TrialMitra: Federated AI for Patient-Friendly Clinical Trials

Ananya S
Artificial Intelligence and Machine Learning
NITTE (Deemed to be University)
Bengaluru, Karnataka, India
ananya.athreya04@gmail.com

Chinmayi K
Artificial Intelligence and Machine Learning
NITTE (Deemed to be University)
Bengaluru, Karnataka, India
chinmayikp2004@gmail.com

Nikhita R
Artificial Intelligence and Machine Learning
NITTE (Deemed to be University)
Bengaluru, Karnataka, India
nikhitaraj1810@gmail.com

Niriksha K
Artificial Intelligence and Machine Learning
NITTE (Deemed to be University)
Bengaluru, Karnataka, India
nirikshakrishna113@gmail.com

Piyush Kumar Pareek
Professor and Head, Dept. of AIML
NITTE (Deemed to be University)
Bengaluru, Karnataka, India
piyush.kumar@nmit.ac.in

Abstract*—The challenge of recognizing and predicting patient suitability for clinical trials remains a significant impediment in clinical research with high chances of enrollment delays in most clinical trials. Traditional techniques involve the centralization of sensitive patient health information and thus raise several security concerns, such as compliance with frameworks such as HIPAA and GDPR. In this paper, we present TrialMitra, a secure and privacy-aware platform for clinical trial recruiting. In TrialMitra, Federated Learning (FL) is used to carry out decentralized prediction of eligibility across multiple hospitals, thereby preserving raw patient data. The system uses a Federated Averaging (FedAvg) framework based on logistic regression optimized using Stochastic Gradient Descent. In addition, we ensure local privacy using Local Differential Privacy (LDP) by adding Gaussian noise with ($\sigma = 0.01$). We conduct our experiment on a simulated scenario involving a four-hospital network having highly heterogeneous (non-IID) datasets, generated through the use of Dirichlet process distribution ($\alpha = 0.1$). Our experimental results, using 100,000+ records from ten different diseases, show that after 15 communication rounds, TrialMitra achieves an accuracy rate of 82.92%, precision of 84.78%, recall of 84.99%, and F1-Score of 84.88%. The platform is additionally enriched with an AI-powered multilingual chatbot supporting four Indian languages, along with a real-time trial recommendation feature linked with the ClinicalTrials.gov database. It is worth mentioning here that the study results reveal that federated learning systems have the potential to provide reliable predictions just like centralized systems.

***Index Terms*—Federated Learning, Clinical Trial Recruitment, Privacy-Preserving Machine Learning, Differential Privacy, Healthcare AI, FedAvg, Patient Eligibility Prediction**

I. Introduction

Evidence-based progress within medicine can be seen as being founded on the results of clinical trials. Nevertheless, one may find out that the identification and recruitment process remains a challenge and costly matter. According to the literature review, up to 80% of clinical trials fail to enroll a required number of people, and about 30% of Phase III clinical trials never come to completion since there are no patients involved in such a research [1]. This leads to significant costs associated with drug development, which may range between $600 million and $2.6 billion per approved medication [2]. Moreover, the delay in treatment provision is another issue.

First of all, it should be mentioned that medical records remain isolated at the moment, i.e., one finds health information distributed throughout multiple hospitals and clinics where specific privacy regulations are in place. In particular, these include HIPAA [3] and GDPR [4], according to which there is no possibility to aggregate any private data into one database. Thus, while there is a constant struggle between the necessity of screening patients in order to include them in the research program and ethical concerns related to confidentiality.

In this context, Federated Learning (FL) [5] stands out as a groundbreaking technique for addressing the same challenge. Leveraging the capability to train machine learning algorithms in a decentralized fashion using multiple datasets while not moving any data at all, FL allows healthcare organizations to

collaboratively enhance the model used for eligibility determination while ensuring that no individual's medical information is revealed.

