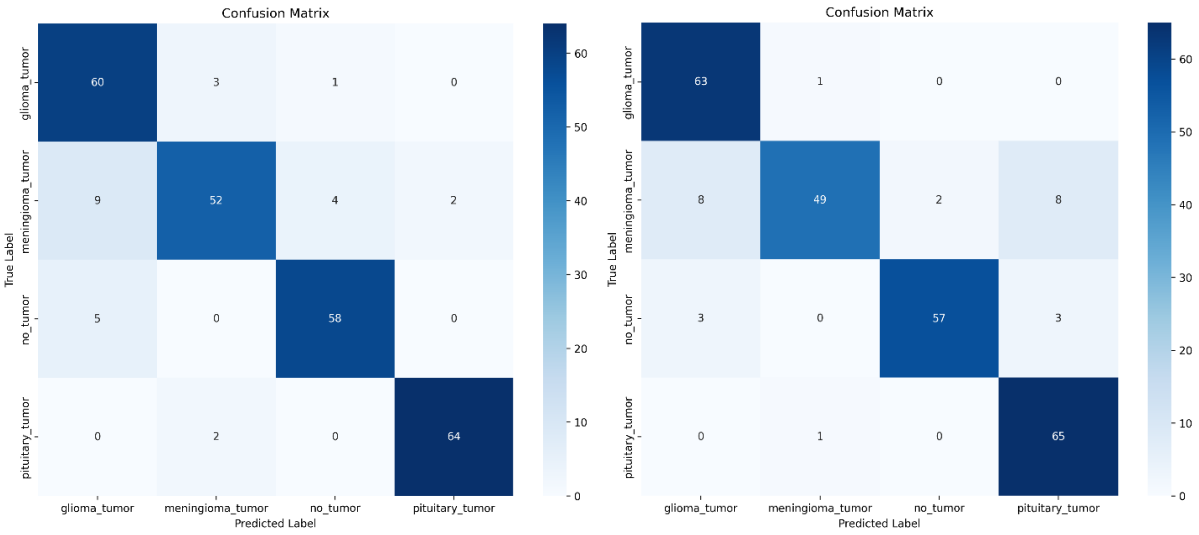
## ResNet-50

In a synchronous FedAvg setup with 20 rounds and 2 local epochs per round, ResNet-50 was trained on data from four non-IID clients. The optimization behaved well: the training loss decreased steadily, while accuracy and F1 scores rose and then stabilized. The validation curves closely followed the training curves, suggesting that regularization was effective and there were no clear signs of overfitting (Fig. 5, Fig. 6). On the held-out test data, the model achieved a mean accuracy of 0.9031 ± 0.0119, a macro-F1 of 0.9020 ± 0.0125, and a weighted-F1 of 0.9011 ± 0.0126 (mean ± SD across clients). The close agreement between accuracy and both F1 scores indicates that the model performs fairly across classes rather than relying heavily on any single majority class. The client-wise dispersion was moderate (ACC range: 0.8923-0.9202; span ~2.79 percentage points; Table 4), showing that the model works well across different local distributions, but there is still opportunity to make it more robust for the hardest site. In general, the residual architecture converged stably in the federated context and showed reliable mid-90s performance trends in the learning curves. However, there is room for targeted improvements (such client-aware augmentation or adaptive aggregation) to close the gap between clients.

***Table 4: Client-wise performance of ResNet-50 in the synchronous FedAvg setup (20 rounds, 2 local epochs/round, 4 clients).***

| Client | ACC | Macro F1 | Weighted F1 |
|---|---|---|---|
| 1 | 0.9000 | 0.8984 | 0.8976 |
| 2 | 0.9000 | 0.8997 | 0.8994 |
| 3 | 0.8923 | 0.8902 | 0.8888 |
| 4 | 0.9202 | 0.9197 | 0.9187 |


***Figure 5: Federated training dynamics of ResNet-50: loss and accuracy curves, macro-F1, server-side aggregation time, and per-client data profiles across 20 FedAvg rounds (2 epochs/round).***




(a) ResNet-50 Client 1 (b) ResNet-50 Client 2

(c) ResNet-50 Client 3 (d) ResNet-50 Client 4

***Figure 6: Per-client confusion matrix results for ResNet-50, illustrating class-level performance on held-out data in the federated brain MRI setting.***

## Swin-T + DenseNet121

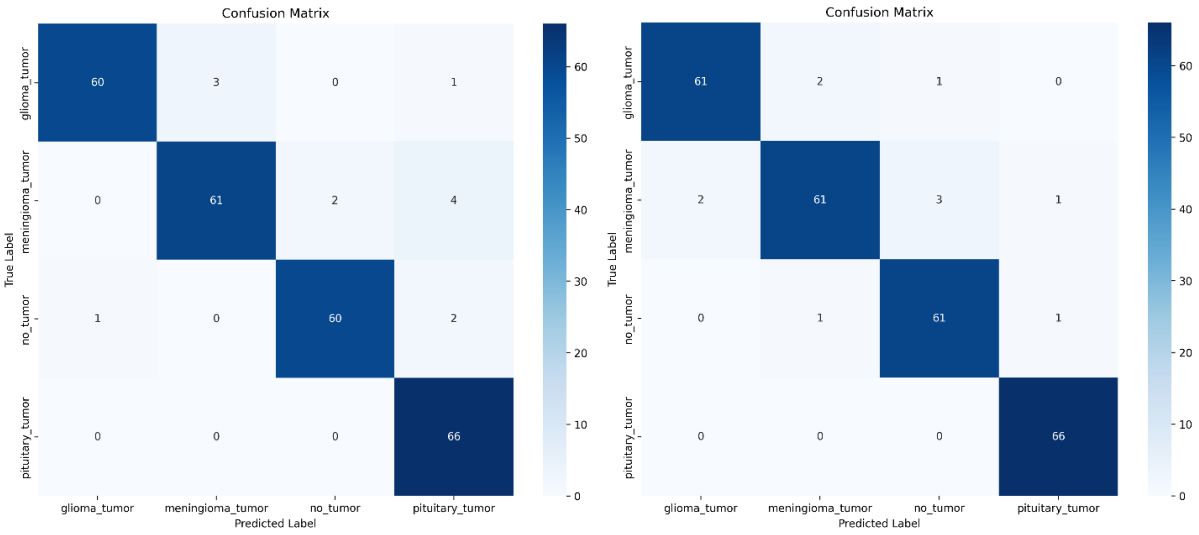
We trained the hybrid Swin-T+DenseNet121 model on four non-IID clients using synchronous FedAvg with 20 communication rounds and 2 local epochs per round. The optimization worked as planned: the training loss went down steadily, and the accuracy and F1 curves went up and then leveled off. The validation curves stayed very close to the training curves, suggesting that our regularization was effective, and that overfitting was not a major issue (Fig. 7, Fig. 8). On the held-out test data, the model achieved a mean ACC of 0.9578 ± 0.0048, a macro-F1 of 0.9580 ± 0.0049, and a weighted-F1 of 0.9578 ± 0.0048 (mean ± SD across clients). Performance variation between clients remained small (ACC range: 0.9500–0.9620; span ≈ 1.20 percentage points; Table 5), and the close alignment between accuracy and both F1 scores suggests balanced behavior across classes rather than reliance on a single dominant class. This stability is consistent with the design of the hybrid: Swin-T's windowed self-attention captures long-range contextual information, while DenseNet's dense connectivity encourages feature reuse and strengthens local representations. Together, these properties support robust generalization in a federated, heterogeneous setting, and the transformer–CNN hybrid converged reliably and maintained strong performance even as the number of participating

clients changed.

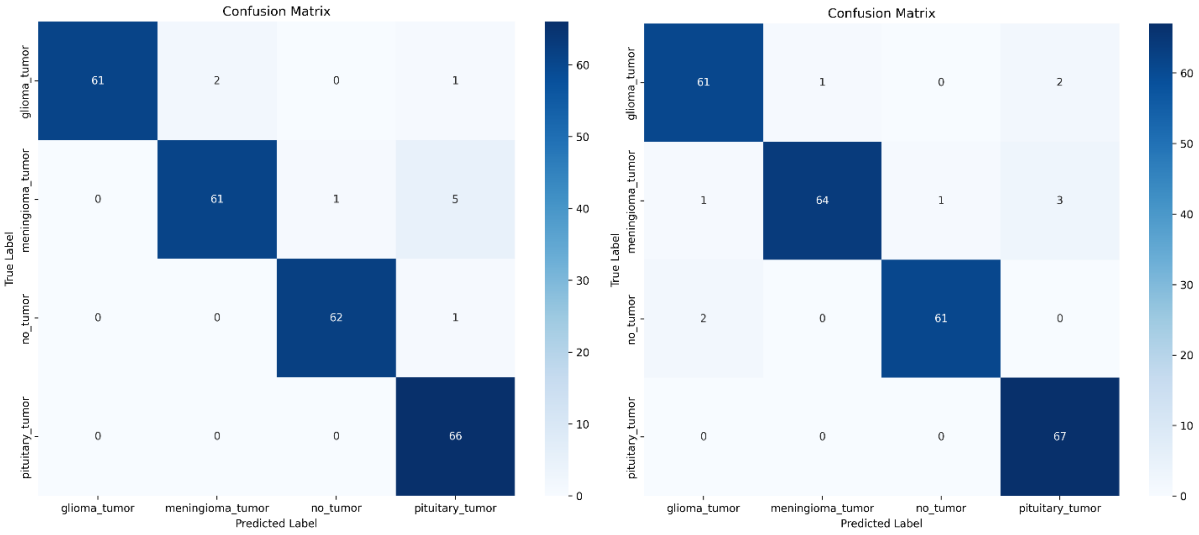
***Table 5: Client-wise performance of Swin-T + DenseNet121 in the synchronous FedAvg setup (20 rounds, 2 local epochs/round, 4 clients).***

| Client | ACC | Macro F1 | Weighted F1 |
|---|---|---|---|
| 1 | 0.9500 | 0.9502 | 0.9500 |
| 2 | 0.9577 | 0.9575 | 0.9575 |
| 3 | 0.9615 | 0.9621 | 0.9616 |
| 4 | 0.9620 | 0.9621 | 0.9619 |


