TVFed-P: Tversky-based Federated Learning with Personalized Loss Parameterization for Medical Imbalanced Data

Samar Samir Khalil[1,2] (✉), Noha S. Tawfik[1], and Marco Spruit[2,3]

[1] Computer Engineering Department, Arab Academy for Science, Technology and Maritime Transport, Alexandria, Egypt {samar, noha.abdelsalam}@aast.edu

[2] Leiden Institute of Advanced Computer Science, Leiden University, 2333CC Leiden, The Netherlands {s.s.khalil, m.r.spruit}@liacs.leidenuniv.nl

[3] Public Health & Primary Care, Leiden University Medical Center, 2333ZA Leiden, The Netherlands m.r.spruit@lumc.nl

Abstract. Schizophrenia is a severe mental health condition that disrupts personal, social, and professional life, impacting millions worldwide. Early and accurate diagnosis is essential for improving patients' quality of life. While artificial intelligence (AI) holds significant promise in this area, conventional AI approaches typically rely on centralized data, which poses challenges related to privacy, scalability, and accessibility. Federated learning (FL) addresses these issues by enabling decentralised model training across institutions while preserving patient confidentiality. This study implements a horizontal federated learning model to enable collaboration between different healthcare institutions (clients) for schizophrenia diagnosis using resting-state functional MRI (rs-fMRI) data. To tackle the inherent class imbalance in schizophrenia data, where positive cases are typically under-represented, our approach first integrates the Tversky loss function. Building on this, we propose a personalization scheme that allows each client to adapt their own Tversky loss parameters, enabling more tailored optimization and improved performance across heterogeneous data distributions. Combined with FL aggregation algorithms tailored for non-IID data, TVFed-P effectively addresses both quantity and feature skew, which are commonly observed in many medical datasets. We present a comprehensive comparative analysis of our proposed model, TVFed-P, against both locally trained and centralized models reported in the literature. Our best-performing variant surpasses all baseline approaches, demonstrating superior performance. These findings reinforce the promise of federated learning as a secure and effective strategy for early schizophrenia diagnosis.

Keywords: Federated learning · Personalization · Non-IID data · Data Imbalance · Tveresky loss · Schizophrenia Diagnosis

1 Introduction

According to the World Health Organisation (WHO) [23], schizophrenia is a severe mental illness that affects about 24 million people around the world, 0.32%

of all people, with a slightly higher prevalence among adults, at 0.45% of all adults. While its global prevalence varies, it has a huge impact, causing significant distress and impairment across various aspects of life, including personal relationships, education, and employment. People with schizophrenia have a life span that is two to three times lower than the average person, mainly due to associated physical health problems and related challenges [4]. In addition to health problems, they often have to deal with violations of their human rights, extreme stigma, and systemic discrimination that limit their access to essential services. Improving their quality of life necessitates an early and precise diagnosis to start the treatment process.

Artificial Intelligence (AI) has shown great potential in various domains, including the detection of mental health disorders. Traditional machine learning (ML) approaches rely on centralised learning schemes where patient data from multiple healthcare institutions are collected into a single centralised point for model training. However, this approach not only raises critical concerns about patient data privacy but also limits the scope of available datasets, as many healthcare providers refuse to and/or are legally forbidden from sharing sensitive patient information. To address these challenges, Federated Learning (FL), a decentralised learning paradigm introduced in [11], enables machine and deep learning models to be trained across distributed datasets without transferring raw data from their original sources. With the ability to learn collaboratively while keeping data localised, FL offers a privacy-preserving alternative that is particularly advantageous in sensitive applications, such as schizophrenia detection. Since patient records remain with their respective institutions, FL ensures compliance with data protection regulations while maximising participation from diverse healthcare providers. Furthermore, FL mitigates the issue of data scarcity in schizophrenia classification by utilising small, fragmented datasets in the training process, thereby enhancing model generalisability and diagnostic accuracy.

Federated learning faces many challenges arising from its decentralised nature, including the non-independent and identically distributed (non-IID) data and security vulnerabilities that malicious actors can exploit. Non-IIDness of data can manifest in three forms: label skew, feature skew and quantity skew. In label skew, clients have uneven class distribution where some clients have a disproportionate amount of data for certain classes compared to others. Feature skew is the difference in how the input features, the data itself, are spread out among different clients, while quantity skew is the difference in the amount of data held by each client [7]. These challenges are particularly prevalent in the medical domain, due to the decentralised nature of healthcare systems, variations in clinical practices, and patient populations [1]. Such heterogeneity can significantly impact the performance and convergence of FL models in medical applications in general and mental health in specific.

