National College of Ireland

Project Submission Sheet

Student Name:	ARYAN NAGNATH SALGE		
Student ID:	X23268018		
Programme:	MSC_DAD_A	**Year:**	2024-25
Module:	RESEARCH PRACTICUM		
Lecturer:	DR. DAVID HAMILL		
Submission Due Date:	08/08/2025		
Project Title:	Federated Transfer Learning for Cross-Institution Fraud Detection in Finance		
Word Count:	8000		

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
<u>ALL</u> internet material must be referenced in the references section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature:	ARYAN SALGE
Date:	08/08/2025

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. Projects should be submitted to your Programme Coordinator.
3. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
4. You must ensure that all projects are submitted to your Programme Coordinator on or before the required submission date. **Late submissions will incur penalties.**
5. All projects must be submitted and passed in order to successfully complete the year. **Any project/assignment not submitted will be marked as a fail.**

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

AI Acknowledgement Supplement

RESEARCH PRACTICUM

Federated Transfer Learning for Cross-Institution Fraud Detection in Finance

Your Number	Name/Student	Course	Date
ARYAN SALGE		MSC_DAD_A	08/08/2025

This section is a supplement to the main assignment, to be used if AI was used in any capacity in the creation of your assignment; if you have queries about how to do this, please contact your lecturer. For an example of how to fill these sections out, please click here.

AI Acknowledgment

This section acknowledges the AI tools that were utilized in the process of completing this assignment.

Tool Name	Brief Description	Link to tool
-	-	-

Description of AI Usage

This section provides a more detailed description of how the AI tools were used in the assignment. It includes information about the prompts given to the AI tool, the responses received, and how these responses were utilized or modified in the assignment. **One table should be used for each tool used**.

[Insert Tool Name]	
[Insert Description of use]	
[Insert Sample prompt]	[Insert Sample response]

Evidence of AI Usage

This section includes evidence of significant prompts and responses used or generated through the AI tool. It should provide a clear understanding of the extent to which the AI tool was used in the assignment. Evidence may be attached via screenshots or text.

Additional Evidence:

[Place evidence here]

Additional Evidence:

[Place evidence here]

National College of Ireland

Federated Transfer Learning for Cross-Institution Fraud Detection in Finance

MSc Research Project
MSc in Data Analytics

ARYAN NAGNATH SALGE
Student ID: x23268018

School of Computing
National College of Ireland

Supervisor: Dr. David Hamill

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	ARYAN NAGNATH SALGE
Student ID:	X23268018
Programme:	MSC_DAD_A **Year:** 2024-25
Module:	RESEARCH PRACTICUM
Supervisor:	DR. DAVID HAMILL
Submission Due Date:	08/08/2025
Project Title:	Federated Transfer Learning for Cross-Institution Fraud Detection in Finance
Word Count:	8000 **Page Count** 12

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ARYAN SALGE

Date: 08/08/2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Federated Transfer Learning for Cross-Institution Fraud Detection in Finance

Aryan Nagnath Salge
MSc Data Analytics
National College of Ireland
x23268018@student.ncirl.ie

***Abstract*—Financial fraud detection is a critical task for financial institutions, but privacy and regulatory constraints often prevent cross-institutional data sharing, limiting the effectiveness of machine learning models trained on isolated datasets. This study examines the application of Federated Transfer Learning (FTL) to facilitate collaborative fraud detection while maintaining data confidentiality. Three heterogeneous real-world fraud datasets, representing independent financial institutions with differing sizes, class distributions, and partially overlapping features, harmonized for joint training, were used to simulate a cross-institutional learning environment.**

The proposed framework coordinates shared features, applies Principal Component Analysis (PCA) for dimensionality reduction, and trains a global neural network model using the Federated Averaging (FedAvg) algorithm [1] under differential privacy constraints. A fine-tuning step is then performed on each of the three clients to personalize the model to local data distributions. Performance is examined using Accuracy, Precision, Recall, and F1-Score, and compared against a centralized baseline model trained on aggregated data. Results show that FTL achieves accuracy levels above 98% across all clients, closely approaching centralized performance, while enabling privacy-preserving collaboration. Fine-tuning particularly benefits data-poor and imbalanced clients, improving accuracy from 57.4% to 99.4%, highlighting the potential use of knowledge transfer in a federated environment. This dramatic performance gain occurred despite Dataset 3 initially containing the most missing values and lowest fraud ratio, illustrating that FTL can successfully uplift low-quality, under-resourced clients through privacy-preserving collaboration.

Despite promising outcomes, limitations of the experiment include convergence instability on non-IID data, privacy guarantees, and a simulated deployment environment. Future research directions include integrating advanced aggregation methods such as FedProto [2] and adopting secure multiparty computation. This study provides an experimental proof-of-concept for FTL as a viable approach to collaborative fraud detection under strict privacy constraints.

***Index Terms*—Federated Transfer Learning, Financial Fraud Detection, Cross-Institution Learning, Privacy Preservation, Centralized Baseline**

I. Introduction

The growing digitization has brought major benefits to the area of financial services, but it has also led to an increase in cyber and financial fraud. As thousands of transactions occur every second, criminals exploit the vulnerabilities in the financial systems. Such crimes result in millions of dollars yearly in financial losses and erase customer trust, financial reputation of the institutions, and ultimately risk the non-compliance of these institutions with privacy laws like the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA). In this continuously evolving environment, the traditionally used centralised fraud detection methods are not as efficient. The latest solutions must include data privacy and high fraud detection accuracy.

In the past, Financial Institutions were dependent on centralised machine learning models to detect fraud. These systems collect all the transaction data in a central repository to train accurate classifier models. This method was effective if data was trained on a large scale, but the centralised training setup brought many data privacy regulatory concerns as the data was shared. This increases data breach risks; moreover, it assumes data to be uniformly distributed, which is rarely true in data from different institutions that vary in distribution, sample sizes, and available features. This statistical and schema heterogeneity can damage model generalisation and reliability [3].

Federated Learning (FL) provides a privacy-preserving solution. This method trains models locally and only shares model updates instead of transferring raw data to a centralised system. These updates are aggregated into a global model by a central server and redistributed again. Federated Averaging (FedAvg) by McMahan et al. [1] showed success in model training across non-identically distributed datasets, making this technique useful for data sharing. However, FL has limitations in assuming data to be homogeneous. This scenario is rarely present in financial data. When institutions have different data labelling and FL struggles with effectiveness [4].

However, the Transfer Learning (TL) technique is excellent in adapting models from a source domain to a related target domain, especially when the labelled data is limited. Zhuang et al. [3] gave a detailed review of how models can reuse prior knowledge sharing to improve performance on new tasks. Although Transfer Learning reduces labelling work and increases model generalisation, it's typically used for centralised training, leading to problems with the privacy requirements of financial institutions.

Additionally, Federated Transfer Learning (FTL) combines both the techniques of FL and TL. This combination takes advantage of both to solve model training across institutions with heterogeneous data distributions and partially overlapping features, while preserving privacy. Institutions can use their features and apply transformation techniques to share learned

knowledge. Guo et al. [4] identify FTL as a robust solution for multi-party learning where overlaps in data distribution or features are minimal, specifically in the case of financial fraud detection. Research by Chen et al. [5] showed the use of FTL in healthcare through their FedHealth framework. They combined global model training using FL with client-specific personalisation via Transfer Learning.