***Figure 7: Federated training dynamics of the Swin-T+DenseNet121 hybrid: loss and accuracy curves, macro-F1, server-side aggregation time, and per-client dataset profiles across 20 FedAvg rounds (2 epochs/round).***




(a) Swin-T + DenseNet121 Client 1 (b) Swin-T + DenseNet121 Client 2




(c) Swin-T + DenseNet121 Client 3 (d) Swin-T + DenseNet121 Client 4

***Figure 8: Per-client confusion matrix results for the Swin-T + DenseNet121 hybrid, illustrating class-level performance on held-out data in the federated brain MRI setting.***

## DenseNet-121

We trained DenseNet-121 on four non-IID clients using a synchronous FedAvg setup with 20 rounds and 2 epochs per round. The optimization process was successful. The training loss decreased, the accuracy and F1 curves increased and then leveled off, and the validation trajectories closely followed the training ones. This pattern indicates that the regularization strategy was effective, and that overfitting was not a major concern (Fig. 9, Fig. 10). The model got a mean ACC of 0.9645 ± 0.0074, a macro-F1 of 0.9648 ± 0.0074, and a weighted-F1 of 0.9644 ± 0.0076 (mean ± SD across customers) when it was kept out for evaluation. The dispersion among clients was minimal (ACC range: 0.9538-0.9696; span ∼1.58 percentage points; Table 6), and the close correlation between ACC and both F1 variants indicates a balanced per-class performance rather than dependence on a singular class. In general, DenseNet-121 converged reliably in the federated regime and consistently gave high cross-site performance when the data was spread out across different sites.

***Table 6: Client-wise performance of DenseNet-121 in the synchronous FedAvg setup (20 rounds, 2 local epochs/round, 4 clients).***

| Client | ACC | Macro F1 | Weighted F1 |
|---|---|---|---|
| 1 | 0.9692 | 0.9695 | 0.9692 |
| 2 | 0.9654 | 0.9656 | 0.9654 |
| 3 | 0.9538 | 0.9540 | 0.9534 |
| 4 | 0.9696 | 0.9699 | 0.9696 |

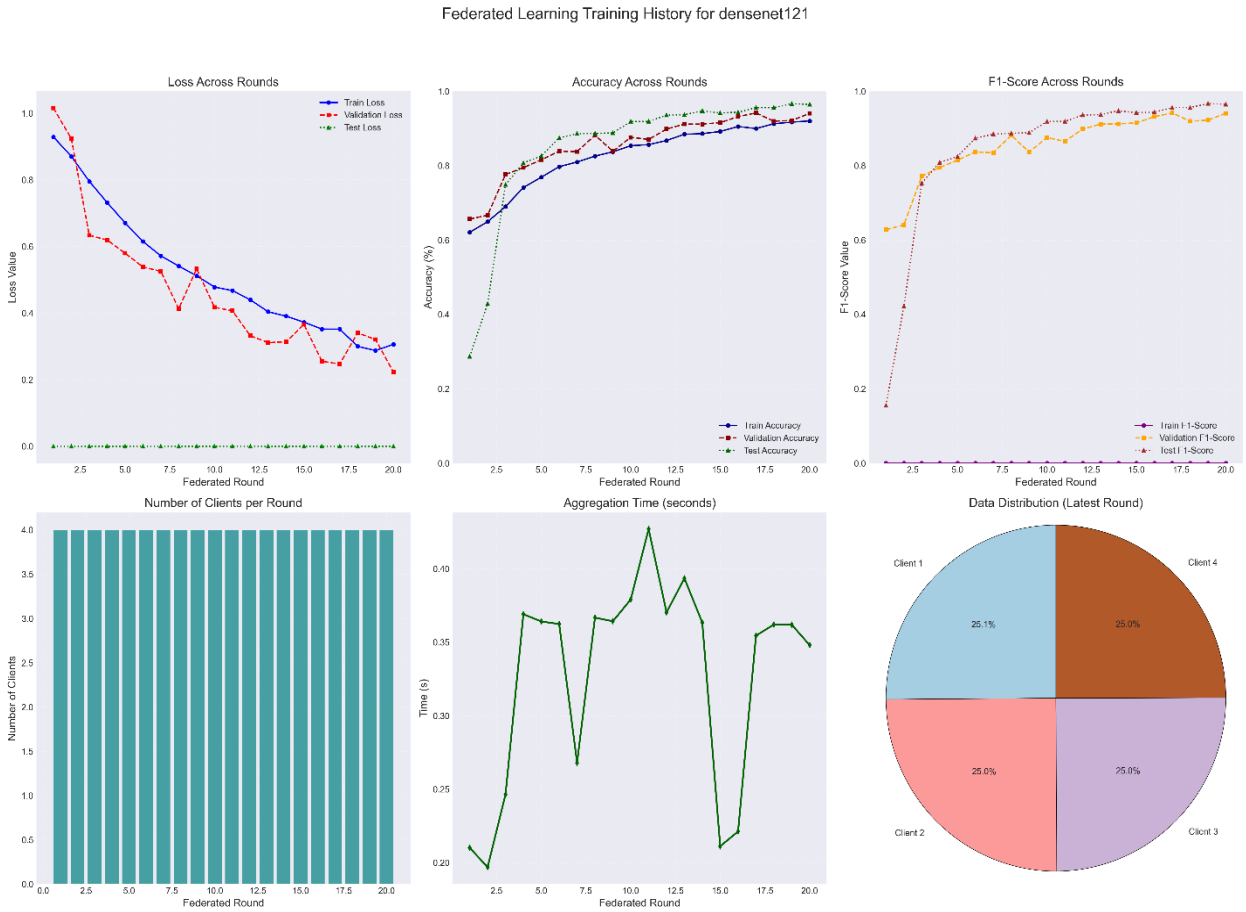
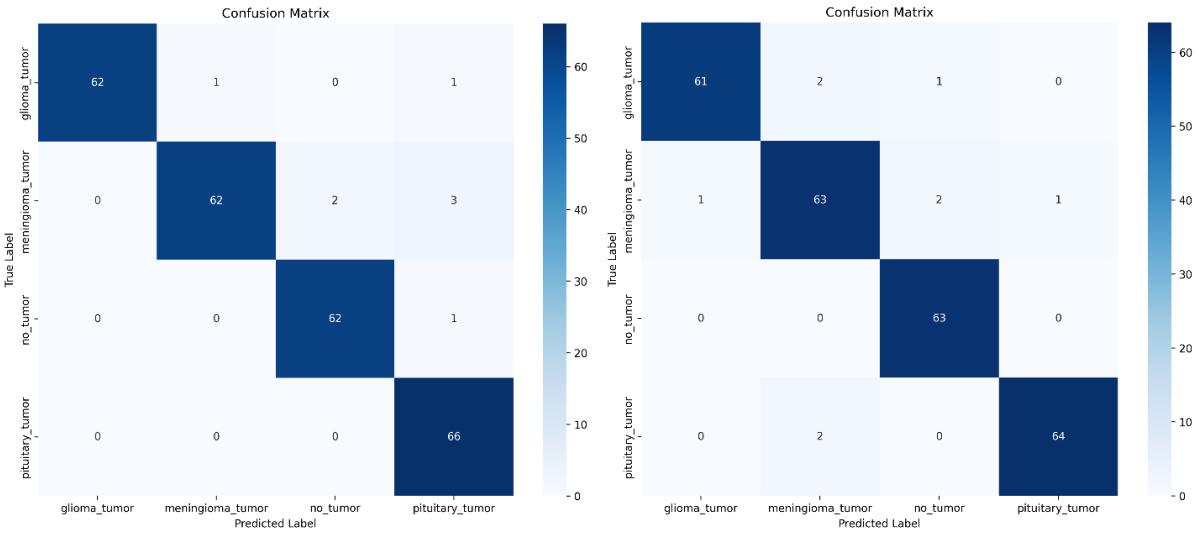



***Figure 9: Federated training dynamics of DenseNet-121: loss and accuracy curves, macro-F1, server-side aggregation time, and perclient dataset profiles across 20 FedAvg rounds (2 epochs/round).***




(a) DenseNet121 Client 1 (b) DenseNet121 Client 2

(c) DenseNet121 Client 3 (d) DenseNet121 Client 4

***Figure 10: Per-client confusion matrix results for DenseNet-121, illustrating class-level performance on held-out data in the federated brain MRI setting.***

## MobileNetV3-Large

Using synchronous FedAvg with 20 rounds and 2 local epochs per round, MobileNetV3-Large was trained on four non-IID clients. The optimization was stable: the training loss kept going down, and the accuracy and F1 scores kept going up until they hit a plateau. The validation curves were very similar to the training curves, which means that the regularization strategy worked and there wasn't much overfitting (Fig. 11, Fig. 12). The model got a mean ACC of 0.9732 ± 0.0063, a macro-F1 of 0.9734 ± 0.0061, and a weighted-F1 of 0.9733±0.0062 (mean±SD across clients). The dispersion among clients was minimal (ACC range: 0.9654-0.9808; span ∼1.54 percentage points; Table 7), and the close correlation between accuracy and both F1 variants indicates a balanced per-class performance rather than dependence on a singular dominant class. In the federated regime, MobileNetV3-Large converged reliably and performed well and consistently over a wide range of local distributions.