Motivated by these challenges, we propose a federated learning framework tailored to tackle data heterogeneity and client-level disparities. We summarize our key contributions in this paper as follows:

- We propose a novel federated learning framework that incorporates Tversky loss function [17] to effectively handle the data skewness where positive samples (diagnosed patients) are scarce across distributed clients. To validate the proposed **TVFed** model, this research utilizes the Strategic Research Program for Brain Science (SRPBS) multi-site connectivity dataset [20]. designed for schizophrenia classification through resting-state functional MRI (rs-fMRI) data.
- To further mitigate non-IID data distributions, a comprehensive empirical study to identify the most effective aggregation mechanism for imbalanced and non-IID neuroimaging data. Our findings reveal the optimal combination of loss formulation and aggregation scheme for robust decentralised training.
- Building on these insights, we introduce a personalization layer that enables optimal client-specific adaptation through regularized loss parameterization. This extension leads to our final framework, **TV**ersky-based **Fed**erated learning with **P**ersonalized loss parameterization **(TVFed-P)**, which achieves superior performance while preserving privacy and enhancing generalization across institutions.

Unlike traditional centralised models, the proposed privacy-preserving model ensures that rs-fMRI data remains localised at each institution while achieving superior classification performance. Experimental results show that the proposed model overcomes class imbalance, quantity skew, and feature skew without compromising data privacy.

2 Related Work

Traditional and advanced machine learning techniques have been employed in schizophrenia research, improving classification performance using resting-state functional MRI (rs-fMRI) and other data modalities. Studies leveraging large-scale datasets such as SRPBS have focused on identifying biomarkers and enhancing model architectures. Nevertheless, much of this work overlooks institutional heterogeneity and data privacy, which are crucial in medical applications. This section summarises previous work in three main areas: performance of traditional centralised methods for schizophrenia classification; FL models tailored for schizophrenia diagnosis and their challenges, and finally, methods and techniques to reduce class imbalance in FL for reliable real-world deployment.

2.1 Centralised Learning Approaches for Schizophrenia

This section specifically limits its scope to the employed SRPBS dataset, as centralised learning on other datasets and/or modalities falls outside the current research objectives. Takahara et al. [19] obtained objective classification biomarkers for patients with major depressive disorder and extended these biomarkers to other disorders, including schizophrenia. The SRPBS rs-fMRI dataset was used to implement multiple pipelines combining six parcellation methods and four functional connectivity (FC) calculation methods including Pearson's correlation and partial correlation. The authors provided a comparative analysis of five machine learning algorithms, including SVM and Random Forest. The highest achieved a composite score of 0.852, an evaluation metric reported by the authors that combines both the area under the curve (AUC) and the Matthews correlation coefficient [3]. Zhu et al. [26] proposed Temporal-BCGCN, a graph neural network (GNN) model to classify schizophrenia using dynamic functional connectivity (dFC) from rs-fMRI. Only a subset of the SRPBS dataset consisting of 200 subjects – 92 schizophrenic patients and 108 healthy controls – was used, scoring 0.85 F1-score. Wang et al. [22] designed the EA-GNAS (Evolutionary Algorithm-based Graph Neural Architecture Search) framework based on high-performance Graph Neural Network (GNN) architectures. They used GNNExplainer to interpret the model's results and locate important brain regions linked to schizophrenia diagnosis. To address class imbalance, the training set was balanced by selecting an equal number of healthy controls and schizophrenia patients. The model achieved an accuracy of 0.8246, an f1-score of 0.8438, and an AUC of 0.8258.

2.2 Federated Learning Approaches for Schizophrenia

Federated learning was previously utilised in various mental health research [7], including schizophrenia classification. Huang et al. [6] presented a Federated Multi-Task Learning (FMTLJD) framework to jointly diagnose autism spectrum disorder, attention deficit/hyperactivity disorder and schizophrenia utilising resting-state fMRI data from the ABIDE, ADHD-200, and COBRE datasets. Combining multi-task learning and FL helps to overcome domain shift and data scarcity by using shared knowledge across diseases. They employ a multi-gate mixture of experts (FMMoE) classifier for adaptive task weighting, a federated contrastive learning feature extractor (FCLFE) for dimensionality reduction, and differential privacy (DP-SGD) for safe model training. The model attained accuracies of 69.48% (ASD), 71.44% (ADHD), and 83.29% (SCZ). One of the constraints of this model to generalisation is the small sample size as the COBRE dataset contains only 72 SCZ patients and 74 healthy controls.