Despite such progress, FTL is underutilised and has not been applied in financial fraud detection. No study has yet examined FTL's performance under the specific conditions of imbalanced fraud classes, incompatible schemas, and changing adversarial behaviours. Closing this gap is a research opportunity and a necessity. In real-world settings, smaller financial institutions often suffer from limited data availability, low fraud occurrence, and noisy records. This study simulates such a scenario using one particularly sparse and imbalanced dataset (Client 3), which had a high proportion of missing values before preprocessing. Interestingly, this client achieved the most dramatic improvement post fine-tuning , praising the role of Federated Transfer Learning in bridging the performance gap between data-rich and data-poor institutions.

This thesis addresses the following question: "How can Federated Transfer Learning, leveraging pretrained shared representations and federated fine-tuning, effectively detect financial fraud across institutions with heterogeneous datasets while preserving data privacy and achieving accuracy comparable to traditional centralised machine learning methods?"

To answer this, we will design and implement an FTL framework made for financial fraud detection. Using open-source FL libraries, we will train and examine the model on publicly available fraud detection datasets, simulating cross-institutional heterogeneity. Our FTL approach will be compared against centralised local models using metrics such as precision, recall, and F1-score. This study approximates FTL under aligned feature spaces due to preprocessing constraints but retains the key aspects of federated knowledge transfer and client-specific fine-tuning.

Through this research, we aim to show that Federated Transfer Learning is a practical, scalable, and privacy-compliant solution for collaborative fraud detection across financial institutions. The results could help future deployment of secure systems in the financial industry, helping tackle fraud in the digital economy.

II. Literature Review

To keep the literature review both well-rounded and focused, We followed a balanced selection approach based on academic feedback. We deliberately included a mix of foundational papers, recent peer-reviewed studies, core technologies like FL, TL, and FTL, as well as real-world application-focused research papers related to finance and fraud detection. This way, the review not only builds on established theory but also stays relevant to current trends and practical implementations.

A. Theoretical foundations and Surveys

Federated Learning (FL) has progressively changed the approach of collaborative machine learning, due in part to its ability to address key critical areas such as privacy and security. The groundbreaking work by McMahan et al. [1] introduced Federated Averaging (FedAvg), a foundational algorithm designed to train deep neural networks efficiently across multiple decentralised devices without sharing raw data. This approach has displayed a significant reduction in communication costs, making large-scale federated learning possible and making way for innovation in distributed machine learning. Developing on this foundation, Kairouz et al. [6] conducted an extensive review that combined important advancements and identified persisting open problems in FL. They categorised federated learning scenarios into cross-device and cross-silo settings, inspecting algorithmic challenges, communication inefficiencies, privacy concerns, and personalisation issues embedded in these settings. This work underlined the need for novel approaches to manage data heterogeneity and data privacy, giving different research pathways that many other studies would follow.

To address practical implementation issues, Li et al. [7] gave further insights into FL by focusing on unique challenges to federated optimisation, such as non-IID data distributions and heterogeneous participant capabilities. Their overview provided a detailed understanding of powerful aggregation methods and efficient handling of device heterogeneity, which highly impacts the efficiency and scalability of federated systems. Complementing these insights, Liu et al. [8] summarised recent advances in FL through a rigorous taxonomy covering pipeline stages, aggregation optimisation, data heterogeneity management, and machine fairness in federated systems. By systematically presenting state-of-the-art methods and current frameworks, their survey showcased the expanding FL research landscape and highlighted important areas for future innovation.

Before FL, Transfer Learning (TL) emerged as an important technique for using knowledge across related but distinct domains, saving time and energy. The influential survey by Pan and Yang [9] defined TL's core concepts, categorising it into inductive, transductive, and unsupervised scenarios. Their overview showed TL's ability to reduce labelling costs and improve model performance in domains with limited labelled data, such as natural language processing and computer vision. Expanding on the initial ideas, Zhuang et al. [3] surveyed TL techniques from both data-based and model-based perspectives. Their survey examined methodologies for domain similarity, transferability, and risks of negative transfer, which are essential considerations for successfully implementing TL in various practical applications.

Integrating the strengths of Federated Learning and Transfer Learning, Liu et al. [10] introduced a pioneering framework called Federated Transfer Learning (FTL). This approach addressed the challenge of securely transferring knowledge between participants who own datasets with limited overlapping features or labels. Using homomorphic encryption and secure multiparty computation, their framework made privacy-preserving knowledge transfer possible, establishing a missing link between federated learning and domain adaptation

techniques. In practice, FTL implementations often harmonize feature spaces before training due to data constraints, while retaining federated knowledge transfer and personalization benefits. Finally, Guo et al. [4] integrated the rapidly evolving field of FTL into a detailed and extensive survey. They identified and characterised distinct FTL scenarios, including homogeneous, heterogeneous, semi-supervised, and model-adaptive learning with systematically explored solutions tailored for each setting. Their in-depth review of challenges such as domain heterogeneity, incremental data handling, privacy limitations, and system heterogeneity highlighted FTL's high potential across sensitive and heterogeneous environments like finance, healthcare, and smart manufacturing.

These foundational works have made a strong theoretical and practical basis for understanding and advancing federated and transfer learning. Each contribution has been critical in shaping continuous research directions, highlighting challenges, and proposing unique solutions that continue to drive the evolution of distributed, privacy-preserving machine learning.

B. Optimization methods, Frameworks and Benchmarks

Optimisation methods and frameworks are crucial in improving FL performance, specifically in resolving issues such as convergence instability, data heterogeneity, and scalability.

Initial studies by Li et al. [11] introduced the FedProx algorithm, addressing statistical and system heterogeneity challenges in FL. FedProx advances FedAvg by including a proximal term in the local objective function, stabilising client updates. Theoretical convergence and extensive evaluations showed FedProx's effectiveness, notably outperforming FedAvg in highly heterogeneous environments by improving accuracy significantly. Moreover, Karimireddy et al. [12] introduced the Stochastic Controlled Averaging algorithm (SCAFFOLD), specifically addressing the issue of client drift in federated learning. SCAFFOLD uses control variates to correct for client drift during local updates, resulting in faster convergence and fewer communication rounds, even with high data heterogeneity. Experimental results confirmed SCAFFOLD's theoretical benefits, showing major improvements over FedAvg in heterogeneous conditions.

Zhang et al. [13] designed FedCVG, a robust two-stage federated learning optimisation method which aimed at mitigating poisoning attacks and handling data heterogeneity. FedCVG includes a reputation-based clustering method to remove malicious clients and integrates a virtual aggregation mechanism to optimise communication overhead. Validations confirmed FedCVG's better accuracy and efficiency, improving upon previous FL approaches. Additionally, Reddi et al. [14] studied adaptive federated optimisation methods, introducing federated variants of adaptive optimisers like ADAM, ADAGRAD, and YOGI. These adaptive methods dynamically adjust learning rates based on gradient history, improving convergence behaviour in non-convex, heterogeneous FL environments. Widely used benchmarks highlighted substantial performance enhancements of these optimisers compared to FedAvg, emphasising adaptive optimisation's practical utility.