***Table 7: Client-wise performance of MobileNetV3-Large in the synchronous FedAvg setup (20 rounds, 2 local epochs/round, 4 clients).***

| Client | ACC | Macro F1 | Weighted F1 |
|---|---|---|---|
| 1 | 0.9731 | 0.9735 | 0.9733 |
| 2 | 0.9808 | 0.9808 | 0.9808 |
| 3 | 0.9654 | 0.9659 | 0.9656 |
| 4 | 0.9734 | 0.9736 | 0.9734 |

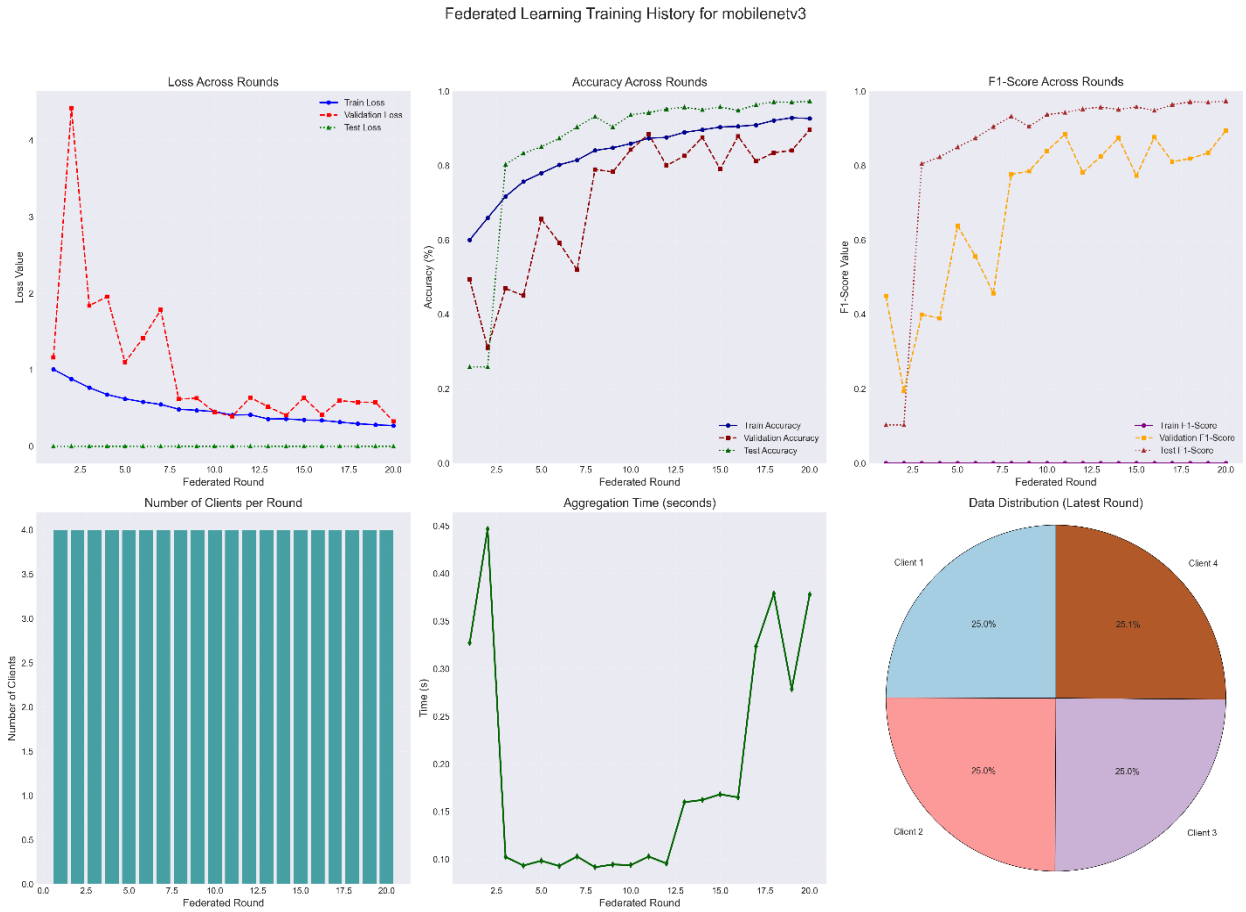
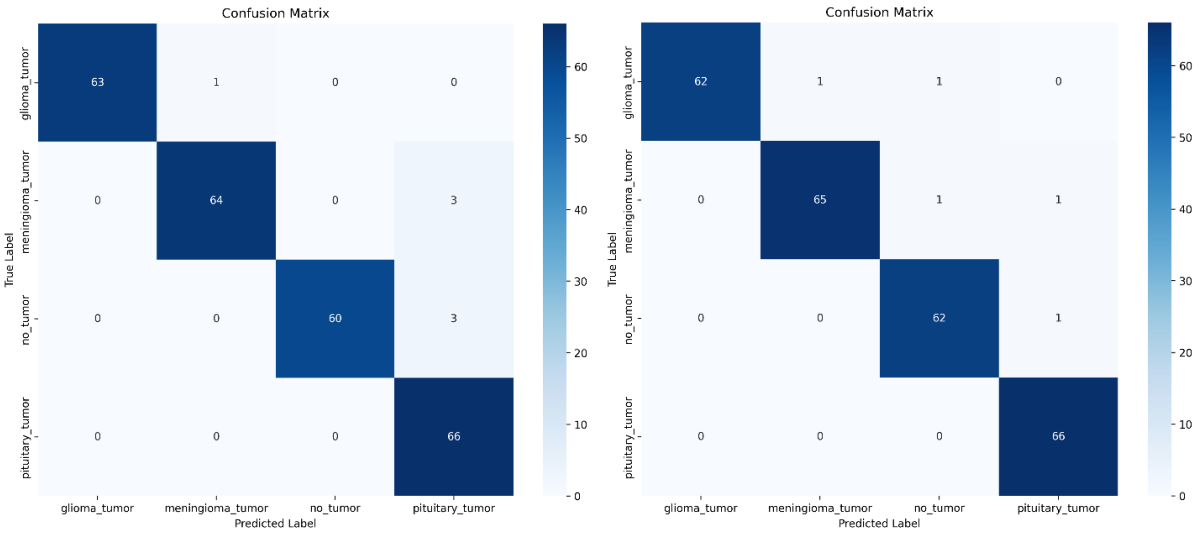
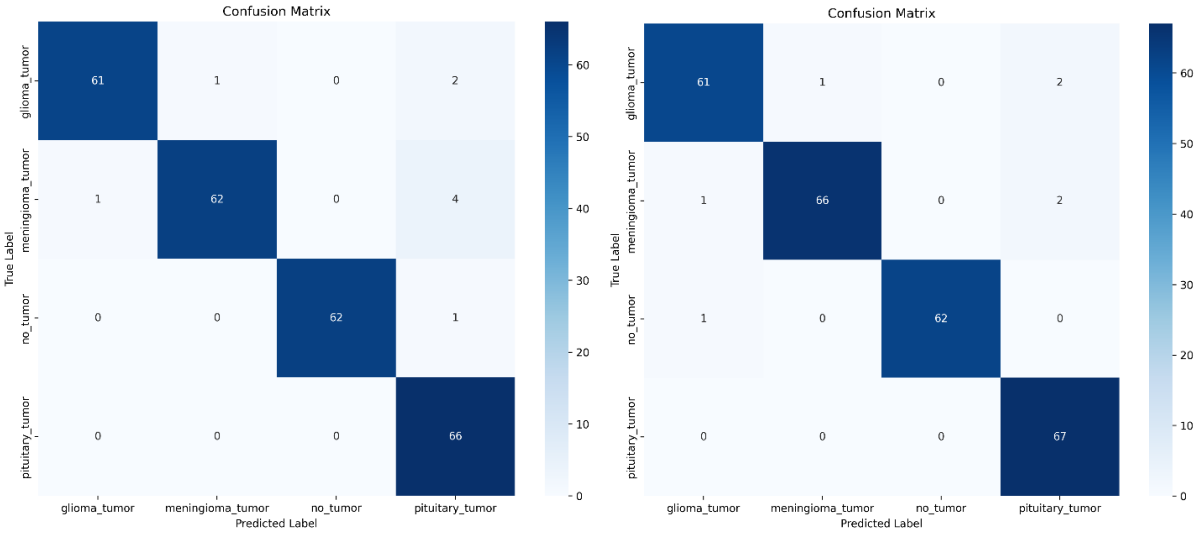



***Figure 11: Federated training dynamics of MobileNetV3-Large: loss and accuracy curves, macro-F1, server-side aggregation time, and per-client dataset profiles across 20 FedAvg rounds (2 epochs/round).***




(a) MobileNetV3-Large Client 1 (b) MobileNetV3-Large Client 2




(c) MobileNetV3-Large Client 3 (d) MobileNetV3-Large Client 4

***Figure 12: Per-client confusion matrix results for MobileNetV3- Large, illustrating class-level performance on held-out data in the federated brain MRI setting.***

## ShufflenetV2

ShufflenetV2 was trained on four non-IID clients in a synchronous FedAvg setup with 20 rounds and 2 epochs per round. The optimization process went well: losses went down consistently, while accuracy and F1 curves went up and then leveled off. Validation closely followed training evidence of good regularization and minimized overfitting (Fig. 13, Fig. 14). In held-out testing, the model reached a mean accuracy (ACC) of 0.9789 ± 0.0067, a macro-F1 of 0.9791 ± 0.0067, and a weighted-F1 of 0.9789 ± 0.0068 (mean ± SD across clients). The variation between clients was small (ACC range: 0.9692–0.9848; span ≈ 1.56 percentage points; Table 8), and the close match between ACC and both F1 scores suggests that the model treats all classes fairly instead of leaning on a single dominant class. The training curves also show smooth aggregation and consistent improvements over rounds, which fits well with ShuffleNetV2's lightweight design (depthwise separable convolutions and channel shuffle) that is well-suited to federated settings with bandwidth and compute constraints. Overall, ShuffleNetV2 converged reliably and delivered strong performance across sites with differing local data distributions.