Wang et al. [21] proposed Fed-FAAD, a federated few-shot domain-adaptive anomaly detection framework for diagnosing schizophrenia. They used multi-site fMRI data from the Human Connectome Project dataset as a source domain (HC) and six clinical sites as target domains (SCZ). By combining federated learning with Gaussian differential privacy and deep SVDD for hypersphere-based anomaly detection, Fed-FAAD aligns feature distributions across sites and improves generalisation. Experiments show it achieves 78.1% (10-shot) and 82.5% (20-shot) AUC. A key consideration is that the model relies on only 10–20 labelled SCZ samples per site for training. While this addresses data scarcity, it risks overfitting or bias toward site-specific noise.

Salam et al. [16] developed a FL framework for diagnosing schizophrenia utilising functional and structural MRI data from the 10th annual MLSP competition dataset, comprising 144 subjects: 75 (HC) and 69 (SCZ). K-NN, SVM, Random Forest, and other classifiers are augmented with swarm intelligence algorithms (Firefly and Jaya) for feature selection, reducing the number of features from 410 to 143–186. K-NN integrated with swarm intelligence outperformed other models achieving 100% accuracy and AUC. Similar to [6] the limited sample size constrains the model's generalisability.

2.3 Class Imbalance in Federated Learning

Class imbalance in federated learning occur when all the clients share the same classes, but one or more of those classes are significantly under-represented across all clients. Models trained on imbalanced datasets often exhibit poor generalisation, particularly for minority classes, which can be neglected during the learning process. Many research was published to mitigate the imbalance using sampling, algorithm-centred or system-centred stratigies [25]. The recent work of Wu et al. proposed FedIIC [24], a loss function that utilises difficulty-aware logit adjustment (DALA), Inter- and Intra-client contrastive learning. Together with FedAvg, FedIIC outperformed multiple federated algorithms in class-imbalanced medical image classification. Sahoo et al. [15] proposed AdaFedProx, an FL framework designed to address system heterogeneity (variations in client hardware) and data heterogeneity (class imbalance) in medical image classification. It employs a reinforcement learning (RL)-based Deep Q-Network (DQN) on the server side to dynamically adjust the proximal term μ in the FedProx [9] algorithm, optimising regularisation based on client-specific states such as data distribution entropy, class ratios, computational capacity, and performance feedback. FedIIC and AdaFedProx represent two different approaches in handling class imbalance, one targeting the local loss function at each client while the other adjusts the aggregation parameters at the server.

3 Methodology

Schizophrenia and other mental health problems suffer from intrinsic data skewness that arises from natural variations in data distribution across clients (healthcare institutions), including class skew, quantity skew, and feature distribution shifts. Class/label skew is reflected in the scarcity of schizophrenia patients in individual healthcare institutions. Quantity skew manifests in the uneven data contributions across clients, with large healthcare providers contributing substantially more data compared to smaller institutions. Feature skew in federated fMRI studies can result from both technical variability (different scanners) and clinical variability (symptom severity).

3.1 Data Preprocessing and Local Model Configuration

Within this diagnostic framework for schizophrenia, individual clients contribute resting-state functional MRI (rs-fMRI) data, a non-invasive neuroimaging modality that captures brain activity by measuring blood flow changes. For preprocessing, each brain image is divided into 140 parcels following the sulci-based anatomical atlas of the Extended Brainvisa Sulci Atlas [13]. Resting-state functional connectivity is then assessed by calculating Pearson correlation coefficients [14] between the mean fMRI time series of every possible parcel pair, resulting in a 140 × 140 connectivity matrix, which includes 9,730 unique functional connections (FCs). Despite using different parcelation atlases, Shevchenko et al. [18] proved that using raw FCs in classification outperformed low-dimensional derivatives; thus, the complete set of 9730 FCs is used as input for our learning model. Each site/client trains a feedforward neural network on its local imbalanced dataset. The network consists of an input layer for the 9730 FCs, two fully connected hidden layers of sizes 512 and 265 with ReLU activation functions, and an output layer of 1 neuron and a sigmoid activation function. A dropout layer is added after each hidden layer to mitigate overfitting, given the high feature-to-sample ratio in the dataset.