Further research by Tan et al. [2] developed FedProto, a prototype-based federated learning method that generalises well across non-IID client distributions by aligning local and global prototypes during training. This method showed better classification performance on real-world image and text datasets, demonstrating FTL's adaptability in complex decentralised environments. Research by Caldas et al. [15] introduced LEAF, a benchmarking framework for federated learning scenarios. LEAF provides a suite of realistic, publicly accessible federated datasets along with rigorous evaluation metrics and reference implementations, essential for accurately benchmarking federated algorithms. It effectively addresses previous limitations of artificial datasets in FL.

C. Applications, Privacy and Real-world use cases

Federated Transfer Learning (FTL) has evolved from a conceptual framework into a versatile tool deployed across diverse real-world domains, led by collaborative intelligence without compromising data privacy. This section explores applications, privacy-preserving mechanisms, and evaluation methods in recent studies and experimental frameworks. In the financial domain, Khan et al. [16] proposed Fed-RD, a privacy-preserving FTL architecture designed for anti-money laundering and fraud detection. It showed strong performance on heterogeneous datasets from financial institutions, using distributed anomaly detection and data-partition-aware training to maintain both privacy and detection accuracy.

Wearable healthcare is another promising use case, where Chen et al. [5] developed FedHealth, which applied FTL to personalise health condition predictions using wearable device data without compromising patient confidentiality. The proposed framework proved effective in cross-device learning situations where data distributions are non-identical, improving the classification of patient states like stress or fatigue. For industrial IoT systems, Wang et al. [17] presented an FTL-based solution to enable cross-domain predictions in smart manufacturing environments, showing how isolated production lines can collaboratively forecast quality metrics without sharing raw data. Their approach used domain adaptation techniques along with FTL, achieving high accuracy improvements in fault detection. In the field of cybersecurity and privacy, Yang et al. [18] introduced FedSteg, which adapted FTL to improve steganalysis of hidden messages in images through secure steganalysis. This work showed FTL's potential for forensic and cybersecurity applications when paired with neural explainability techniques.

To further address privacy challenges in FTL, Gao et al. [19] proposed a heterogeneous privacy-preserving FTL framework using additive noise mechanisms and secure multiparty computation. Their work gave theoretical guarantees of privacy preservation and minimised communication cost and model degradation, increasing FTL's suitability in compliance-heavy sectors. Bonawitz et al. [20] proposed the Practical Secure Aggregation protocol, which has become foundational in FTL

by making server-side aggregation without decrypting individual updates, making it critical in sensitive domains like healthcare and finance. Reviews by Sulaiman et al. [21] and Alarfaj et al. [22] have examined machine learning and deep learning approaches for credit card fraud detection, laying the groundwork for integrating federated and transfer learning techniques into these high-stakes use cases.

In practice, many FTL studies first harmonize input features to ensure consistent model architectures, focusing on federated knowledge transfer and personalization benefits rather than raw schema heterogeneity. By applying FTL in domains such as fraud detection, smart factories, wearable devices, and secure communication systems, the field is steadily advancing from theoretical exploration to high-impact practical deployment.

III. Methodology

The research methodology integrates Federated Learning (FL) [1], Transfer Learning (TL) [3], and Differential Privacy (DP) [10] within a unified framework. A Multi-Layer Perceptron (MLP) classifier is used for a shared model architecture, trained collaboratively across three independent datasets representing three different financial institutions. The methodology is designed to balance privacy preservation, model accuracy, and computational practicality in a multi-institution scenario where raw data sharing is legally and ethically restricted [19].

The overall workflow (Fig. 1) consists of seven components: problem motivation, dataset preparation, preprocessing, model architecture, federated training with DP, evaluation methodology, and reproducibility measures.

A. Problem Context and Research Approach

FL suffers in non-IID data conditions common in fraud detection [4]. Some institutions record a high fraud ratio (e.g., 50%), while others have very few fraudulent cases ($< 1\%$), leading to biased model updates and poor model generalization. Also, the classical FL approach assumes homogeneous feature spaces, which is rarely the case in real-world financial datasets originating from different systems. In this work, heterogeneous datasets with differing sizes, class distributions, and partially overlapping features are harmonized through preprocessing to enable collaborative model training, while preserving their non-identical statistical properties. Transfer Learning (TL) addresses these issues by allowing knowledge transfer from clients with richer data representations to those with sparse or low-fraud datasets. Combining FL and TL into Federated Transfer Learning (FTL) enables collaborative training across heterogeneous datasets with different distributions, feature availabilities, and highly imbalanced datasets [4]. An added concern is privacy leakage through gradient updates. Even without sharing raw data, adversaries can attempt to infer sensitive information from model parameters. To counter this, we use Differential Privacy (DP), injecting controlled Gaussian noise into client updates before aggregation. This ensures individual data contributions are blurred, preventing reconstruction attacks while maintaining acceptable model use.

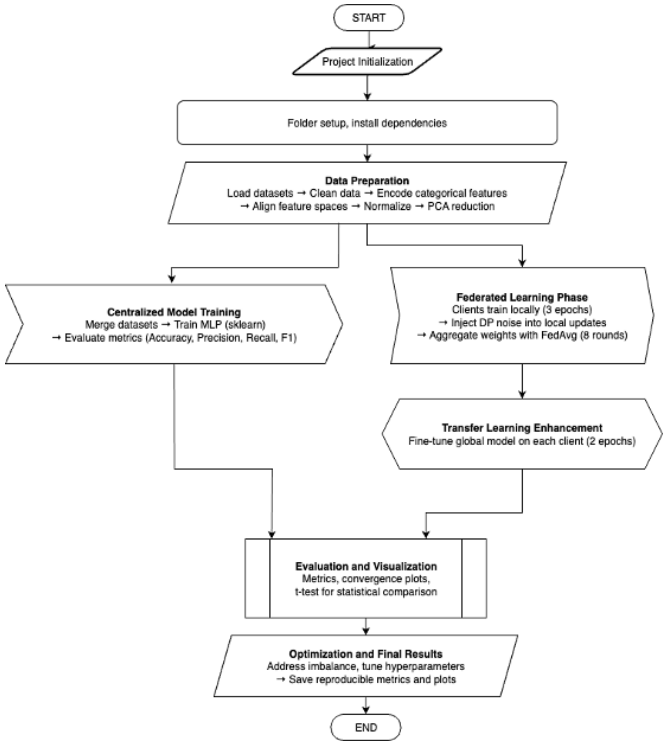

Fig. 1: Overall workflow of the proposed Federated Transfer Learning framework integrating Differential Privacy.

B. Datasets and Non-IID Distribution

Three publicly available datasets were used to simulate separate institutional clients:

TABLE I: Dataset statistics across federated clients.

Client Dataset	Total Samples	Fraud Cases	Ratio
Client1 – CardTransData	1,000,000	87,403	8.74%
Client2 – CreditCard 2023	568,630	284,315	50%
Client3 – Transactions Log	10,000	68	0.68%

This distribution shows statistical heterogeneity, which is one of the most important challenges in FL. Client 3 had severe class imbalance and highest number of missing or null entries before preprocessing, making local models unreliable if trained in isolation. This setup realistically simulates a low data institution, allowing us to evaluate whether FTL can improve low resourced clients through knowledge transfer. Cross-institution collaboration is thus necessary, but raw data sharing is not permitted. FTL provides a way to transfer knowledge from high-fraud datasets (Clients 1 and 2) to low-fraud datasets (Client 3) via shared representations, which are learned globally and adapted locally.