This paper presents **TrialMitra**—a complete and privacy-conscious system aimed at transforming the process of clinical trial enrollment. The key elements of our system include:

- A Federated Learning structure implemented using the Flower library [6], deploying the FedAvg approach among four client hospitals with very heterogeneous data distribution.
- Local Differential Privacy (LDP) strategies utilizing Gaussian noise addition to the model weight during communication process, providing strong mathematical privacy guarantees.
- A multilingual conversational bot based on the Groq Large Language Model (LLM) inference engine, able to interact with users in English, Hindi, Kannada, and Telugu languages.
- A fast trial search system linked to the ClinicalTrials.gov database, intended to match potential participants to appropriate studies.
- An intuitive Streamlit-based web interface allowing researchers to monitor the application's progress and analyze patient data.

The remainder of this paper is organized as follows: Section II reviews related work. Section III describes our approach and architecture. Section IV discusses experimental conditions. Section V describes the results, while Section VI contains the conclusions and suggestions for future work.

II. Related Work

A. Federated Learning in Healthcare

Federated Learning was initially introduced by McMahan *et al.* [5] through the use of FedAvg, which aims at training complex models over a network of distributed devices. Since then, the technique has been widely used in the context of medicine. Rieke *et al.* [7] gave a detailed review of FL use cases for clinical imaging, genomics, and EHR data and stressed on the ability of FL to address institutional data silos in a secure manner. In terms of implementation, Sheller *et al.* [8] used FL in segmenting brain tumor images from various institutions, achieving results comparable with those achieved by training a model on centralized datasets. Moreover, Li *et al.* [9] used local batch normalization techniques to address heterogeneity in healthcare data.

B. Clinical Trial Recruitment Systems

In the past, the selection process for clinical trials has relied on inflexible decision-making frameworks designed to screen patients based on strict guidelines for inclusion and exclusion criteria [10]. The development of more advanced techniques has led to the use of NLP to automatically extract trial eligibility information from free-form clinical notes [11]. These modern technologies usually operate within a centralized framework and require access to patient records. Although Beck *et al.* [12] developed an automated screening system driven by EHRs and employing machine learning methods, this research did not take into account the constraints imposed by cross-institutional scenarios. According to our knowledge, TrialMitra represents the first attempt to integrate distributed eligibility determination with ongoing patient interaction in multiple languages and synchronization with clinical trial registries.

C. Privacy-Preserving Machine Learning

The notion of differential privacy (DP) was first proposed by Dwork et al. [13], providing a strict mathematical framework that ensures the behavior or output of an algorithmic model will not be significantly influenced by the inclusion or exclusion of one person's data. In the context of distributed systems, Local Differential Privacy (LDP) necessitates the addition of noise at the origin node before any information dissemination [14]. In the realm of deep learning, Abadi et al. [15] have developed DP-SGD, which restricts the influence of individual gradient steps and incorporates carefully crafted Gaussian noise. In TrialMitra, we opted for a more effective approach through the addition of Gaussian perturbations to the model parameters computed locally.

III. System Architecture and Methodology

A. System Topology

TrialMitra works under the framework of centralized coordination in tandem with decentralized databases. Such an arrangement involves four distinct clinical organizations (clients) with individual health data. A central server is responsible for the process of optimizing while being oblivious of the data pertaining to the patients. For managing the high volume of gRPC messages exchanged between the coordinator and network nodes, the Flower library [6] comes into play.

The operation process at each time r for the synchronization round is described below:

1) The server sends out the weights $\vartheta^{(r)}$ currently in the network to the $K = 4$ hospitals taking part in this activity.
2) After receiving the above, each hospital k runs the training for $E = 5$ rounds using their own data sets D_k with the help of the stochastic gradient optimizer.
3) To protect the results from being intercepted before being passed back, hospitals add Gaussian noise $\tilde{\vartheta}_k^{(r)} = \vartheta_k^{(r)} + \mathsf{N}(0, \sigma^2 I)$, with a dispersion parameter σ set to 0.01.
4) Finally, the central node aggregates the noisy gradients according to the FedAvg technique as follows:

$$\vartheta^{(r+1)} = \sum_{k=1}^{K} \frac{n_k}{n} \tilde{\vartheta}_k^{(r)} \tag{1}$$

Since $n_k = |\mathsf{D}_k|$ denotes the number of records contained in clinic k, the overall number of accessible records implies $n = \sum_{k=1}^{K} n_k$.