3.2 Tversky loss with personalized parameterization

To mitigate the class imbalance present in each client, the Tversky loss function is utilised as a generalised variant of the Dice loss [12]. It was originally implemented to handle imbalance present in medical image segmentation tasks [17], where the number of positive voxels is often significantly lower than that of negative voxels. It mitigates class imbalance by imposing greater penalties on false negatives (FNs) compared to false positives (FPs). The Tversky loss function is formulated as:

$$\mathcal{T}(\alpha, \beta) = \frac{\sum_{i=1}^{N} p_{0i} g_{0i}}{\sum_{i=1}^{N} p_{0i} g_{0i} + \alpha \sum_{i=1}^{N} p_{0i} g_{1i} + \beta \sum_{i=1}^{N} p_{1i} g_{0i}} \tag{1}$$

where α and β control FP and FN penalties, respectively. p_{0i} and p_{1i} are the probabilities for positive/negative samples, while g_{0i} and g_{1i} are the ground truth labels. By tuning α and β, recall is prioritised (reducing FNs), which is crucial for detecting small positives in imbalanced data. Notably, setting $\alpha = \beta = 0.5$ recovers the Dice loss while $\alpha + \beta = 1$ yielding, F_β scores.

We further address client-specific data imbalance in schizophrenia detection, where each client autonomously computes an adaptive weighting coefficient α based on its local class distribution. Specifically, $\alpha = \frac{n_+}{n_+ + n_-}$, where n_+ and n_- denote the number of positive and negative samples, respectively. This locally optimized client-specific α allows the loss to adapt to each client's class imbalance, prioritizing recall in clients with scarce positive samples in imbalanced federated settings. The automatic personalization reduces the impact of over-represented negative samples while strongly penalising false negatives, ensuring that the global model remains highly sensitive to schizophrenia cases, where missing true positives poses a substantially greater risk than false predictions.

3.3 TVFed-P learning framework

As illustrated in Figure 1, our proposed TVFed-P is a horizontal federated learning (HFL) that allows for cooperative model training across several healthcare institutions without requiring direct data exchange. In HFL, a centralised server sends an initial model seed W_i to the participating clients K; each client trains its version of the model on its local data for a number of epochs and then sends back the trained model W_{i+1}^K to the server. The server completes one communication round when it aggregates the local models into one updated global model W_{i+1}. The updated model is then shared again until the learning phase ends. TVFed-P exploits FedAvgM [5] for enhanced performance as it improves upon standard FedAvg by maintaining a momentum buffer at the server level, which effectively reduces update oscillations and mitigates client drift in heterogeneous non-IID data environments. The source code is available at GitHub.

4 Experiments

4.1 SRPBS Dataset

To validate the proposed framework, we use The Multi-disorder Connectivity Dataset from the Strategic Research Program for the Promotion of Brain Science (SRPBS) [20]. The dataset comprises resting-state functional MRI data from both patients and healthy control volunteers. It include data on six disorders from eight institutions across Japan. The dataset is further divided into