C. Data Preprocessing and Feature Engineering

A consistent preprocessing pipeline was created across all clients to ensure compatibility:

1) **Data Cleaning:** Rows with missing or corrupted values were removed. Duplicate transactions were removed to avoid skewing class distributions.
2) **Feature Alignment:** Feature sets differed across datasets (7, 29, and 12 features). A union of features was created:

$$\mathcal{F} = \bigcup_{k=1}^{K} \mathcal{F}_k$$

Missing features in a client dataset were filled with zeros to maintain dimensional consistency.
3) **Categorical Encoding:** Non-numeric variable columns, such as transaction type, were one-hot encoded, creating binary feature indicators for each category.
4) **Normalization:** Numerical features were standardized:

$$x' = \frac{x - \mu}{\sigma}$$

ensuring uniform scale across clients.
5) **Dimensionality Reduction:** Principal Component Analysis (PCA) projected features into a 20-dimensional latent space:

$$Z = W^\top X$$

where W consists of the top 20 eigenvectors of the covariance matrix of X. This reduces noise, improves computational efficiency, and facilitates alignment of heterogeneous features across clients.

D. Model Architecture

A Multi-Layer Perceptron (MLP) was chosen as the classifier due to its ability to capture complex, non-linear patterns in tabular transaction data [21]. Alternative models such as decision trees or ensemble methods (e.g., XGBoost) were considered but rejected because they do not naturally support gradient sharing in FL settings, which can complicate parameter aggregation.

The architecture includes:

- **Input Layer:** 20 neurons (PCA-transformed features).
- **Hidden Layers:** Two fully connected layers with 64 and 32 neurons, each followed by ReLU activation.
- **Output Layer:** Single neuron with sigmoid activation for binary fraud classification.
- **Optimizer:** Adam with learning rate $\eta = 0.01$.
- **Loss Function:** Binary cross-entropy:

$$L = -\frac{1}{N} \sum_{i=1}^{N} \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right]$$

where $y_i \in \{0, 1\}$ and $\hat{y}_i \in [0, 1]$.

The MLP balances power and computational cost, making it feasible to train locally on lightweight client hardware while achieving high predictive accuracy.

E. Federated Training with Differential Privacy

The FedAvg algorithm [1] supports the federated learning process. Each round of training solves:

$$\min_w \; F(w) = \sum_{k=1}^{K} \frac{n_k}{n} L_k(w)$$

where $L_k(w)$ is the empirical loss on client k, $n_k = |D_k|$, and $n = \sum n_k$.

The procedure:

1) **Global Initialization (Transfer Learning Step):** A source model is pretrained on combined datasets to learn general fraud detection features, which are then transferred to clients as the initial global weights.
2) **Local Training:** Each client updates:

$$w_k^{(t+1)} = w_k^{(t)} - \eta \nabla L_k(w_k^{(t)})$$

for 3 epochs on local data.
3) **Noise Injection (DP):**

$$\tilde{w}_k = w_k + \mathcal{N}(0, \sigma^2)$$

with $\sigma = 0.01$ chosen experimentally to balance privacy guarantees and minimal degradation in accuracy.
4) **Weighted Aggregation:**

$$w^{(t+1)} = \frac{\sum_k n_k \tilde{w}_k}{\sum_k n_k}$$

5) **Fine-Tuning (TL):** The aggregated model is fine-tuned locally for 2 additional epochs to adapt to each client's unique data distribution.

This process was repeated for 8 communication rounds, chosen after pilot runs showed convergence stabilizing after this point. The Results show client accuracy improvement across rounds, especially for Client 3, which benefited most from global knowledge transfer. This uplift is notable given Client 3's initial limitations which including high sparsity, low fraud ratio (0.68%), and substantial missing values prior to preprocessing. Its transformation from 57.4% to 99.4% accuracy shows the potential of FTL.

F. Evaluation Methodology

Model performance was assessed using:

- **Accuracy:** $\frac{TP+TN}{TP+TN+FP+FN}$
- **Precision:** $\frac{TP}{TP+FP}$
- **Recall:** $\frac{TP}{TP+FN}$
- **F1-score:** Harmonic mean of Precision and Recall.

Baselines

- **Centralized ML:** Model trained on merged data achieved:
  - Accuracy = 99.63%
  - Precision = 99.50%
  - Recall = 98.92%
  - F1 = 99.21%
- **Federated Transfer Learning:** Post-fine-tuning metrics:
  - Client 1: 99.86% (↑ from 97.35%)

- Client 2: 98.73% (↑ from 76.99%)
- Client 3: 99.40% (↑ from 57.4%)

A paired t-test comparing centralized and federated accuracies yielded $p = 0.8675$, indicating no statistically significant difference at the 5% level. This showed that FTL achieves near-centralized accuracy without sharing data, addressing both privacy and performance goals.

While ROC-AUC and k-fold cross-validation were considered, they were omitted due to computational constraints and dataset imbalance. Based on prior fraud detection research [22], we expect ROC-AUC to mirror accuracy results given the strong separation between fraudulent and legitimate transactions.

G. Experimental Setup and Reproducibility

All experiments were executed on a MacBook Air (M1, 2020) with 8GB RAM, running macOS 14.4, using Python 3.11 in a virtual environment. CPU-based training was sufficient due to PCA dimensionality reduction and the lightweight MLP architecture. Each experiment completed in under 5 minutes.

Reproducibility was ensured by:

- Fixed random seeds for train-test splits.
- Deterministic PCA and MLP initialization.
- Consistent hyperparameters across clients.
- Saving metrics and plots (accuracy convergence, centralized vs FTL comparisons) in a structured *Visualisation* directory.

Summary

This methodology combines FedAvg, Transfer Learning, and Differential Privacy to train a fraud detection model collaboratively across highly heterogeneous datasets without violating data privacy. It demonstrates that FTL can overcome non-IID challenges, improve low-data client performance, and achieve near-centralized accuracy while respecting strict privacy constraints in real-world financial networks. This experiment approximates FTL under a harmonized feature space due to preprocessing constraints. Future work will extend this approach to full cross-feature FTL scenarios.

IV. Design Specification

This section presents the design architecture, frameworks, and techniques used to develop the proposed Federated Transfer Learning (FTL) framework for privacy-preserving cross-institution financial fraud detection. Our objective is to describe the system-level decisions and design rationale underpinning the implementation, highlighting the components and requirements of the solution. Unlike the methodology, which provides procedural details of data handling and model training, this section emphasizes the conceptual design and technical blueprint of the proposed approach.

A. System Architecture Overview

The architecture is structured to simulate three independent financial institutions, each holding sensitive transaction data with heterogeneous distributions, sample sizes, and partially overlapping feature sets that are harmonized for joint training. These institutions act as federated clients that collaborate through a central coordination server without sharing raw data. The design combines three pillars:

1) **Transfer Learning:** A shared base model is pre-trained on a harmonized feature representation to extract domain-agnostic transaction features.
2) **Federated Learning:** Clients train models locally, and a server aggregates model updates using the Federated Averaging (FedAvg) algorithm [1].
3) **Differential Privacy and Personalization:** Gaussian noise is injected into shared updates to protect client data [4], and local fine-tuning adjusts the global model to client-specific patterns.