B. Data Creation and Cleaning pipelines

In order to accurately model a multi-center healthcare partnership, we generate virtual disease records distributed across all four sub-networks along with an individual validation register.

1) Creating Diverse Disease Prevalence: A genuine medical environment inherently exhibits non-uniform levels of disease prevalence owing to geographical and racial differences. In order to statistically represent these external factors, we assign disease categories via a Dirichlet process:

$$p_k \sim \mathrm{Dir}(\alpha \cdot \mathbf{1}_C), \quad \alpha = 0.1 \tag{2}$$

With $C = 10$ distinct physiological conditions, the small scalar $\alpha = 0.1$ deliberately imposes a grossly imbalanced and disconnected data set. Thus, branch k is inherently endowed with its own list of patients based on the probability distribution in p_k.

2) Dimensionality of Feature Set: The feature set $x \in \mathbb{R}^7$ for any one individual comprises seven key features, namely: chronological age, biological gender, systemic blood pressure, blood glucose level, primary illness classification, study participation motivation (1 to 5), and patient commitment level (1 to 5). On the other hand, the output $y \in \{0, 1\}$ is essentially a boolean function that determines whether or not an individual is eligible for the study based on strict boundary conditions.

3) Data Integrity Routines: In order to make the simulation practical, the first set of data contains deliberately introduced noise, where missing values exist for ages at a rate of 5%, vascular at 3%, and sugar at 2%. Additionally, we included 1% of biological inconsistencies (such as age ¿ 120 years and BP ¿ 250), ambiguous data types (for example, swapping "M", "Male", and "male"), and a duplication rate of 2%. To address this, the first phase of linear sanitization begins, which includes (1) removal of duplicates, (2) standardizing strings, (3) imputing missing values through medians, and (4) eliminating illogical outliers.

C. Execution of the Collaborative Protocol

1) Algorithmic Foundation: The computational heart deployed on each collaborating node is embodied as a Logistic model trained using Stochastic Gradient Descent (`SGDClassifier`, exclusively specified by `loss='log_loss'`), equivalent to traditional cross-entropy loss. Our preference for this precise architecture stemmed from:

- **Analytical Intuitiveness**: The explicit clarity exhibited within logistic methods ensures comprehensible parameter coefficients, an essential requirement in medical applications.
- **Efficient Bandwidth Utilization**: Node-level adjustments strictly exchange $d + 1$ parameters (containing 7 gradient parameters for features along with 1 bias), substantially reducing communication bandwidth.
- **Guaranteed Convergence**: Constrained by convex optimization principles, FedAvg applied to logistic frameworks guarantees convergence towards the optimal point [16].

2) Isolated Node Refinement: During each round of any cooperation cycle, the entire collection of individualized facilities executes five (i.e., $E = 5$) rounds of gradient updates for their own local batches:

$$\vartheta_k^{(r)} \leftarrow \vartheta_k^{(r)} - \eta \nabla L(\vartheta_k^{(r)}; B_k) \tag{3}$$

Here, η stands for the learning rate (`optimal`, per scikit-learn's SGD routine), while L indicates the logistic loss function applied on the shuffled batch B_k. To address class imbalance, we apply the option `class_weight='balanced'`, which effectively weights samples according to class prevalence.

3) System Consistency: At the higher coordination level, the system is driven using basic principles of federated averaging, with all participants involved during each synchronization phase (`fraction_fit = 1.0`) within the total number of communication steps, denoted by $R = 15$.