***Table 8: Client-wise performance of ShuffleNetV2 in the synchronous FedAvg setup (20 rounds, 2 local epochs/round, 4 clients).***

| Client | ACC | Macro F1 | Weighted F1 |
|---|---|---|---|
| 1 | 0.9692 | 0.9694 | 0.9691 |
| 2 | 0.9808 | 0.9808 | 0.9808 |
| 3 | 0.9808 | 0.9810 | 0.9808 |
| 4 | 0.9848 | 0.9851 | 0.9848 |

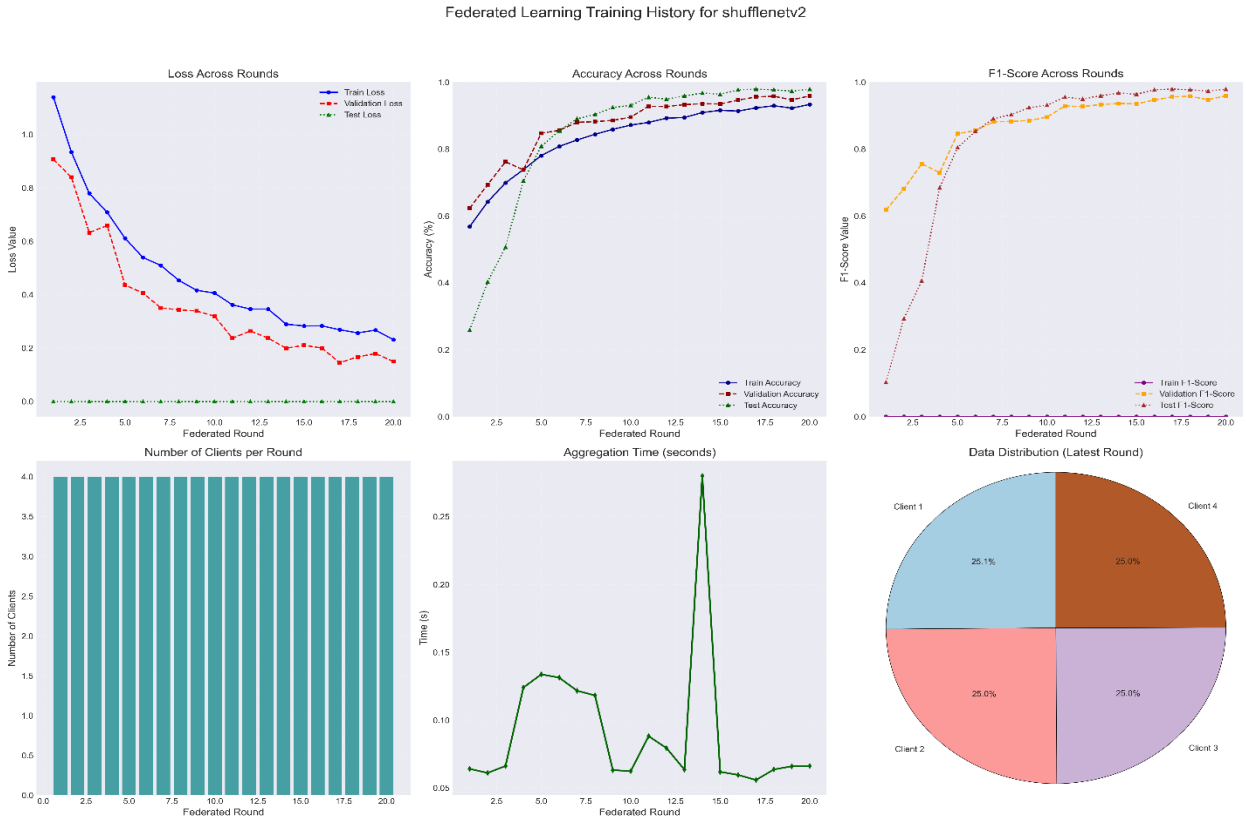
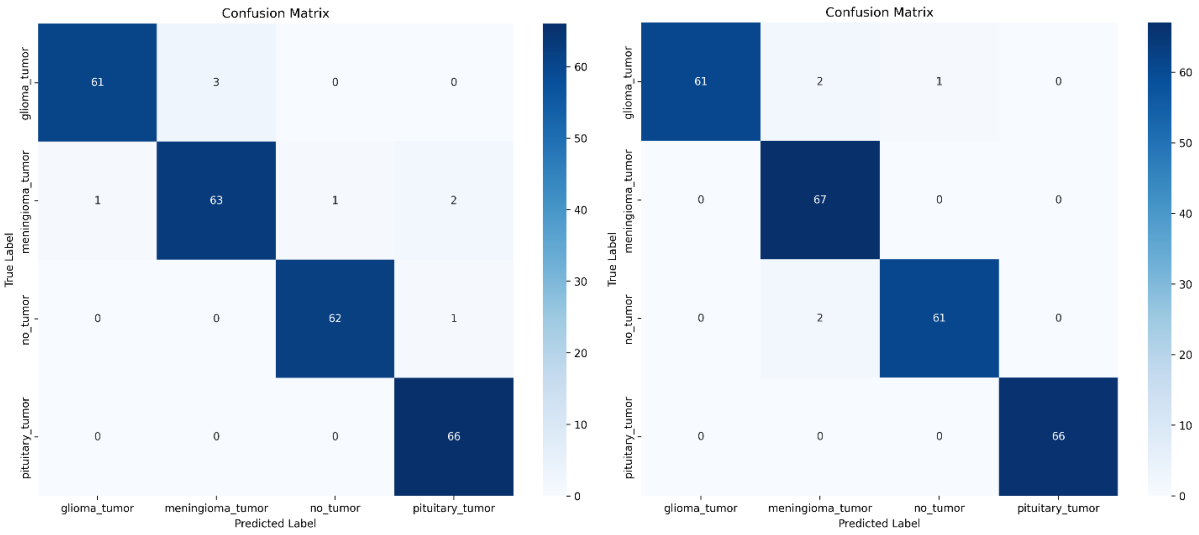
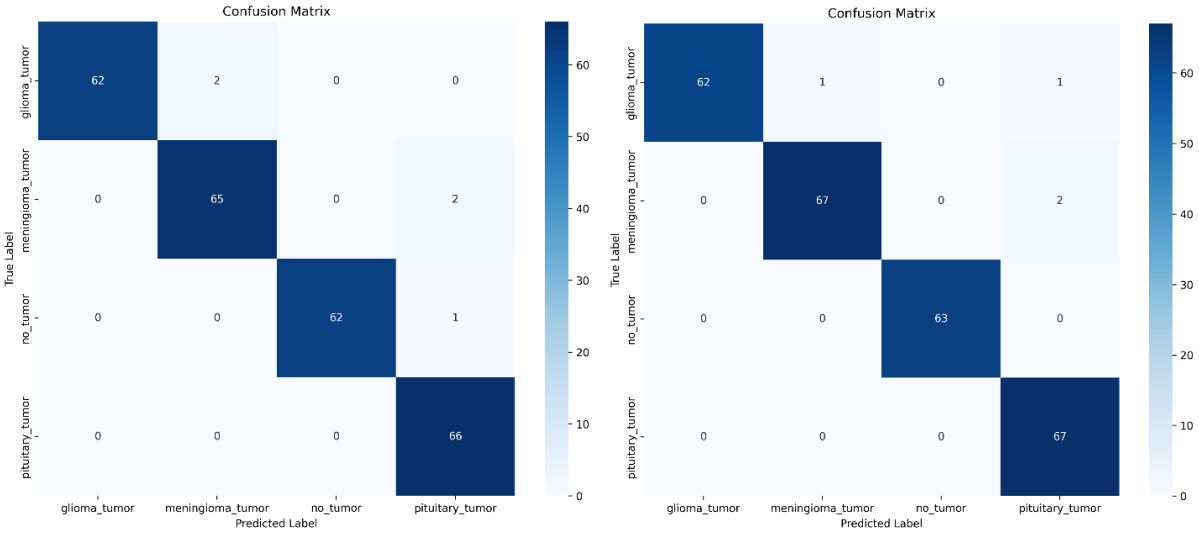



***Figure 13: Federated training dynamics of ShuffleNetV2: loss and accuracy curves, macro-F1, server-side aggregation time, and perclient dataset profiles across 20 FedAvg rounds (2 epochs/round).***




(a) ShuffleNetV2 Client 1 (b) ShuffleNetV2 Client 2




(c) ShuffleNetV2 Client 3 (d) ShuffleNetV2 Client 4

***Figure 14: Per-client confusion matrix results for ShuffleNetV2, illustrating class-level performance on held-out data in the federated brain MRI setting.***

## RegNetY400

In a synchronous FedAvg setup with 20 rounds and 2 local epochs per round, RegNetY400 was trained on four non-IID clients. The optimization behaved well: training loss decreased steadily, while accuracy and F1 scores increased and then plateaued. The validation curves closely tracked the training curves, indicating that regularization was effective, and overfitting was minimal (Fig. 15, Fig. 16). On held-out data, the model reached a mean accuracy of 0.9828 ± 0.0058, a macro-F1 of 0.9827 ± 0.0058, and a weighted-F1 of 0.9827 ± 0.0058 (mean ± SD across clients). Client-wise variation was very small (ACC range: 0.9769–0.9923; span ≈ 1.54 percentage points; Table 9), and the tight alignment between ACC and both F1 variants suggests balanced performance across all classes rather than reliance on a dominant class. The smooth training traces fit well with RegNetY's architectural design regularized stage widths and depths, bottleneck blocks, and squeeze-and-excitation modules which together promote efficient capacity usage and strong generalization under heterogeneous local data. Overall, RegNetY400 converged reliably and delivered excellent cross-site performance in the federated setting.