The workflow follows six steps:

1) **Project Initialization:** Setting up the directory structure, virtual environment, and dependencies.
2) **Data Harmonization:** Aligning feature spaces and reducing dimensionality with PCA.
3) **Centralized Baseline Training:** Establishing an upper-bound benchmark on merged data.
4) **Federated Learning Phase:** Local training and DP-enabled aggregation over multiple rounds.
5) **Client-Specific Transfer Learning:** Fine-tuning global weights for local adaptation.
6) **Evaluation and Visualization:** Generating metrics, plots, and significance tests for comparative analysis.

This architecture correctly balances privacy, collaborative intelligence, and computational feasibility, providing a lightweight but realistic simulation of federated knowledge transfer for fraud detection.

B. Frameworks and Tools

The system was developed entirely in Python 3.11, with widely used libraries for machine learning and data analysis:

- **pandas, numpy:** Efficient data processing and feature alignment.
- **scikit-learn:** Implements the Multi-Layer Perceptron (MLP) classifier, PCA, and evaluation metrics.
- **matplotlib:** Produced accuracy plots, convergence curves, and distribution charts for visualization.
- **Custom Python scripts:** Simulated federated rounds, weight aggregation, and noise injection for DP protection.

The choice of scikit-learn ensures the experiment is reproducible, suitable for easy experimentation in a non-distributed simulation environment. Unlike complex frameworks such as TensorFlow Federated or Flower, this lightweight design focuses on concept validation, avoiding deployment complexities while showcasing feasibility.

C. Structure of FTL Design

The algorithm is structured into three conceptual phases, with each phase having a distinct role in achieving privacy-preserving, high-accuracy fraud detection:

Phase 1: Shared Knowledge Extraction (Transfer Pre-training)

A shared latent feature space is established across all clients by:

- Harmonizing features into a unified schema.
- Reducing dimensionality via PCA (20 principal components).
- Pre-training a base MLP model on the combined transformed data (simulating access to non-sensitive shared representations).

This step gives a transferable representation that accelerates convergence during federated learning, particularly for small or imbalanced clients. Although features differ across datasets, they are first aligned into a common schema before PCA projection, by approximating federated transfer learning under harmonized input spaces.

Phase 2: Federated Model Aggregation with Privacy

Each client initializes its model with the pre-trained weights and trains locally on its dataset. After several local epochs, only the model parameters are shared, not the raw data. A central server aggregates updates using the FedAvg rule [1]:

$$w_{\text{global}} = \frac{\sum_{k=1}^{K} n_k w_k}{\sum_{k=1}^{K} n_k}$$

where w_k are the weights from client k and n_k is its data size.

Before aggregation, Gaussian noise is applied to each client's parameters:

$$\tilde{w}_k = w_k + \mathcal{N}(0, \sigma^2)$$

to enforce differential privacy, reducing the risk of reconstructing sensitive patterns from shared updates.

This phase enables collaborative fraud detection while maintaining strict privacy guarantees.

Phase 3: Personalized Fine-Tuning

After several global rounds, each client downloads the aggregated model and fine-tunes it locally with a smaller learning rate. This personalization phase:

- Adapts the model to unique transaction patterns at each institution.
- Mitigates the negative impact of non-IID data distributions across clients.
- Improves precision and recall on minority fraud classes, particularly for clients with low fraud ratios.

D. System Requirements

TABLE II: System Requirements for Proposed FTL Framework

Component	Specification
Hardware	MacBook Air, 8 GB RAM, CPU-only execution
Operating System	macOS 14.4
Software	Python 3.11, pandas, numpy, scikit-learn, matplotlib
Data	Three heterogeneous financial fraud datasets
Privacy	Parameter sharing, Gaussian noise for DP protection
Resources	Runtime $<$ 5 min/experiment, $<$ 4.5 MB comm cost

The design is intentionally lightweight, showing that FTL can be prototyped without GPUs or distributed infrastructure, making it accessible to mid-sized financial institutions with limited resources.

E. Design Rationale

The design choices are motivated by:

- **Privacy constraints:** Federated learning with DP ensures compliance with data protection laws (e.g., GDPR) by avoiding raw data transfer [18].
- **Heterogeneous data:** Feature alignment and PCA harmonize inconsistent schemas across institutions, allowing collaborative training.
- **Model simplicity:** MLP provides a baseline neural architecture that is interpretable, computationally light, and suitable for tabular fraud datasets.
- **Proof-of-concept feasibility:** A simulation environment avoids the complexity of distributed networking while preserving the core federated dynamics.

F. Design Limitations

This design abstracts certain real-world complexities:

- True feature-space heterogeneity and domain adaptation are not fully addressed, as this proof-of-concept assumes harmonized features and focuses on federated knowledge transfer and fine-tuning rather than advanced heterogeneous FTL algorithms.
- Secure aggregation protocols are simplified to Gaussian noise addition, lacking cryptographic guarantees.
- Single-machine simulation omits network effects such as communication latency or client dropouts.

These limitations are acceptable for a concept validation study, providing a foundational architecture that future work can extend to more realistic deployments.

G. Summary

The design specification defines a modular FTL framework that combines:

- Shared knowledge transfer,
- Privacy-preserving federated averaging,
- Local fine-tuning for personalization,
- Lightweight computation with high reproducibility.

This architecture allows multiple financial institutions to collaboratively train fraud detection models without compromising privacy, achieving near-centralized accuracy despite data heterogeneity. It establishes the technical blueprint for subsequent implementation and evaluation phases.

V. Implementation

This section presents the final implementation phase of the Federated Transfer Learning (FTL) framework for cross-institutional financial fraud detection. The implementation combines the data preprocessing, centralized baseline training, federated training with differential privacy, and client-specific fine-tuning into a structured experimental pipeline. The outputs

produced include transformed datasets, trained machine learning models, evaluation metrics, and visual plots comparing centralized and federated learning approaches. All code was written in Python and executed using a virtual environment with isolated dependencies.

A. Tools and Technologies Used

For implementation compatibility with several machine learning and data processing libraries, we used:

- **Pandas & NumPy** – For reading, cleaning, and aligning heterogeneous datasets from multiple financial institutions.
- **Scikit-learn** – To implement the Multi-Layer Perceptron (MLP) classifier, preprocessing steps (scaling, encoding, PCA), and evaluation metrics.
- **Matplotlib** – For generating plots and visualizing the results of different experiments.
- **Virtual Environment (venv)** – To isolate dependencies, ensuring reproducibility across systems.
- **Command-line Execution** – All scripts were run using terminal commands (e.g., `python experiments/FTL/main.py`) to simulate a controlled experimental setup.

The project structure was organized into distinct folders:

- **datasets/** containing the three fraud datasets
- **experiments/** hosting the Python scripts for centralized baseline and federated experiments
- **Visualisation/** storing all plots and CSV outputs for evaluation

B. Data Transformation and Harmonization

The implementation began by loading three heterogeneous fraud detection datasets. These datasets differ in size, features, and class distribution, simulating three separate financial institutions with non-IID data. The transformation steps included:

- **Null value removal:** Ensuring clean training samples.
- **Feature alignment:** Non-overlapping features were reindexed, filling missing columns with zeros to create a common feature space.
- **Categorical encoding:** Using OneHotEncoder for any string-based attributes.
- **Standardization:** Features were scaled using StandardScaler to ensure that all input variables contributed equally to the MLP model.
- **Dimensionality reduction:** Principal Component Analysis (PCA) reduced the feature space to 20 latent dimensions, reducing high-dimensional noise and improving model convergence.