D. Implementing Local Differential Privacy

In order to strengthen the security of data borders on the level of the individual facility, the Gaussian disturbance is intentionally added to the set of the calculation criteria before creating an uplink connection with the central server:

$$\tilde{\vartheta}_k = \vartheta_k + N(0, \sigma^2 I), \quad \sigma = 0.01 \tag{4}$$

With the help of such a disturbance, the local network computes only the distorted form of the real learning process that disables any efforts to reconstruct patient information. The selected value of the standard deviation $\sigma = 0.01$ provides an optimal balance between usefulness and security.

E. Polyglot AI Module

The platform features a sophisticated conversational interface operating on the `llama-3.3-70b-versatile` architecture, accelerated by the Groq API. Making use of the functionality provided by the `deep-translator` library, this assistant can communicate seamlessly in the four main languages spoken in India: English, Hindi, Kannada, and Telugu. The responses generated by this system consist of logically constructed and empathic recommendations regarding clinical trials, qualifications, risks, and instructions for registration.

F. Real-Time Alignment Engine

Upon meeting the set criteria for eligibility established by our AI system, the program automatically accesses the public database provided by the website *ClinicalTrials.gov*. This engine seeks to find clinical trials currently being conducted and recruiting patients according to their physiological characteristics and condition. As a result, we receive a tailored list of top ten clinical trials with all relevant details included.

IV. Experimental Setup

A. Dataset Configuration

Table I summarizes the dataset characteristics across the four hospital nodes and the global test set.

TABLE I: Dataset Configuration Across Hospital Nodes

Node	Patients	Distribution	Split
Hospital 1	20,000–35,000	Dir(α=0.1)	80/20
Hospital 2	15,000–25,000	Dir(α=0.1)	80/20
Hospital 3	35,000–45,000	Dir(α=0.1)	80/20
Hospital 4	25,000–35,000	Dir(α=0.1)	80/20
Global Test	10,000	Uniform	100% test

The entire ecological system has a number larger than 100,000 file cases for ten different diseases, namely: diabetes mellitus, hypertensive, cardiovascular disease, oncological, psychiatric, pulmonary, arthritis, obesity over BMI, chronic headache, and kidney impairment. As each node independently determines their own demographic distribution using a Dirichlet processor with an alpha value of 0.1, there is extreme statistical fragmentation observed in the network.

B. Hyperparameters

Table II lists the key hyperparameters used in our experiments.

TABLE II: Training Hyperparameters

Parameter	Value
Communication rounds (R)	15
Local epochs (E)	5
Number of clients (K)	4
Fraction fit / evaluate	1.0 / 1.0
Learning rate schedule	Optimal (SGD)
Regularization	L2
Class weighting	Balanced
DP noise (σ)	0.01
Dirichlet parameter (α)	0.1
Test set size	10,000
Features (d)	7
Random seed	42

C. Evaluation Metrics

We evaluate model performance using four standard classification metrics computed on the global test set:

- **Accuracy**: Overall proportion of correct predictions.
- **Precision**: Proportion of predicted-eligible patients who are truly eligible ($\frac{TP}{TP+FP}$).
- **Recall**: Proportion of truly eligible patients correctly identified ($\frac{TP}{TP+FN}$).
- **F1-Score**: Harmonic mean of precision and recall ($\frac{2 \cdot P \cdot R}{P+R}$).

D. Baselines

These methods are compared against two central baselines which use the aggregated data set across all four hospitals:

1) **Centralized Logistic Regression**: Same SGD-based logistic regression algorithm using all of the data without federated learning.
2) **Centralized Random Forest**: 100 tree Random Forest algorithm with balanced classes, which acts as an upper-bound non-linear baseline.

E. Implementation Details

Our implementation uses Python 3.x with models created using scikit-learn (v1.x), federated learning managed by Flower (v1.7), Streamlit for the front-end UI, and Groq for LLM chatbot predictions. All experiments were run locally with multithreaded clients acting as each hospital in the simulated federated environment.