***Table 9: Client-wise performance of RegNetY400 in the synchronous FedAvg setup (20 rounds, 2 local epochs/round, 4 clients).***

| Client | ACC | Macro F1 | Weighted F1 |
|---|---|---|---|
| 1 | 0.9923 | 0.9924 | 0.9923 |
| 2 | 0.9808 | 0.9807 | 0.9808 |
| 3 | 0.9769 | 0.9769 | 0.9768 |
| 4 | 0.9810 | 0.9809 | 0.9810 |

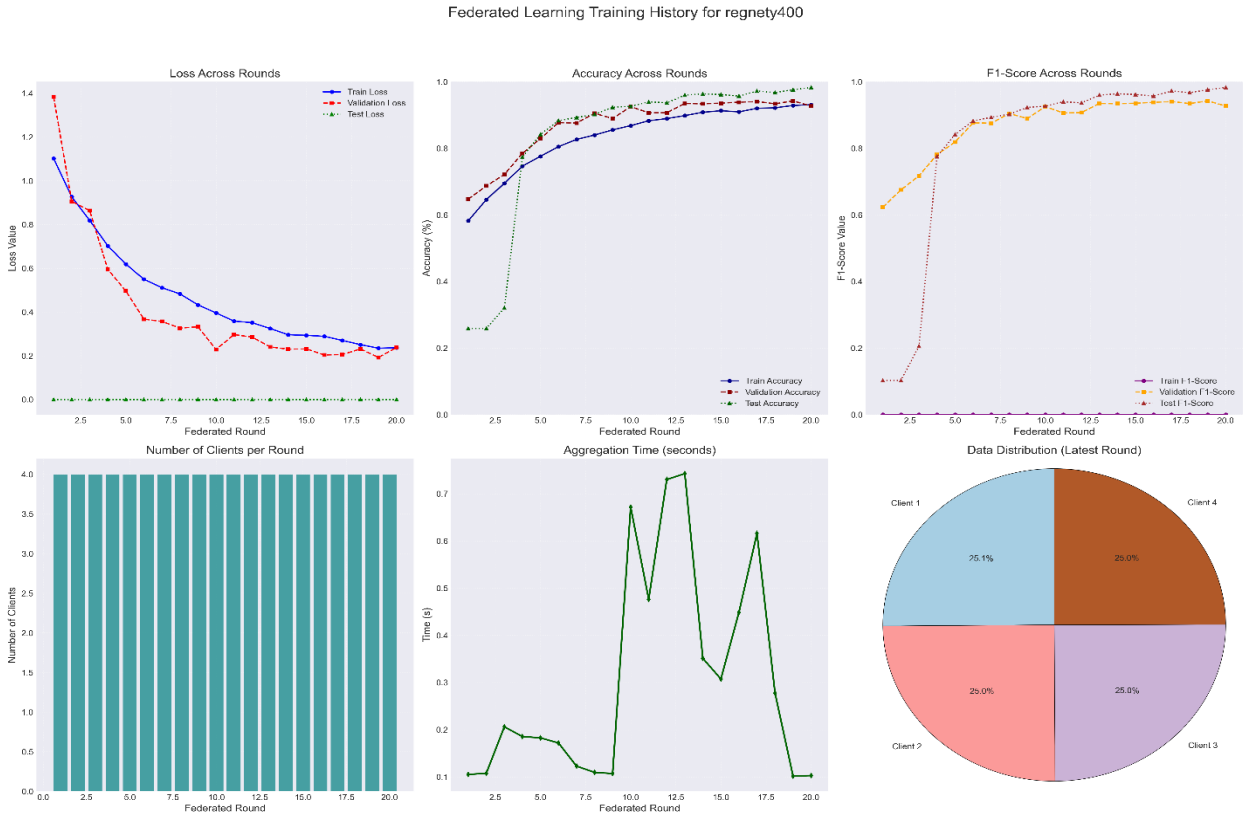
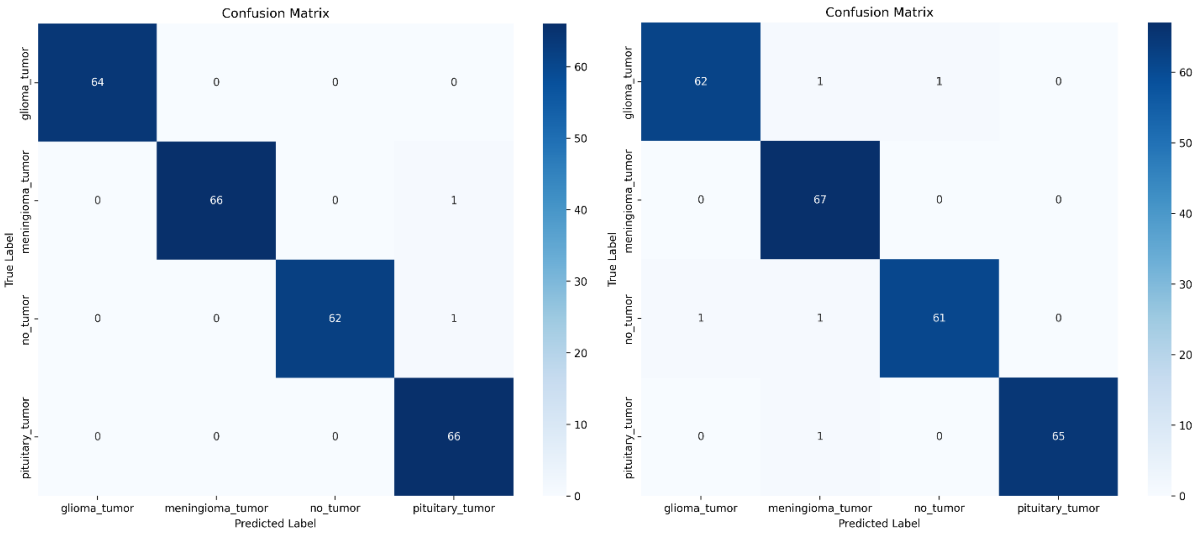
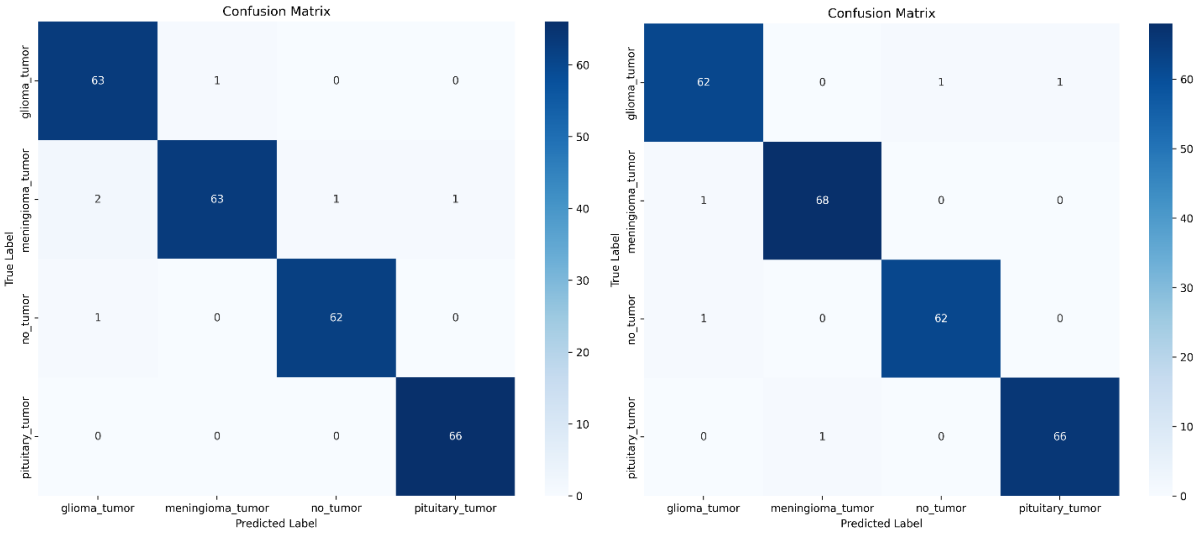



***Figure 15: Federated training dynamics of RegNetY400: loss and accuracy curves, macro-F1, server-side aggregation time, and perclient dataset profiles across 20 FedAvg rounds (2 epochs/round).***




(a) RegNetY400 Client 1 (b) RegNetY400 Client 2




(c) RegNetY400 Client 3 (d) RegNetY400 Client 4

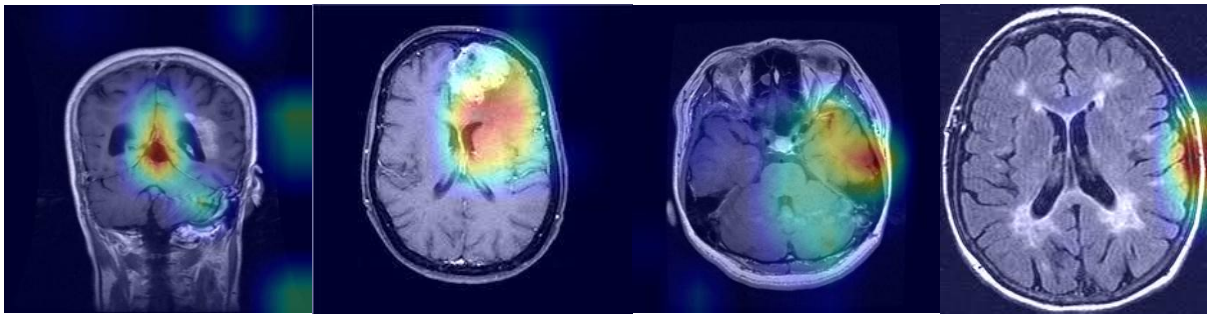
***Figure 16: Per-client confusion matrix results for RegNetY400, illustrating class-level performance on held-out data in the federated brain MRI setting.***

# Explainability Evaluation

We evaluated explanation faithfulness by employing Deletion AUC over rounds and incorporating Grad-CAM overlays. As demonstrated in Table 12, ShuffleNetV2 had the lowest Deletion AUC mean (0.38 at R18) and the highest Val F1 (0.9592, R20). This means that its optimal checkpoint is both accurate and true to informative pixels. With a DelAUC of 0.41 and a Val F1 of 0.9424, RegNetY400 finished in second. The mid-tier models, Swin-T+DenseNet121 and DenseNet-121, had moderate DelAUC peaks (0.47 and 0.55) and thereafter stayed at stable validation F1 plateaus. The MobileNetV3-Large and ResNet-50 were close (0.73 and 0.85), but the Custom CNN was behind (0.89 at R16), which makes sense because it has less power and peaks later in F1. In general, a higher Val F1 score was linked to a higher Deletion AUC score near convergence. This suggests that models that generalize better also create saliency that is more faithful.