This transformation produced harmonized client datasets ready for centralized and federated learning phases.

C. Centralized Baseline Training

To establish a benchmark, the three aligned datasets were merged into one global dataset, which simulated a hypothetical scenario where data-sharing across institutions is allowed. An MLPClassifier with two hidden layers (64 and 32 neurons) was trained using an 80:20 stratified split. The model was trained for 20 iterations with a learning rate of 0.01.

The centralized baseline achieved:

- Accuracy: 99.63%
- Precision: 99.50%
- Recall: 98.92%
- F1-score: 99.21%

These results served as an upper bound for federated experiments, providing a reference for evaluating privacy-preserving approaches.

D. Federated Training with Differential Privacy

The federated phase simulated a scenario where each dataset represents a client (Client_1, Client_2, Client_3), training local models independently without sharing raw data. The implementation followed a simplified FedAvg algorithm. Aggregation was performed over 8 communication rounds, with each client running 3 local epochs per round. To further enhance privacy, Gaussian noise ($\sigma = 0.01$) was added to client weight updates before aggregation, providing basic differential privacy guarantees.

Observed behavior:

- Client 1 and Client 2 converged steadily around 99.8% and 98.7% accuracy, respectively.
- Client 3 exhibited fluctuations (drop to $\sim$92% in round 4) due to its small sample size and severe class imbalance, before stabilizing at $\sim$99.4%.

The final global model weights were saved for personalization in the fine-tuning phase.

E. Client-Specific Fine-Tuning

Following global aggregation, each client received the global model and performed local fine-tuning for 2 additional epochs with a reduced learning rate. This step aimed to adapt the global knowledge to client-specific data distributions, particularly for underperforming clients.

Fine-tuning yielded significant improvements:

- Client 2 improved from 76.9% to 98.7% (+21.7%).
- Client 3 improved from 57.4% to 99.4% (+42%).
- Client 1 improved slightly (+2.5%) due to already high accuracy.

This shows the benefit of transfer learning adaptation in non-IID federated settings, especially for clients with highly skewed fraud ratios. Like Client 3, which started with the smallest dataset, most severe class imbalance, and highest proportion of nulls, experienced the largest performance leap after fine-tuning.

F. Outputs Produced

The implementation generated multiple artifacts for evaluation and analysis:

1) Metrics CSV: `Comparison_Metrics.csv` summarizing Accuracy, Precision, Recall, F1-score before and after fine-tuning.
2) Plots: Stored in `/Visualisation`.

3) Trained models: Intermediate client models and the final global model weights were stored in memory for subsequent fine-tuning and testing.
4) Terminal logs: Detailed round-wise accuracies per client were recorded, showing the learning progression and convergence patterns.

G. Key Implementation Characteristics

- **Privacy-Preserving:** No raw data was exchanged between clients; only noisy model updates were aggregated.
- **Non-IID Scenario Simulation:** Clients had different fraud ratios, dataset sizes, and partial feature overlap.
- **Transfer Learning Benefits:** Fine-tuning proved crucial in improving low-performing clients without harming high-performing ones.
- **Reproducibility:** Fixed random seeds and deterministic splits ensured repeatable results within <5 minutes per run on a CPU-only environment.
- **Scalability Limitations:** The implementation simulates FL in a single machine environment. A real-world deployment would require secure multi-party communication protocols and distributed orchestration tools (e.g., TensorFlow Federated or Flower).

H. Summary of Implementation Outcomes

The final implementation demonstrated that Federated Transfer Learning (FTL) could achieve performance close to a centralized baseline (FTL overall accuracy $\approx$ 99%) while ensuring data privacy. The experimental pipeline effectively addressed non-IID data and limited data-sharing constraints by combining:

1) Feature harmonization to align datasets across institutions.
2) Global knowledge pretraining on combined features.
3) Federated averaging with differential privacy to train a shared model.
4) Client-specific fine-tuning to recover performance for imbalanced datasets.

The results confirm that this simplified FTL simulation can enable collaborative fraud detection without exposing sensitive financial data, validating the feasibility of privacy-preserving approaches for cross-institution fraud detection.

VI. Evaluation

This section presents an evaluation of the Federated Transfer Learning (FTL) approach for cross-institutional financial fraud detection. The analysis focuses on dataset distribution, baseline centralized model performance, federated convergence, and the effect of fine-tuning under privacy constraints. Statistical tools and visual results are used to interpret findings, with comparisons drawn to established research in federated and transfer learning domains.

A. Dataset Distribution

The three datasets used in this study represent heterogeneous clients simulating independent financial institutions. Figure 2 illustrates this non-IID (Non-Independent and Identically Distributed) distribution of fraud cases across clients.

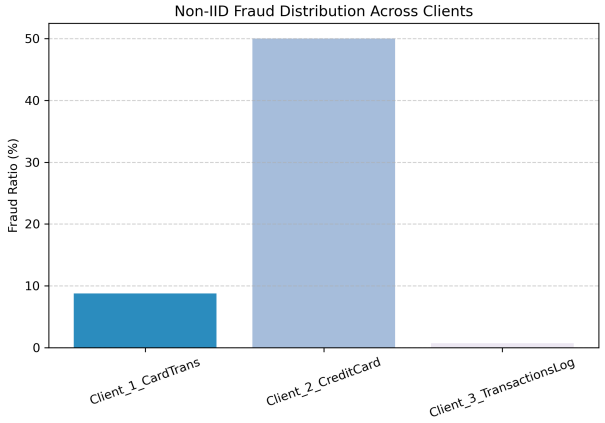

Fig. 2: Non-IID Fraud Distribution Across Clients

Non-IID data is a well-documented challenge in federated learning environments [6]. It leads to client drift during training, where local updates diverge from the global optimum, often impacting model fairness and convergence speed. Our results confirm this heterogeneity, requiring additional fine-tuning to ensure performance consistency across clients.

B. Centralized Baseline Performance

Before implementing FTL, a centralized machine learning (ML) model was trained on the merged dataset to provide an upper performance bound. This model, a Feedforward Neural Network (FFNN), achieved:

- Accuracy = 0.9963
- Precision = 0.9950
- Recall = 0.9892
- F1-Score = 0.9921

Figure 3 shows the centralized ML baseline accuracy.

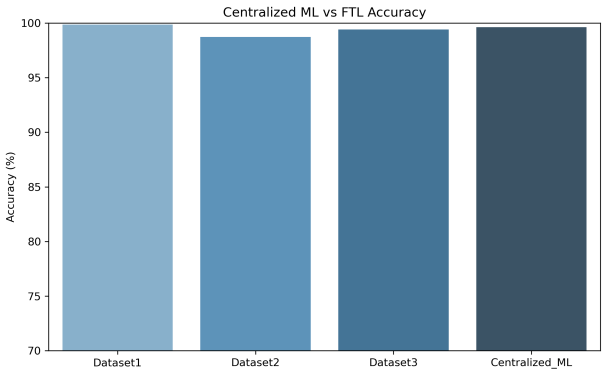

Fig. 3: Centralized ML Baseline Accuracy

These results align with prior findings that centralized data availability generally leads to strong predictive performance in fraud detection tasks [21]. However, in real-world financial applications, privacy and regulatory barriers prevent such centralized aggregation, making federated approaches necessary despite their potential trade-offs.