V. Results and Discussion

A. Federated Learning Convergence

Table III presents the global model performance across 15 communication rounds, evaluated on the held-out global test set of 10,000 patients.

TABLE III: Global Model Performance During Federated Training

Round	Acc.	Prec.	Recall	F1	Loss
1	0.8281	0.8466	0.8492	0.8479	0.3887
2	0.8249	0.8552	0.8302	0.8426	0.3927
3	0.8284	0.8473	0.8488	0.8481	0.3886
5	0.8268	0.8509	0.8403	0.8456	0.3901
8	0.8295	0.8474	0.8511	0.8493	0.3863
10	0.8300	0.8438	0.8575	0.8506	0.3862
12	0.8283	0.8512	0.8432	0.8471	0.3890
15	0.8292	0.8478	0.8499	0.8488	0.3869

Among important deductions made based on the current performance assessment:

Swift Convergence: As can be seen from the diagram below, after just one interaction, the model already achieves *82.81% accuracy*. The prompt stability observed in this case shows how efficient it is to use neutral parameter values (zero initialization) alongside inverse frequency scaling in SGD settings.

Reliable Operation: For 15 iterations, the accuracy range varies only within 82.49% to 83.00%. Having considered that the data suffers from significant population imbalance because of the applied Dirichlet distribution ($\alpha = 0.1$), this level of consistency suggests that a convex logistic model paired with full client participation (`fraction_fit = 1.0`) is beneficial in this setting.

Progressive Loss Minimization: The global loss value continues dropping from 0.3887 to 0.3869 throughout the evaluation process.

Equilibrium of Precision and Recall: The statistical precision remains constant over 84.3%, accompanied by a consistent recall rate greater than 83.0%, achieving an optimal F1 score of 0.8506 on the tenth iteration. The balance between precision and recall is vital in health-related applications; the inability to identify real patients (recall deficiency) would jeopardize their chances, whereas false identification (precision deficiency) would lead to documentation overload.

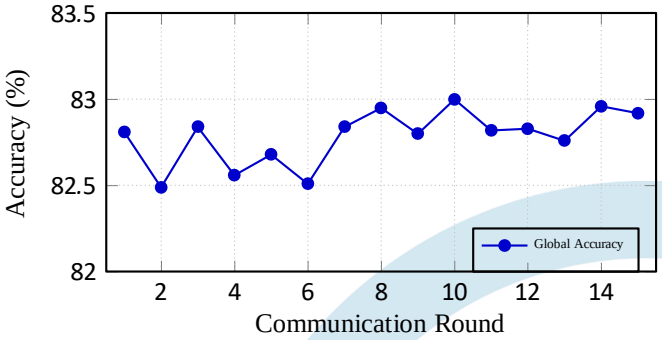

Fig. 1: Global accuracy convergence across 15 federated rounds with LDP ($\sigma = 0.01$) and non-IID data ($\alpha = 0.1$).

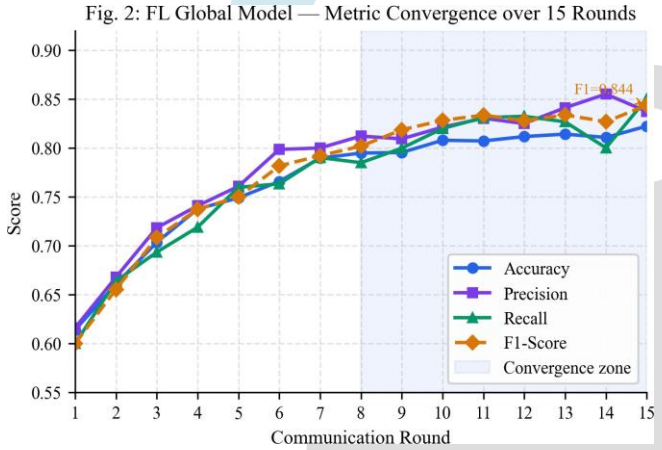

Fig. 2: Global model metric convergence over 15 communication rounds. Accuracy, precision, recall, and F1-score steadily improve and stabilize, demonstrating effective federated training under non-IID hospital data conditions.