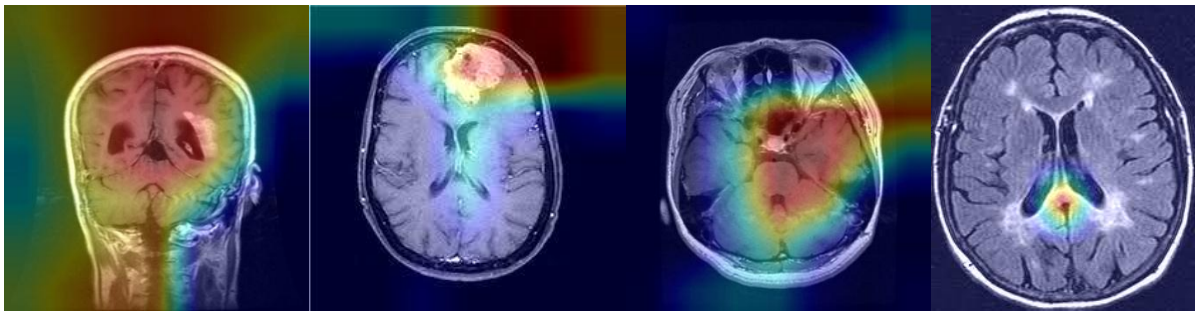
Grad-CAMs back up these tendencies in a qualitative way. In glioma, meningioma, and pituitary, heatmaps for ShuffleNetV2 (Fig. 17) are compact and aligned with lesions. However, in no_tumor, they stay low and diffuse, which cuts down on false positives. RegNetY400 (Fig. 18) exhibits comparably concentrated attributions with marginally expanded coverage at lesion

peripheries, aligning with its robust validation scores and near-optimal fidelity. The quantitative Deletion AUC peaks and the class-wise Grad-CAMs show that the top models not only do an excellent job of sorting things into groups, but they also use data that makes sense from an anatomical point of view to come to their findings.



a) Glioma_tumor (b) Meningioma_tumor (c) Pituitary_tumor (d) No_tumor

***Figure 17: Grad-CAM visualizations for ShuffleNetV2 on representative brain MRI slices, highlighting class-discriminative regions for glioma, meningioma, pituitary, and no_tumor.***



a) Glioma_tumor (b) Meningioma_tumor (c) Pituitary_tumor (d) No_tumor

***Figure 18: Grad-CAM visualizations for RegNetY400 on representative brain MRI slices, highlighting class-discriminative regions for glioma, meningioma, pituitary, and no_tumor.***

# Interpretation Of Results

To identify the most effective architecture under our federated learning (FL) setup, we evaluated seven backbones: Custom CNN, ResNet-50, Swin-T + DenseNet121, DenseNet-121, MobileNetV3-Large, ShuffleNetV2, and RegNetY400. Unless otherwise stated, we used synchronous FedAvg for 20 rounds with 2 local epochs per round. We tracked client-side means ± SD (Table 11), round-wise validation peaks and held-out Test ACC (Table 12), and explanation faithfulness via Deletion AUC (Table 10).

Client-side ranking (means ± SD). Across clients, RegNetY400 ranks first with the highest mean ACC/Macro-F1/Weighted-F1 (0.9828 ± 0.0066, 0.9827 ± 0.0067, 0.9827 ± 0.0067), closely followed by ShuffleNetV2 (0.9789 ± 0.0067 ACC). MobileNetV3-Large is third (0.9732 ± 0.0063), while DenseNet-121 and the hybrid Swin-T + DenseNet121 form a strong middle tier (0.9645 ± 0.0074 and 0.9578 ± 0.0055 ACC, respectively). ResNet-50 and Custom CNN trail but remain stable (0.9031 ± 0.0119 and 0.8944 ± 0.0360 ACC). Notably, the smallest dispersion is observed for the hybrid Swin-T + DenseNet121 (ACC SD = 0.0055), indicating particularly consistent cross-site behavior (Table 11).

Best checkpoints and held-out generalization. When selecting the best round per model, ShuffleNetV2 attains the strongest validation maxima (Val ACC/F1 = 0.9587/0.9592 at R20)

and competitive generalization (Test ACC = 0.9789). RegNetY400 peaks slightly lower on validation (Val ACC/F1 = 0.9424/0.9424 at R19) but delivers the best held-out performance (Test ACC = 0.9827). MobileNetV3-Large has a significant gap between its modest validation peaks (Val F1 = 0.8940 at R20) and its high-Test ACC (0.9732), which suggests that it is very robust at evaluation time even though its validation gains are conservative. Both DenseNet-121 (Val F1 = 0.9414 at R17; Test ACC = 0.9664) and Swin-T + DenseNet121 (Val F1 = 0.9241 at R20; Test ACC = 0.9645) are suitable options. ResNet-50 and Custom CNN improve over time, with their best scores at R19 and R18, respectively, and Test ACC scores of 0.9223 and 0.9003 (Table 12). Together, these results highlight that parameter-efficient families (RegNetY/ShuffleNet/MobileNet) provide the best blend of optimization stability and cross-site generalization, while the Swin-T hybrid yields low client-to-client variance.

Explainability faithfulness . Deletion AUC rankings mirror the accuracy hierarchy at the top: ShuffleNetV2 is first (mean 0.38 at R18) with the highest co-occurring validation F1 (0.9592 at R20), and RegNetY400 is second (mean 0.41 at R20; Val F1 = 0.9424 at R19). Mid-tier Hybrid Swin-T + DenseNet-121 and DenseNet-121 follow (means 0.47 and 0.55), while ResNet-50 and MobileNetV3-Large are comparable (means 0.85 and 0.73); CustomCNN lags (mean 0.89). Overall, we observe a positive coupling between generalization and faithfulness near the best rounds (Val F1 ↑ aligns with DelAUC ↓; Table 10).

- RegNetY400 has the best held-out accuracy, and the best client-side means, while ShuffleNetV2 has the most accurate and highest validation solution at convergence.
- Lightweight, bandwidth-friendly backbones (ShuffleNet/RegNet/MobileNet) excel under FL constraints.
- The hybrid Swin-T + DenseNet121 has the lowest cross-client dispersion, which is helpful for sites with heterogeneous clients.
- Lower Deletion AUC tends to co-occur with higher validation F1, supporting the reliability of model attributions at the best checkpoints.

For clarity, Tables 10–12 jointly report

- client-side means ± SD,
- best per-round checkpoints and Test ACC, and
- explanation faithfulness with peak round indices.

***Table 10: Round-wise explainability summary using Deletion AUC. The table lists, per model, the lowest observed Deletion AUC mean (with std) and the highest validation F1, with the corresponding round in parentheses.***

| Model | xai_del_auc_mean | xai_del_auc_std | Validation F1 |
|---|---|---|---|
| Custom CNN | 0.89 (R16) | 0.88 (R20) | 0.8667 (R18) |
| ResNet-50 | 0.85 (R18) | 0.81 (R19) | 0.8839 (R19) |
| MobileNetV3-Large | 0.73 (R19) | 0.64 (R18) | 0.8940 (R20) |
| Swin-T + DenseNet121 | 0.47 (R20) | 0.49 (R18) | 0.9241 (R20) |
| DenseNet-121 | 0.55 (R20) | 0.45 (R17) | 0.9414 (R17) |
| RegNetY400 | 0.41 (R20) | 0.48 (R19) | 0.9424 (R19) |
| **ShuffleNetV2** | **0.38 (R18)** | **0.31 (R18)** | **0.9592 (R20)** |

***Table 11: Client-side performance under synchronous FedAvg (20 rounds; 2 epochs/round). Entries report mean ± SD across the four clients.***

| Model | ACC | Macro F1 | Weighted F1 |
|---|---|---|---|
| Custom CNN | 0.8944 ± 0.0360 | 0.8926 ± 0.0378 | 0.8914 ± 0.0381 |
| ResNet-50 | 0.9031 ± 0.0119 | 0.9020 ± 0.0125 | 0.9011 ± 0.0126 |
| Swin-T + DenseNet121 | 0.9578 ± 0.0055 | 0.9580 ± 0.0056 | 0.9577 ± 0.0055 |
| DenseNet-121 | 0.9645 ± 0.0074 | 0.9647 ± 0.0074 | 0.9644 ± 0.0076 |
| MobileNetV3-Large | 0.9732 ± 0.0063 | 0.9734 ± 0.0061 | 0.9733 ± 0.0062 |
| ShufflenetV2 | 0.9789 ± 0.0067 | 0.9791 ± 0.0067 | 0.9789 ± 0.0068 |
| **RegNetY400** | **0.9828 ± 0.0066** | **0.9827 ± 0.0067** | **0.9827 ± 0.0067** |

***Table 12: Best of round FL performance across all backbones. For validation metrics, the peak round is given in parentheses.***

| Model | ACC | Validation ACC | Validation F1 | Test ACC |
|---|---|---|---|---|
| Custom CNN | 0.8611 | 0.8674 (R18) | 0.8667 (R18) | 0.9003 |
| ResNet-50 | 0.8625 | 0.8838 (R19) | 0.8839 (R19) | 0.9223 |
| Swin-T + DenseNet121 | 0.9056 | 0.9241 (R20) | 0.9241 (R20) | 0.9645 |
| DenseNet-121 | 0.9201 | 0.9414 (R17) | 0.9414 (R17) | 0.9664 |
| MobileNetV3-Large | 0.9287 | 0.8962 (R20) | 0.8940 (R20) | 0.9732 |
| ShufflenetV2 | **0.9332** | **0.9587** (R20) | **0.9592** (R20) | 0.9789 |
| **RegNetY400** | 0.9309 | 0.9424 (R19) | 0.9424 (R19) | **0.9827** |

# CHAPTER 5
# Conclusion

## Impact & Relevance

Clinical imaging requires both strict privacy across hospitals and clear, transparent decision-making. Our architecture addresses this by offering synchronous federated learning (FL) for four-class brain MRI classification together with a locally computed XAI signal that is monitored globally.
Instead of treating saliency as a cosmetic, after-the-fact picture, we rely on a quantitative faithfulness measure (Deletion AUC). Each client computes round-wise faithfulness summaries and sends only these scalars to the server. The server monitors trust just like accuracy, aligns checkpoint selection with both signals, and thus enables transparency-aware training.