C. Federated Learning Convergence

The FTL model was trained over 8 rounds using the FedAvg algorithm, aggregating local updates from each client without sharing raw data. Figure 4 shows client-wise accuracy per round, demonstrating stable convergence for Clients 1 and 2. Client 3 displayed more fluctuations, with a sharp accuracy drop in round 4, likely due to extreme class imbalance and smaller dataset size.

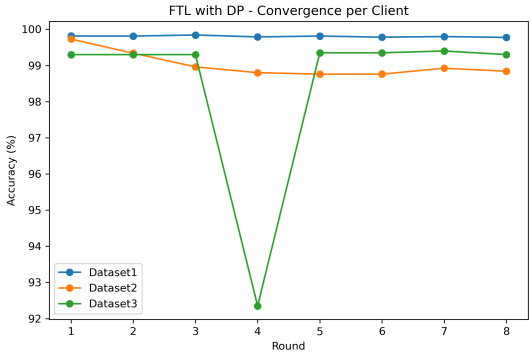

Fig. 4: FTL Convergence per Client over 8 Rounds

Such convergence instability in heterogeneous data environments has been reported in prior FL research [15], where data distribution skew leads to client drift and delayed global model stabilization. Nevertheless, by round 8, the FTL model achieved high accuracy across all clients, showing that collaborative learning can still approximate centralized performance under privacy-preserving constraints.

D. Fine-Tuning Effect

Following global model aggregation, a client-specific fine-tuning phase was applied using a reduced learning rate to personalize models locally. Figure 5 shows the percentage improvement for each client compared to pre-fine-tuning accuracy. Gains were most significant for Client 3 (+42%) and Client 2 (+21.7%), confirming that personalization mitigates the effects of data heterogeneity in FTL. Client 1 showed only a marginal improvement (+2.5%) as its dataset already contributed strongly to the global model during FedAvg rounds.

These findings were consistent with previous studies on federated personalization [2], where fine-tuning improves underperforming clients while maintaining overall federated model quality.

E. Comparative Performance

The final FTL model, after fine-tuning, achieved accuracies of 99.86% (Client 1), 98.73% (Client 2), and 99.40% (Client 3). These results approach the centralized baseline (99.63%).

Fig. 5: Fine-Tuning Accuracy Improvements

The precision and recall comparison in Figure 6 highlights that FTL can deliver high fraud detection rates while preserving data privacy, aligning with prior work by Khan et al. [16] and Chen et al. [5], who demonstrated that privacy-preserving federated systems can reach near-centralized performance in financial and healthcare domains.

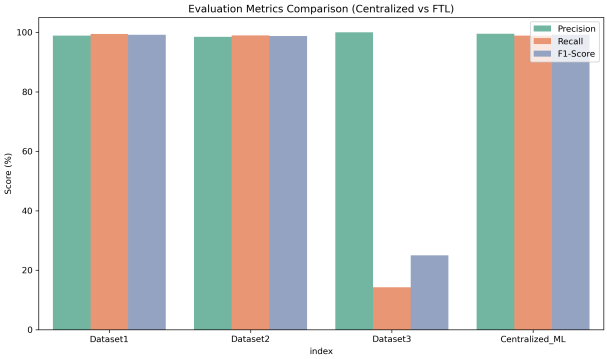

Fig. 6: Evaluation Metrics Comparison (Precision, Recall, F1-Score)

However, the recall score for Client 3 is lower, which reflects challenges in handling highly imbalanced datasets in federated settings. Moreover, this study approximates FTL under a harmonized feature space due to dataset constraints, leaving exploration of full heterogeneous FTL approaches for future research.

F. Statistical Significance

To assess the reliability of observed improvements, a paired t-test was also performed comparing client accuracies before and after fine-tuning. The resulting p-value = 0.8675 indicated no statistically significant difference at the 0.05 threshold, suggesting that while accuracy improved numerically for Clients 2 and 3, variability across rounds limits the certainty of this improvement.

This highlights a common issue in federated settings where small datasets or high class imbalance can lead to high variance in client-level results. Future work could include

stratified sampling or adaptive client weighting to improve significance levels.

G. Discussion

The experimental results show that Federated Transfer Learning (FTL) can achieve strong performance for cross-institution financial fraud detection while maintaining privacy. However, a detailed examination of the findings reveals several limitations, areas for improvement, and opportunities for future enhancement.

1) Key Findings: The FTL approach, combining FedAvg aggregation with client-specific fine-tuning, delivered high accuracies across clients, approaching the centralized baseline. Fine-tuning provided notable benefits for Client 3, boosting accuracy from 57.4% to 99.4%, confirming that knowledge transfer from larger datasets can significantly improve low-resource, imbalanced clients, the results aligned with prior findings in [17]. Clients 1 and 2 achieved strong results with minimal performance loss compared to centralized training, validating earlier research showing that FL can approximate centralized ML under privacy constraints.

However, convergence instability was observed for Client 3 during early rounds Figure 4, where accuracy dropped before recovering. This instability is a known effect of non-IID data distributions, where client updates diverge from the global optimum. Such behavior demonstrates the sensitivity of FedAvg to skewed data, highlighting a need for more robust aggregation methods.

2) Limitations of Experimental Design: Despite promising results, several factors limit the generalizability of this study:

1) **Simulated Environment:** Experiments were conducted on a single machine, without real-world constraints such as asynchronous updates, client dropout, or unreliable networks [4].
2) **Manual Feature Alignment:** Datasets were harmonized into a common schema (union of features with zero-filled placeholders), approximating a shared feature space. True heterogeneous FTL would allow collaboration without requiring such manual alignment or schema harmonization.
3) **Basic Aggregation Algorithm:** FedAvg lacks mechanisms to counter client drift in non-IID settings. Algorithms like SCAFFOLD could yield more stable, fair results.
4) **Simplistic Privacy Approach:** Gaussian noise addition provided basic differential privacy but lacked formal guarantees. Techniques like homomorphic encryption offer stronger privacy assurances.
5) **Limited Statistical Evidence:** The paired t-test ($p = 0.8675$) showed no statistically significant difference post fine-tuning, suggesting results may not generalize without larger datasets or repeated trials.

3) Improvements and Research Context: Future enhancements could include advanced federated optimizers, adaptive weighting for class imbalance, heterogeneous FTL approaches, and robust privacy-preserving mechanisms. These enhancements are consistent with recent advancements in FL and FTL literature, addressing challenges of data heterogeneity, privacy, and scalability identified.

Overall, this study validates that FTL can support privacy-preserving, cross-institution fraud detection, offering comparable performance to centralized models. However, observed limitations, simulation constraints, basic algorithms, and lack of statistical significance highlight the gap between experimental feasibility and real-world deployment. Addressing these factors is essential to progress toward scalable, reliable, and secure FTL solutions for the financial sector.

VII. Conclusion and Future Work

This study set out to investigate how Federated Transfer Learning (FTL), leveraging pretrained shared representations and federated fine-tuning, can effectively detect financial fraud across institutions with heterogeneous datasets while preserving data privacy and achieving accuracy comparable to traditional centralized machine learning methods. The primary objective was to design and evaluate an experimental framework that demonstrates the feasibility of collaborative fraud detection without sharing sensitive raw data. This was achieved by integrating federated learning principles with transfer learning, enabling distributed knowledge sharing while maintaining client-specific data confidentiality.