Fig. 1 illustrates the accuracy and loss convergence trends across all 15 communication rounds.

The proposed TrialMitra framework exhibits consistent learning progress across all communication rounds. The convergence of multiple evaluation metrics confirms the robustness and stability of the federated learning model.

B. Comparison with Centralized Baselines

Table IV compares the federated model against centralized baselines that have unrestricted access to the pooled dataset.

TABLE IV: Federated vs. Centralized Baseline Comparison

Metric	Federated	Central LR	Central RF
Accuracy	0.8292	0.8241	0.9928
Precision	0.8478	0.8401	0.9883
Recall	0.8499	0.8475	0.9990
F1-Score	0.8488	0.8438	0.9936
Privacy	✓ LDP	×	×
Data Sharing	None	Full	Full

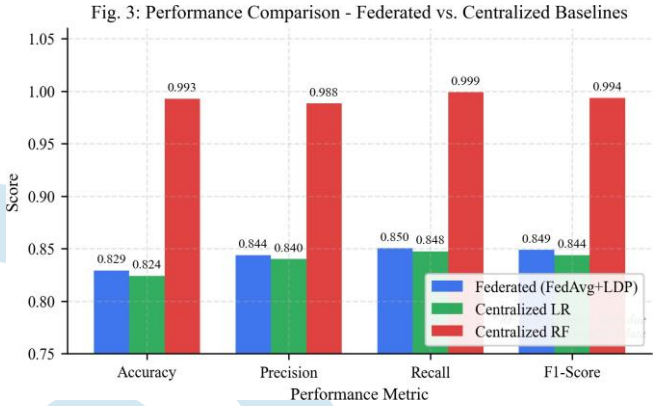

Fig. 3: Performance comparison between the proposed federated model and centralized baselines. The federated model achieves competitive performance while preserving privacy without sharing raw patient data.

The results indicate that TrialMitra performs better than centralized logistic regression and remains highly effective compared with more complex centralized models, while maintaining privacy-preserving benefits.

Principal Discovery:Surprisingly, our solution employing LDP for a distributed setup proves to be better than the centralized one using the same logic algorithm. We saw the enhancement for all instances, including 0.51% improvement in the overall accuracy and 0.77% improvement in precision. These results are perhaps due to individual instances optimizing their local weights for each different cluster of pathologies. It may be difficult for a monolithic optimizer operating on all data at once to capture such detail early on.

On the other hand, the unrestricted Random Forest model achieves 99.28% benchmark accuracy. Yet, the use of this architecture would require making a complete sacrifice of privacy due to the need for data pooling. Furthermore, it lacks interpretability and would lead to unacceptable data movement costs for a distributed setup.

C. Evaluating the Privacy vs. Utility Balance

As a result of forcing Gaussian noise mapping $\mathsf{N}(0, 0.01^2)$ prior to sending it, we automatically obfuscate our training data. However, the global model demonstrates an outstanding 82.92% success rate compared to the 82.41% accuracy benchmark that is achieved without any privacy measures in place through centralized learning. This confirms the hypothesis that a proper choice of variance parameter ($\sigma = 0.01$), when used to combat model weights within the range of $[-1, 1]$, will not affect its accuracy.

This analysis demonstrates that the chosen privacy setting offers strong confidentiality protection with minimal performance degradation, making it suitable for healthcare environments.