- Trust telemetry at scale: Per-round faithfulness logs provide a continuous "trust trace," helping to flag potential issues such as spurious reasoning, data drift, or client instability.

- Better model selection: When Macro-F1 is similar across checkpoints, we prefer those with higher and more stable faithfulness, reducing the risk of "right answer, wrong reason" behavior.

- Fit for heterogeneous sites: Parameter-efficient backbones (RegNetY, ShuffleNet, MobileNet) perform well on both accuracy and faithfulness, while the Swin-T hybrid exhibits low cross-client variance, which is valuable in non-identical hospital settings.

- Auditable and private: Only model weights and scalar metrics are shared with the server; raw images and heatmaps never leave the clients. This creates an auditable record that can support governance and compliance without compromising patient privacy.

In summary, quantifying and federating faithfulness elevates XAI from a visualization tool to a first-class training signal for multi-site MRI: accurate, privacy-preserving, and ready to support real clinical decision-making.

## Limitations & Future Work

### Limitations

- 2D inputs only. We use single channel 224×224 slices, omitting 3D context and multi sequence MRI (T1/T1c/T2/FLAIR); slice labels may blur lesion boundaries and miss case level heterogeneity.

- Mild federation. Experiments use few clients with controlled non-IID splits; real hospitals face stronger domain shift, variable participation, and concept drift, so robustness may not fully transfer.

- Privacy/security. No secure aggregation, differential privacy, or gradient leakage

defenses; model updates and XAI stats could leak information under adversaries.

- XAI scope. Grad-CAM at a single target layer with a fixed deletion-AUC protocol; saliency depends on layer choice and masking schedule, and crosspaper comparability is limited.

- No personalization. A single global model via FedAvg; personalized FL (e.g., local heads/adapters) or multi-objective aggregation might better balance accuracy and faithfulness across sites.

## Future Work

- 3D context & sequences. Move to 2.5D/3D models and multi-sequence MRI (T1/T1c/T2/FLAIR) with sequence-aware normalization to improve boundary sensitivity and voxel-level outputs.

- Personalization & weighting. Add lightweight site-specific adapters/LoRA or last-layer tuning; use drift-aware client weighting and semi/asynchronous FL to reduce stragglers.

- Privacy & efficiency. Integrate update compression (quantization/sketches), secure aggregation, and optional differential privacy; quantify privacy utility trade-offs for both accuracy and faithfulness.

- Broader XAI. Benchmark Integrated Gradients, RISE, attention rollout, and deletion/insertion protocols with layer/scale sweeps; include sanity checks and counterfactual probes.

In summary, our study demonstrates that light, efficient backbones perform strongly under FL and that a federated faithfulness signal is both feasible and informative. Addressing the above limitations especially multi-sequence 3D modeling, privacy-preserving training, standardized XAI evaluation, and trust-aware aggregation will be key steps toward clinically robust, auditable deployments.

# Conclusion

We demonstrated that it is possible to train four-class brain tumor MRI in a synchronous FL system with 10,417 greyscale slices across four clients while maintaining privacy and ensuring trust. Under a unified head and harmonized preprocessing, both parameter-efficient CNNs (ShuffleNetV2, RegNetY400, MobileNetV3-Large) and deeper/hybrid backbones (ResNet-50, DenseNet-121, Hybrid Swin-T + DenseNet-121, MLP) train stably with FedAvg. The Hybrid Swin-T + DenseNet-121 demonstrated low cross-client variance, whereas RegNetY400 produced the best held-out accuracy, followed closely by ShuffleNetV2 and DenseNet121. Importantly, we found that peaks in validation F1 tended to correspond with increased fidelity, suggesting alignment of accuracy and reasoning, when we coupled Grad-CAM++ with a federated deletion-style faith-fulness score. Without disclosing pictures or heatmaps, auditable round-wise curves were produced by sharing only weights and scalars. Explainability is improved from post-hoc graphics to a first-class quantitative signal that is utilized to break checkpoint ties, protect privacy, and increase selection reliability. Standardize perturbation based XAI protocols, investigate lightweight personalization/asynchronous FL, add safe aggregation and differential privacy, and expand to multi-sequence/3D MRI. Further

examination of the influence of federated faithfulness on perceived plausibility and decision support can be conducted via prospective, multi-site reader research.

# References

[1] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), pp. 1273–1282, 2017. Introduces Federated Averaging (FedAvg) for on-device, privacy-aware training.

[2] G. Kaissis, M. Makowski, D. Rückert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," Nature Machine Intelligence, vol. 2, no. 6, pp. 305–311, 2020. Overview of FL opportunities and constraints in medical imaging.

[3] M. J. Sheller, G. A. Reina, B. Edwards, J. Martin, and S. Bakas, "Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data," Scientific Reports, vol. 10, p. 12598, 2020. Demonstrates FL feasibility across institutions for imaging tasks.

[4] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in IEEE International Conference on Computer Vision (ICCV), pp. 618–626, 2017. Widely used method for class-discriminative saliency maps.

[5] V. Petsiuk, A. Das, and K. Saenko, "RISE: Randomized input sampling for explanation of black-box models," in British Machine Vision Conference (BMVC), 2018. Perturbation-based evaluation underpinning deletion/insert metrics.

[6] W. Li, F. Milletarì, D. Xu, N. Rieke, J. Hancox, W. Zhu, M. Baust, Y. Cheng, S. Ourselin, M. J. Cardoso, and A. Feng, "Privacy-preserving federated brain tumour segmentation," in Machine Learning in Medical Imaging (MLMI 2019), Lecture Notes in Computer Science, vol. 11861, pp. 133–141, Springer, 2019. Early demonstration of federated learning for BraTS brain tumour MRI segmentation with differential privacy.

[7] E. Gundogan, "A novel hybrid deep learning model enhanced with explainable AI for brain tumor multi-classification from MRI images," Applied Sciences, vol. 15, no. 10, p. 5412, 2025. Hybrid CNN+XAI approach; multi-class MRI classification.

[8] K. M. Adnan, T. Ghazal, et al., "Deep learning driven interpretable and informed decision making model for brain tumour prediction using explainable AI," Scientific Reports, vol. 15, no. 1, p. 3358, 2025. Integrates XAI with deep models; employs perturbation-style evaluation.

[9] T. Gomez, T. Fréour, and H. Mouchère, "Metrics for saliency map evaluation of deep learning explanation methods," in Pattern Recognition and Artificial Intelligence (ICPRAI 2022), Lecture Notes in Computer Science, vol. 13363, pp. 84–95, Springer, 2022. Critically analyzes Deletion/Insertion AUC and proposes additional metrics for saliency faithfulness.

[10] E. Albalawi, T. R. Mahesh, A. Thakur, V. V. Kumar, M. Gupta, S. B. Khan, and A. Almusharraf, "Integrated approach of federated learning with transfer learning for classification and diagnosis of brain tumor," BMC Medical Imaging, vol. 24, no. 1, p. 110,

2024. Federated VGG16-based classifier for multi-client brain tumor MRI with high accuracy and privacy preservation.

[11] M. Nahiduzzaman, L. F. Abdulrazak, H. B. Kibria, A. Khandakar, M. A. Ayari, M. F. Ahamed, M. Ahsan, J. Haider, M. A. Moni, M. Kowalski, et al., "A hybrid explainable model based on advanced machine learning and deep learning models for classifying brain tumors using MRI images," Scientific Reports, vol. 15, no. 1, p. 8424, 2025. Hybrid XAI framework for brain tumor classification from MRI.

[12] Z. Li and O. Dib, "Empowering brain tumor diagnosis through explainable deep learning," Machine Learning and Knowledge Extraction, vol. 6, no. 4, pp. 2248–2281, 2024. Survey and framework showing how XAI techniques improve trust in brain tumor diagnosis models.

[13] M. A. A. Khandaker et al., "Transforming brain cancer diagnosis with explainable AI," arXiv preprint arXiv:2501.05426, 2025. Background study on combining DL with XAI for tumor detection.

[14] (Frontiers in Oncology), "Explainable AI in medical imaging: an interpretable and collaborative federated learning model for brain tumor classification," Frontiers in Oncology, vol. 15, p. 1535478, 2025. Interpretable, collaborative FL pipeline for brain tumor MRI.

[15] Anonymous, "Enhancing transparency and trust in brain tumor diagnosis: An in-depth analysis of deep learning and explainable AI techniques," Preprint (Research Square/ResearchGate), 2025. Survey of ML/DL and XAI techniques for brain tumor diagnosis.

[16] Anonymous, "Deep learning-based automated system for enhanced brain tumor detection and early diagnosis using MRI images," Journal of Neonatal Surgery, vol. 14, no. 1, p. 2642, 2025. DL system for brain tumor detection; includes XAI aspects.

[17] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," International Journal of Computer Vision, vol. 128, no. 2, pp. 336–359, 2020. Gradient-weighted Class Activation Mapping for post-hoc CNN explanations.

[18] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," International Conference on Learning Representations (ICLR), 2019. AdamW optimizer with decoupled weight decay.

[19] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2980–2988, 2017. Introduces Focal Loss for class imbalance and hard-example emphasis.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 1026–1034, 2015. Introduces He (Kaiming) initialization for ReLU networks.

[21] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in Proceedings of the 32nd International Conference on Machine Learning (ICML), pp. 448–456, 2015. Widely used normalization layer that stabilizes

deep CNN training.

[22] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016. Introduces residual networks (ResNets) enabling very deep CNN training.

[23] M. Musthafa, T. R. Mahesh, V. Vinoth Kumar, and S. Guluwadi, “Enhancing brain tumor detection in MRI images through explainable AI using Grad-CAM with ResNet-50,” BMC Medical Imaging, vol. 24, no. 1, p. 107, 2024. ResNet-50 classifier for brain tumor MRI with Grad-CAM-based visual explanations.