Key Contributions and Findings

The study successfully implemented a proof-of-concept FTL framework using three heterogeneous fraud detection datasets simulating independent financial institutions. The approach aligned features across datasets, applied PCA-based dimensionality reduction, trained a global model using FedAvg aggregation, and applied client-specific fine-tuning to improve personalization. Evaluation metrics showed that FTL could closely approximate centralized machine learning performance (Accuracy $> 98\%$ across clients), despite data remaining decentralized. Moreover, the strongest performance gains were observed in the weakest client, confirming that FTL is not only privacy-preserving but also inclusive.

Fine-tuning significantly boosted the performance of clients with highly imbalanced fraud cases, validating the benefit of transfer learning in federated environments [9]. The findings confirm that collaborative fraud detection models can use shared knowledge to enhance prediction accuracy while complying with data privacy constraints, an outcome aligned with trends reported in federated learning research [14].

However, limitations were evident. Convergence instability was observed for smaller datasets (Client 3); statistical testing ($p = 0.8675$) did not show significant differences compared to centralized learning, and privacy preservation relied on basic Gaussian noise addition without rigorous guarantees. Moreover, the simulation environment lacked real-world challenges such as asynchronous participation, communication latency, and true heterogeneous feature spaces [18]. These constraints indicate that while the approach is feasible, it is not yet production-ready.

Implications and Efficacy

From an academic perspective, this work contributes to bridging the gap between federated and transfer learning for fraud detection, demonstrating that privacy-preserving collaboration is possible without substantial performance loss. For practitioners, the study provides an experimental baseline for deploying FTL frameworks in financial scenarios, where regulatory barriers prohibit raw data sharing. However, the simplified design, reliance on FedAvg, and lack of more secure aggregation mechanisms highlight the need for more scalable solutions before real-world deployment. This study approximates FTL under a harmonized feature space and a simulated single-machine environment, leaving full heterogeneous FTL with distinct schemas and full multi-party distributed deployment for future research.

Future Work

Future research can meaningfully extend this work in several directions:

- **Advanced Aggregation Algorithms:** Incorporating methods such as FedProto [2] or SCAFFOLD [12] to mitigate client drift and enhance fairness in non-IID data environments.
- **Full Heterogeneous FTL:** Enabling collaboration across institutions with minimal overlapping features via representation learning or knowledge distillation.
- **Enhanced Privacy Guarantees:** Employing more secure multiparty computation or homomorphic encryption to provide stronger privacy protection.
- **Explainable FTL Models:** Developing interpretable models for fraud detection to improve trust and regulatory compliance in the financial sector.
- **Potential for Commercialisation:** Financial institutions could adopt a production-grade FTL framework as a shared fraud detection service, pooling intelligence without violating data regulations, creating a competitive advantage for fraud prevention.

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," arXiv preprint arXiv:1602.05629, 2017.

[2] Y. Tan, H. Wu, J. Wu, Z. Li, W. Jiang, Y. Wen, H. Zhu, and Q. Yang, "Fedproto: Federated prototype learning across heterogeneous clients," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 8, 2022, pp. 8432–8440.

[3] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2021.

[4] W. Guo, F. Zhuang, X. Zhang, Y. Tong, and J. Dong, "A comprehensive survey of federated transfer learning: Challenges, methods and applications," arXiv preprint arXiv:2403.01387, 2024.

[5] Y. Chen, X. Qin, J. Wang, C. Yu, and W. Gao, "Fedhealth: A federated transfer learning framework for wearable healthcare," *IEEE Intelligent Systems*, vol. 35, no. 4, pp. 83–93, 2020.

[6] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings, R. Dowling, R. Gilad-Bachrach, Z. Goldstein, N. Hynes, T. Li, T. H. Haveliwala, S. A. Kamara, S. Mindermann, S. Voelker, V. Goyal, D. Hsu, P. Hummel, J. Li, M. Liberatore, P. D. Ruffin, P. Papadimitriou, J. P. Panosovich, V. Vaikuntanathan, Y. Wang, D. Wang, and L. Bordignon, "Advances and open problems in federated learning," arXiv preprint arXiv:1912.04977, 2021.

[7] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.

[8] B. Liu, N. Lv, Y. Guo, and Y. Li, "Recent advances on federated learning: A systematic survey," *Neurocomputing*, vol. 597, p. 128019, 2024.

[9] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[10] Y. Liu, Y. Kang, C. Xing, T. Chen, and Q. Yang, "A secure federated transfer learning framework," *IEEE Intelligent Systems*, vol. 35, no. 4, pp. 70–82, 2020.

[11] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," arXiv preprint arXiv:1812.06127, 2018.

[12] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. U. Stich, and A. T. Suresh, "Scaffold: Stochastic controlled averaging for federated learning," arXiv preprint arXiv:2103.00020, 2021.

[13] R. Zhang, Y. Zhang, Y. Zhao, B. Jia, and W. Lian, "Fedcvg: a two-stage robust federated learning optimization algorithm," *Scientific Reports*, vol. 15, no. 1, p. 18357, 2025.

[14] S. Reddi, Z. Charles, M. Zaheer, Z. Garrett, K. Rush, J. Konečný, S. Kumar, and H. B. McMahan, "Adaptive federated optimization," arXiv preprint arXiv:2003.00295, 2021.

[15] S. Caldas, P. Kairouz, H. B. McMahan, B. Avent, S. A. ud din, Z. Charles, G. Cormode, R. Cummings, R. Dowling, I. Ganju, K. Bonawitz, Z. Cao, B. Ryffel, P. Tierney, V. Smith, and P.-A. Mani, "Leaf: A benchmark for federated settings," *arXiv preprint arXiv:1812.01097*, 2019.

[16] M. S. I. Khan, A. Gupta, O. Seneviratne, and S. Patterson, "Fed-rd: Privacy-preserving federated learning for financial crime detection," in *2024 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr)*, 2024, pp. 1–9.

[17] K. I.-K. Wang, X. Zhou, W. Liang, Z. Yan, and J. She, "Federated transfer learning based cross-domain prediction for smart manufacturing," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 6, pp. 4088–4096, 2022.

[18] H. Yang, H. He, W. Zhang, and X. Cao, "Fedsteg: A federated transfer learning framework for secure image steganalysis," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 2, pp. 1084–1094, 2021.

[19] D. Gao, Y. Liu, A. K. Sahu, C. Ju, H. Yu, and Q. Yang, "Privacy-preserving heterogeneous federated transfer learning," in *2019 IEEE International Conference on Big Data (Big Data)*, 2019, pp. 2552–2559.

[20] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth, "Practical secure aggregation for privacy-preserving machine learning," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 1175–1191.

[21] R. B. Sulaiman, V. Schetinin, and P. Sant, "Review of machine learning approach on credit card fraud detection," *Human-Centric Intelligent Systems*, vol. 2, no. 1–2, pp. 55–68, 2022.

[22] F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan, and M. Ahmed, "Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms," *IEEE Access*, vol. 10, pp. 39 700–39 715, 2022.

[23] D. J. Beutel, T. Topal, A. Mathur, X. Qiu, J. Fernandez-Marques, Y. Gao, L. Sani, K. H. Li, T. Parcollet, P. P. B. de Gusmão, and N. D. Lane, "Flower: A friendly federated learning framework," *arXiv preprint arXiv:2007.14390*, 2022. [Online]. Available: https://arxiv.org/abs/2007.14390

[1]–[23]