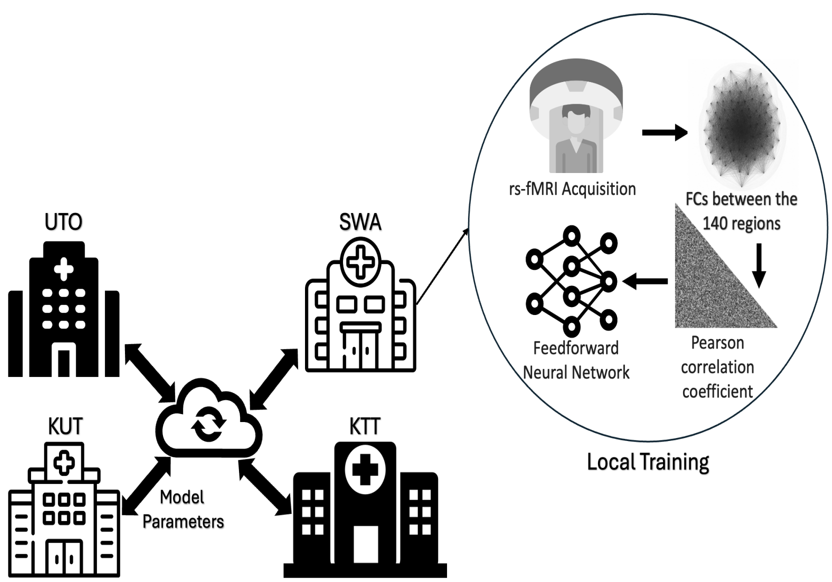

Fig. 1. Proposed TVFed-P framework

sites because participating institutions may utilize multiple scanners. Data on schizophrenia spectrum disorder are sourced from three institutions: University of Tokyo (UTO), Showa University (SWA), and Kyoto University (KUT and KTT), comprising a total of 146 schizophrenic patients (SCZ) and 505 healthy controls (HC). Table 1 summarises the dataset distribution across clients, showing two inherent forms of non-IID data heterogeneity characteristic of real-world medical imaging scenarios. Feature skew arises from variations in scanner types and quantity skew reflects disparities in dataset sizes, with some clients containing up to 50% fewer samples than others.

Table 1. Dataset Summary

Site	Scanner	SCZ	HC	Total	SCZ/Total ratio
UTO	MR750W, GE	35	170	205	17.1%
SWA	Verio, Siemens	19	101	120	15.8%
KTT	Trio, Siemens	45	159	204	22.1%
KUT	TimTrio, Siemens	47	75	122	38.5%

4.2 Experimental Setup

In local training, PyTorch was used to implement and train the entire model end-to-end for 100 epochs using the Adam optimiser [8] with an initial learning rate of 0.001, weight decay of 1e-4, and a batch size of 64. Hyperparameters were systematically tuned by testing various values until optimal settings were identified. The optimisation leveraged Tversky loss to address class imbalance. Flower [2] and PyTorch were used to implement the HFL experiment. In HFL, each client performed one epoch of local training per communication round, while the server executed 100 rounds of fedAvgM with a server momentum of 0.9. All other hyperparameters, including the batch size and Tversky coefficients, remained consistent between both experiments to ensure a fair comparison. All the experiments were conducted on Windows 10 operating system withIntel(R) Core(TM) i7-9750H 2.59 GHz CPU with 16GB RAM and an NVIDIA GeForce RTX 2070 with Max-Q design with 8GB memory.

4.3 Evaluation Metrics

Addressing the class imbalance problem, three complementary evaluation metrics that are robust to class skew are used: the F1-score, Matthews Correlation Coefficient (MCC), and a composite score.

The F1-score focuses on minority class performance by computing the harmonic mean of precision and recall:

$$F_1\text{-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{2}$$

MCC [3] incorporates all four confusion matrix values (unlike F1-score, which ignores TN). It quantifies how well both classes are predicted, even when one class dominates. This metric ranges from -1 (perfect inverse prediction) to +1 (perfect prediction), with 0 indicating random performance. It is formulated as:

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{3}$$

Following Takahara et al. [19], a composite score combining area under the curve (AUC) and normalised MCC is reported. This score normalises MCC to [0,1] range and equally weights it with AUC for comprehensive assessment. It is formulated as:

$$\text{Composite score} = 0.5 \times \text{AUC} + 0.5 \times \frac{\text{MCC} + 1}{2} \tag{4}$$

Collectively, these three metrics address minority class performance (F1), overall classification quality (MCC), and rank discrimination and balanced accuracy (composite score).

4.4 Experimental Results

To evaluate the impact of collaborative learning in healthcare settings, two comparative experiments are conducted:

- **Local Learning,** simulating isolated training where each institution trains a model solely on its local data without collaboration, reflecting scenarios where data sharing is restricted due to privacy or regulatory constraints.
- **Horizontal Federated Learning,** where institutions agree to collaboratively train a global model by sharing only model parameters (not raw data) via a central server that aggregates updates.

The comparative performance of local versus federated learning models is summarised in Table 2. The federated approach demonstrated significant improvements across all participating institutions, with MCC gains ranging from 0.079 at UTO to 0.336 at KTT. Notably, SWA and KUT achieved perfect classification (1 across all metrics) through federated learning. Building upon KUT's already strong local performance (MCC=0.921), this exceptional result probably comes from KUT's balanced dataset, where the ratio of schizophrenia patients to the overall sample size is 38.5%, as shown in table 1. In this case, federated learning primarily reduced variance rather than bias. Similarly, the high classification scores achieved by SWA in the federated setting can be attributed to the substantial performance gain observed across the first three participating sites, with gains ranging from 0.204 to 0.336, and SWA's inherently stronger local performance compared to other sites.