[24] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, “Swin transformer: Hierarchical vision transformer using shifted windows,” in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 10012–10022, 2021. Hierarchical transformer with shifted windows (Swin-T variant).

[25] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4700–4708, 2017. Introduces DenseNet with dense connectivity and transition layers.

[26] S. Abnar and W. Zuidema, “Quantifying attention flow in transformers,” in Proceedings of ACL, pp. 4190–4197, 2020. Attention rollout to visualize token mixing in transformers.

[27] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, “Searching for MobileNetV3,” in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 1314–1324, 2019. Introduces MobileNetV3 with inverted residuals, SE, and hard-swish; includes Large/Small variants.

[28] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, “ShuffleNet V2: Practical guidelines for efficient CNN architecture design,” in Proceedings of the European Conference on Computer Vision (ECCV), pp. 122–138, 2018. Introduces ShuffleNetV2 with channel split and channel shuffle for efficiency.

[29] I. Radosavovic, R. P. Kosaraju, R. Girshick, K. He, and P. Dollár, “Designing network design spaces,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10428–10436, 2020. Introduces the RegNet family via a simple, quantized linear width rule.

[30] F. Yan, Y. Chen, Y. Xia, Z. Wang, and R. Xiao, “An explainable brain tumor detection framework for MRI analysis,” Applied Sciences, vol. 13, no. 6, p. 3438, 2023. End-to-end framework combining brain tumor detection with explainable AI on MRI data.

221-35-1022

ORIGINALITY REPORT

| 16% | 12% | 11% | 10% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | Submitted to Daffodil International University<br>Student Paper | 3% |
| 2 | arxiv.org<br>Internet Source | 1% |
| 3 | Submitted to University of Ulster<br>Student Paper | <1% |
| 4 | Takuto Koyama, Hiroaki Iwata, Ryosuke Kojima, Takao Otsuka et al. "Empowering Federated Learning for Robust Compound-Protein Interaction Prediction across Heterogeneous Cross-Pharma Domains", American Chemical Society (ACS), 2025<br>Publication | <1% |
| 5 | dspace.daffodilvarsity.edu.bd:8080<br>Internet Source | <1% |
| 6 | Submitted to Universiti Malaysia Pahang<br>Student Paper | <1% |
| 7 | github.com<br>Internet Source | <1% |
| 8 | Somayeh Abbasabadi, Parviz Fattahi, Mahdyeh Shiri. "Exploring transfer learning techniques for classifying Alzheimer's disease with rs-fMRI", Computers in Biology and Medicine, 2025<br>Publication | <1% |
| 9 | library.oapen.org<br>Internet Source | <1% |

| | | |
|---|---|---|
| 10 | Submitted to Addis Ababa University<br>Student Paper | <1 % |
| 11 | Submitted to University of Glasgow<br>Student Paper | <1 % |
| 12 | secured-project.eu<br>Internet Source | <1 % |
| 13 | www.hindawi.com<br>Internet Source | <1 % |
| 14 | iris.unitn.it<br>Internet Source | <1 % |
| 15 | d-nb.info<br>Internet Source | <1 % |
| 16 | Wu, Ying. "Deep Learning-Based MRI Analysis for Brain Tumor Classification", The George Washington University<br>Publication | <1 % |
| 17 | ijeecs.iaescore.com<br>Internet Source | <1 % |
| 18 | Submitted to Cornell University<br>Student Paper | <1 % |
| 19 | Submitted to Yeditepe University<br>Student Paper | <1 % |
| 20 | scholarspace.manoa.hawaii.edu<br>Internet Source | <1 % |
| 21 | discovery.ucl.ac.uk<br>Internet Source | <1 % |
| 22 | downloads.hindawi.com<br>Internet Source | <1 % |
| 23 | spiral.imperial.ac.uk<br>Internet Source | <1 % |
| | ijisrt.com | |

| | | |
|---|---|---|
| 24 | Internet Source | <1 % |
| 25 | theses.hal.science<br>Internet Source | <1 % |
| 26 | Submitted to University of Technology<br>Student Paper | <1 % |
| 27 | sistemasi.ftik.unisi.ac.id<br>Internet Source | <1 % |
| 28 | Nada Farhan. "PRIVACY-ENHANCED FEDERATED LEARNING FOR BRAIN TUMOR CLASSIFICATION: AN EVALUATION OF ACCURACY AND PRIVACY", International Journal of Applied Mathematics, 2025<br>Publication | <1 % |
| 29 | Arvind Dagur, Sohit Agarwal, Dhirendra Kumar Shukla, Shabir Ali, Sandhya Sharma. "Artificial Intelligence and Sustainable Innovation - Volume 3", CRC Press, 2026<br>Publication | <1 % |
| 30 | Submitted to islamicuniversity<br>Student Paper | <1 % |
| 31 | Jyotismita Chaki, Marcin Wozniak. "Brain tumor categorization and retrieval using Deep Brain Incep Res Architecture based Reinforcement Learning Network", IEEE Access, 2023<br>Publication | <1 % |
| 32 | www.researchsquare.com<br>Internet Source | <1 % |
| 33 | jneonatalsurg.com<br>Internet Source | <1 % |
| 34 | Submitted to National University of Technology (NUTECH)-Islamabad | <1 % |

# Masrafe Bin Hannan Siam

## 221-35-1022

Quick Submit

Quick Submit

Daffodil International University

### Document Details

Submission ID

trn:oid:::1:3449032109

Submission Date

Dec 20, 2025, 4:34 PM GMT+6

Download Date

Dec 20, 2025, 4:40 PM GMT+6

File Name

221-35-1022.pdf

File Size

2.7 MB

56 Pages

16,203 Words

101,820 Characters



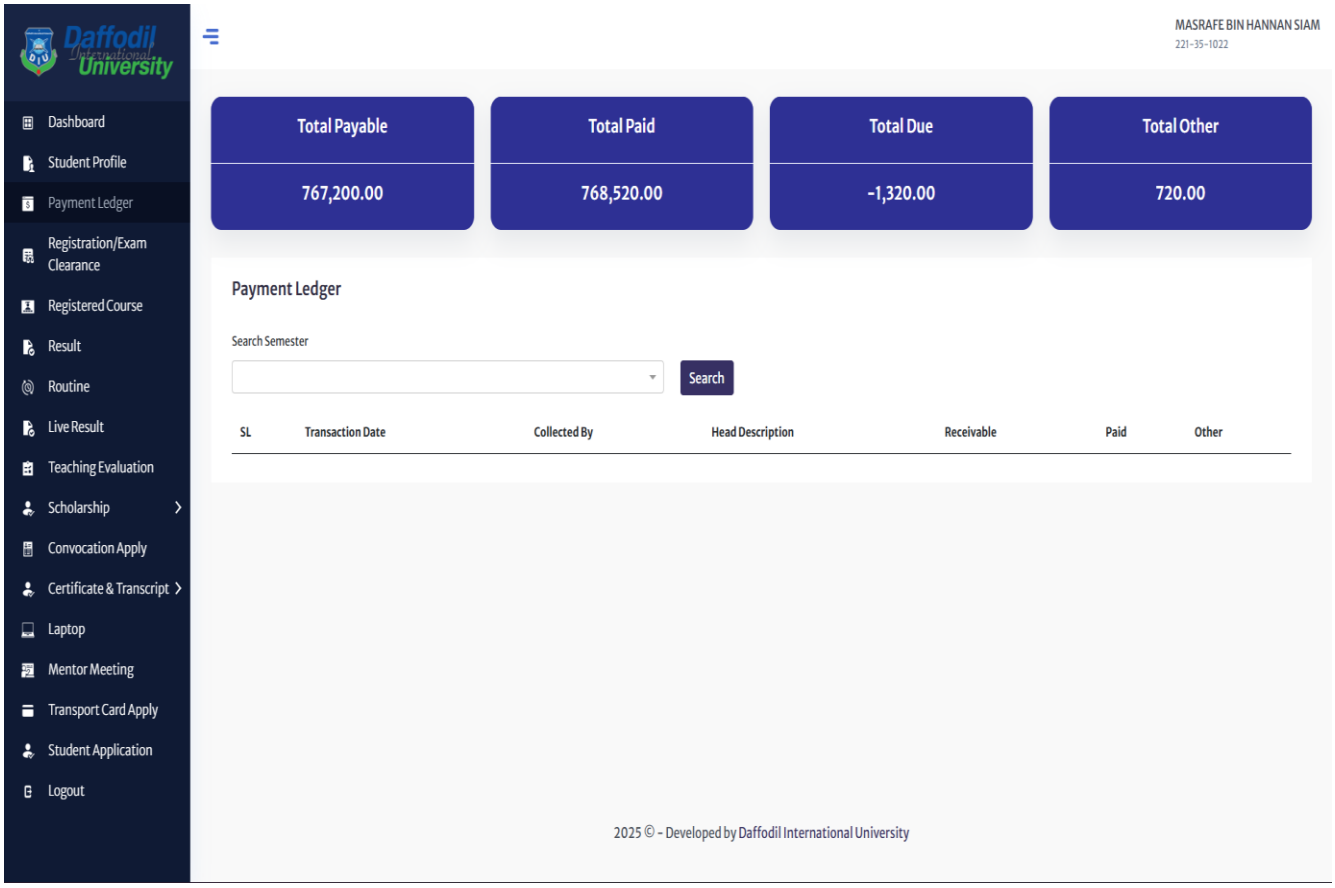


## *% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.


Daffodil International University
MASRAFE BIN HANNAN SIAM
221-35-1022
Dashboard
Student Profile
Payment Ledger
Registration/Exam Clearance
Registered Course
Result
Routine
Live Result
Teaching Evaluation
Scholarship
Convocation Apply
Certificate & Transcript
Laptop
Mentor Meeting
Transport Card Apply
Student Application
Logout
Total Payable
767,200.00
Total Paid
768,520.00
Total Due
-1,320.00
Total Other
720.00
Payment Ledger
Search Semester
Search
SL
Transaction Date
Collected By
Head Description
Receivable
Paid
Other
2025 © - Developed by Daffodil International University