Three key observations emerge from the results: First, KTT—despite moderate data imbalance (scz/total=22.1%)—showed the most dramatic MCC improvement, suggesting its local data contained challenging patterns that re-

Table 2. Model performance comparison between local and federated learning. Gain indicates the rise in the proposed TVFed-P over local performance.

Metric	Local Learning				TVFed-P				Gain			
	UTO	SWA	KTT	KUT	UTO	SWA	KTT	KUT	UTO	SWA	KTT	KUT
F1-score	0.667	0.800	0.632	0.952	0.824	1.0	0.889	1.0	0.157	0.200	0.257	0.048
MCC	0.594	0.775	0.522	0.921	0.799	1.0	0.858	1.0	0.204	0.225	0.336	0.079
Composite Score	0.806	0.906	0.750	0.970	0.920	1.0	0.954	1.0	0.114	0.094	0.204	0.030

quired broader examples for proper generalisation. Second, UTO's relative underperformance (MCC=0.799 vs. site average=0.914) correlates with its different acquisition protocol (non-Siemens scanner) in comparison with the other 3 sites, indicating feature distribution skew that federated learning partially mitigated. Third, although KUT already exhibited strong standalone performance (MCC=0.921), it acted as an anchor site in the federation, contributing valuable gradients to support other clients. Notably, it still experienced a performance boost (MCC=1.0) while retaining 100% data locality. These findings demonstrate federated learning's dual benefit of improving model robustness for institutions with complex data (KTT, UTO) while preserving the privacy advantages of decentralised training (KUT).

While this study excludes centralised learning as a possible solution because it would require healthcare institutions to violate privacy laws and institutional policies. Howerver it is worth noting that TVFed-P demonstrated superior performance not only over local training baselines but also outperformed centralised models reported in literature. The comparison with centralised models was conducted by averaging the scores obtained by the four clients using the proposed HFL. Specifically, the TVFed-P model achieved an F1-Score of 0.928, an MCC of 0.914, and a Composite Score of 0.969. This performance surpasses existing centralised models from literature, such as Zhu et al. [26] with an F1-Score of 0.850, Wang et al. [22] with an F1-Score of 0.844, and Takahara et al. [19] with a Composite Score of 0.852. These results highlight the effectiveness of our federated approach in achieving state-of-the-art performance despite the decentralised and heterogeneous nature of the data.

4.5 Component Evaluation and Model Refinement

To thoroughly evaluate the proposed model, our in-depth analysis consists of three stages. First, we conducted a comparative study of loss functions, benchmarking the Tversky loss against alternatives to address class imbalance. Next, we mitigated non-IID data challenges by testing different federated aggregation algorithms. Finally, we evaluate our proposed personalized local parameterization to assess its efficacy in adapting the global model to local data distributions.

Loss function comparison: Table 3 compares five loss functions for local learning, beginning with binary cross-entropy (BCE) as a baseline, followed by class-imbalance solutions: Focal Loss [10] that adaptively weights hard examples, Dice Loss [12] that optimises for predicted and truth regions overlap, and FedIIC [24] primarly designed for federated medical imaging classification. Results demonstrate Tversky Loss's superior performance, attributed to its tunable $\alpha-\beta$ parameters that explicitly penalise false negatives, a key advanatge for medical diagnosis. Consequently, it was chosen as the optimal loss function for subsequent federated experiments, particularly given the persistent class imbalance across dataset sites. To ensure fair comparisons, a methodical grid search was conducted for all loss functions requiring specific parameters, with all clients constrained to use identical settings (i.e., no personalization). For instance, when using Dice loss, the smoothing parameter was tuned to prevent instability arising from rare class occurrences and skewed label distributions. The results presented reflect the optimal configurations identified through this comprehensive evaluation process.

Table 3. Performance Comparison of Different Loss Functions at SRPBS Sites (Average of Local Training across All 4 Clients)

Metric	**BCE**	**Focal**	**Dice**	**FedIIC**	**Tversky**
F1-score	0.656	0.672	0.750	0.620	**0.791**
MCC	0.595	0.604	0.692	0.535	**0.737**
Composite Score	0.822	0.829	0.868	0.771	**0.879**

Federated aggregation comparison: Using fixed (non-adaptive) Tversky loss parameters α and β across all clients, we compared three federated aggregation approaches under non-IID conditions: FedAvg as the baseline method, FedProx with its proximal term for client stability, and FedAvgM incorporating server-side momentum. The server momentum buffer in FedAvgM accumulates a moving average of past global updates, which helps smooth out fluctuations caused by heterogeneous client data. In the context of federated classification on rs-fMRI data from the SRPBS dataset—characterized by site-specific variability and label imbalance—this momentum mechanism provides a stabilizing effect, enhancing convergence and improving generalisation. Table 4 shows that FedAvgM outperformed both alternatives, demonstrating superior handling of data heterogeneity through its momentum-driven update smoothing.

personalization impact: While the Tversky loss function is well-suited for handling class imbalance in binary classification tasks, applying fixed parameters (α, β) uniformly across all clients assumes a homogeneous class distribution—an assumption that does not hold in federated rs-fMRI settings. Due to site-specific demographic and clinical variations, each client's data exhibits different levels of imbalance between diagnosed and control subjects. Relying on fixed global parameters under such heterogeneity can lead to suboptimal performance. To address this, we introduce a client-specific personalized loss configuration strategy that adapts Tversky loss parameters (α, β) to each client's local class ratio.

Table 4. Performance Comparison of Federated Learning Aggregation Methods Combined with Tversky Loss (Average of Federated Results Across All Clients).

Metric	FedAvg	FedProx	FedAvgM
F1-score	0.873	0.871	**0.897**
MCC	0.863	0.850	**0.878**
Composite Score	0.941	0.936	**0.946**

As demonstrated in Table 5, this personalized approach yielded significant performance gains, with the F1-score improving from 0.897 to 0.928 compared to the consistent global parameterization. The MCC and composite score showed similar improvements, 0.914 and 0.969 respectively. This highlights the effectiveness of client-specific optimization in mitigating heterogeneity while preserving the collaborative strengths of federated learning.

Table 5. Comparison of fixed (non-adaptive) vs. Personalized α and β Tversky's values using FedAvgM

Metric	TVFed	TVFed-P
F1-score	0.897	**0.928**
MCC	0.878	**0.914**
Composite Score	0.946	**0.969**

5 Conclusion and Future Work

This paper proposes TVFed-P, a horizontal federated learning model designed to leverage decentralised multi-site data while ensuring strict privacy compliance. Our proposed model addresses intrinsic class imbalance and client heterogeneity through a comprehensive enhanced architecture: it incorporates a Tversky loss function tailored for imbalanced data, integrates FedAvgM to enhance training stability across clients, and introduces a personalization layer that adapts loss parameters to each client's local class distribution. The model has been validated using resting-state functional MRI (rs-fMRI) data from the Strategic Research Program for Brain Science (SRPBS) multi-site connectivity dataset. Experimental results demonstrate that TVFed-P consistently outperforms both local baselines and existing literature models. In addition, the model demonstrates robustness in various client situations, including severe class imbalance, balanced distributions, and considerable feature skew. The personalization, achieved via the local-optimized Tversky loss, further improved these results, yielding average scores of 0.928 F1-Score, 0.914 MCC, and 0.969 Composite Score. Beyond schizophrenia detection, the proposed TVFed-P framework is generalizable and can be adapted to other medical and non-medical domains involving non-IID, imbalanced, and privacy-sensitive data.Future work involves investigating client selection algorithms, incorporating advanced privacy-preserving techniques such as differential privacy or alternative security mechanisms, and exploring the integration of other data types and multimodality for enhanced schizophrenia diagnosis.

Acknowledgments. Data used in the preparation of this work were obtained from the DecNef Project Brain Data Repository (https://bicr-resource.atr.jp/srpbsopen/), collected as part of the Japanese Strategic Research Program for the Promotion of Brain Science (SRPBS) supported by the Japanese Advanced Research and Development Programs for Medical Innovation (AMED).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Babar, M., Qureshi, B., Koubaa, A.: Investigating the impact of data heterogeneity on the performance of federated learning algorithm using medical imaging. Plos one **19**(5), e0302539 (2024)
2. Beutel, D.J., Topal, T., Mathur, A., Qiu, X., Fernandez-Marques, J., Gao, Y., Sani, L., Li, K.H., Parcollet, T., de Gusmão, P.P.B., et al.: Flower: A friendly federated learning research framework. arXiv preprint arXiv:2007.14390 (2020)
3. Chicco, D.: Ten quick tips for machine learning in computational biology. BioData mining **10**(1), 35 (2017)
4. Hjorthøj, C., Stürup, A.E., McGrath, J.J., Nordentoft, M.: Years of potential life lost and life expectancy in schizophrenia: a systematic review and meta-analysis. The Lancet Psychiatry **4**(4), 295–301 (2017)
5. Hsu, T.M.H., Qi, H., Brown, M.: Measuring the effects of non-identical data distribution for federated visual classification. arXiv preprint arXiv:1909.06335 (2019)
6. Huang, Z.A., Hu, Y., Liu, R., Xue, X., Zhu, Z., Song, L., Tan, K.C.: Federated multi-task learning for joint diagnosis of multiple mental disorders on mri scans. IEEE Transactions on Biomedical Engineering **70**(4), 1137–1149 (2022)
7. Khalil, S.S., Tawfik, N.S., Spruit, M.: Exploring the potential of federated learning in mental health research: a systematic literature review. Applied Intelligence **54**(2), 1619–1636 (2024)
8. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
9. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated optimization in heterogeneous networks. Proceedings of Machine learning and systems **2**, 429–450 (2020)
10. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980–2988 (2017)
11. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial intelligence and statistics. pp. 1273–1282. PMLR (2017)

12. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). pp. 565–571. Ieee (2016)
13. Perrot, M., Rivière, D., Mangin, J.F.: Cortical sulci recognition and spatial normalization. Medical image analysis **15**(4), 529–550 (2011)
14. Poldrack, R.A., Mumford, J.A., Nichols, T.E.: Handbook of Functional MRI Data Analysis. Cambridge University Press (2011)
15. Sahoo, P., Tripathi, A., Saha, S., Mondal, S., Singh, J.P., Sharma, B.: Adafedprox: A heterogeneity-aware federated deep reinforcement learning for medical image classification. IEEE Transactions on Consumer Electronics (2024)
16. Salam, M.A., Badr, E., Monier, E., Mohamed, A.: Schizophrenia diagnosis using optimized federated learning models. IJCSNS **829** (2022)
17. Salehi, S.S.M., Erdogmus, D., Gholipour, A.: Tversky loss function for image segmentation using 3d fully convolutional deep networks. In: International workshop on machine learning in medical imaging. pp. 379–387. Springer (2017)
18. Shevchenko, V., Benn, R.A., Scholz, R., Wei, W., Pallavicini, C., Klatzmann, U., Alberti, F., Satterthwaite, T.D., Wassermann, D., Bazin, P.L., et al.: A comparative machine learning study of schizophrenia biomarkers derived from functional connectivity. Scientific Reports **15**(1), 2849 (2025)
19. Takahara, Y., Kashiwagi, Y., Tokuda, T., Yoshimoto, J., Sakai, Y., Yamashita, A., Yoshioka, T., Takahashi, H., Mizuta, H., Kasai, K., et al.: Comprehensive evaluation of pipelines for classification of psychiatric disorders using multi-site resting-state fmri datasets. Neural Networks **187**, 107335 (2025)
20. Tanaka, S.C., Yamashita, A., Yahata, N., Itahashi, T., Lisi, G., Yamada, T., Ichikawa, N., Takamura, M., Yoshihara, Y., Kunimatsu, A., et al.: A multi-site, multi-disorder resting-state magnetic resonance image database. Scientific data **8**(1), 227 (2021)
21. Wang, C., Su, J., Fan, Z., Hu, D., Zeng, L.L.: Federated few-shot domain-adaptive anomaly detection for diagnostic classification of schizophrenia. In: 2023 China Automation Congress (CAC). pp. 7536–7541. IEEE (2023)
22. Wang, S., Tang, H., Himeno, R., Solé-Casals, J., Caiafa, C.F., Han, S., Aoki, S., Sun, Z.: Optimizing graph neural network architectures for schizophrenia spectrum disorder prediction using evolutionary algorithms. Computer Methods and Programs in Biomedicine **257**, 108419 (2024)
23. WHO: Schizophrenia (2022), `https://www.who.int/news-room/fact-sheets/detail/schizophrenia`
24. Wu, N., Yu, L., Yang, X., Cheng, K.T., Yan, Z.: Fediic: Towards robust federated learning for class-imbalanced medical image classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 692–702. Springer (2023)
25. Zhang, J., Li, C., Qi, J., He, J.: A survey on class imbalance in federated learning. arXiv preprint arXiv:2303.11673 (2023)
26. Zhu, C., Tan, Y., Yang, S., Miao, J., Zhu, J., Huang, H., Yao, D., Luo, C.: Temporal dynamic synchronous functional brain network for schizophrenia classification and lateralization analysis. IEEE Transactions on Medical Imaging (2024)