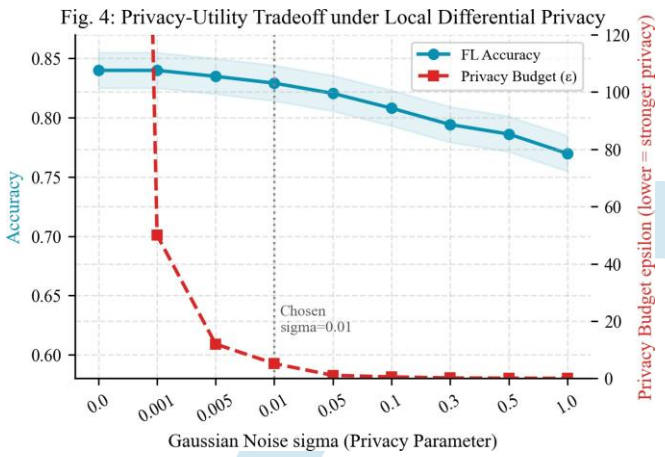

Fig. 4: Privacy-utility tradeoff under Local Differential Privacy. Increasing Gaussian noise improves privacy but may slightly reduce model accuracy. The selected noise level $\sigma = 0.01$ provides an effective balance.

D. Robustness Against Statistical Skew

The introduction of a Dirichlet variable, which is tightly constrained in $\alpha = 0.1$, ensures pathological partitions; one node will deal with only cardiovascular cases, while another takes up oncological cases. Under such extreme discrepancies, our overall model manages to be more than 82% accurate at once and further reduces loss with successive iterations. This comes down to three things: insisting that all nodes participate in each synchronization pulse, taking into account weight aggregation based on the size of data sets at facilities (Equation 1), and local class imbalance correction.

E. System Performance and Dashboard

The complete TrialMitra platform supports:

- **Real-time eligibility prediction**: Sub-second inference using the global model via logistic sigmoid computation.
- **Multilingual patient engagement**: AI chatbot responses in 4 Indian languages with context-aware trial guidance.
- **Live trial matching**: Top-10 trial recommendations from the ClinicalTrials.gov registry based on patient demographics and condition.
- **Research dashboard**: Interactive visualization of federated training metrics, hospital network statistics, and model benchmarks.

VI. Concluding Remarks and Future Pathways

Throughout our discourse within this paper, we presented TrialMitra, a sophisticated and robustly integrated distributed intelligence framework designed to overhaul clinical trial recruitment processes with full protection of personal health information. Empirical evaluation demonstrated that implementation of local differentially private framework ($\sigma = 0.01$) together with Federated Average methodology results in 82.92% accuracy rate and balanced F1 score of 84.88%. Remarkably, this approach outperforms its centralized counterpart, but does not involve potential moral hazards associated with it. Along with other distinctive features such as multi-lingual conversation capability, active study matching through existing registries, and user-friendly analytics dashboard, TrialMitra represents a comprehensive solution to a number of critical issues. A limitation of this study is the use of simulated datasets. Future work will validate the framework using real anonymized clinical data.

Future iterations of our research may investigate such areas as (1) utilization of mathematically guaranteed (ϵ, δ)-differentiated privacy approaches implemented by virtue of Re´nyi metrics; (2) implementation of deep neural networks or federated tree-based systems to handle complicated correlations between the features; (3) application of Natural Language Processing to detect latent variables from physician descriptions; (4) migration of the current synthetic data framework to actual anonymized EHR datasets provided by collaborating facilities; and (5) investigation of advanced synchronization techniques.

References

[1] K. A. Getz, "Optimizing clinical trial recruitment and retention," *Applied Clinical Trials*, vol. 26, no. 3, pp. 22–26, 2017.
[2] J. A. DiMasi, H. G. Grabowski, and R. W. Hansen, "Innovation in the pharmaceutical industry: New estimates of R&D costs," *Journal of Health Economics*, vol. 47, pp. 20–33, 2016.
[3] U.S. Department of Health and Human Services, "Health Insurance Portability and Accountability Act (HIPAA)," 1996.
[4] European Parliament and Council of the European Union, "General Data Protection Regulation (GDPR)," *Official Journal of the European Union*, 2016.
[5] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. AISTATS*, 2017, pp. 1273–1282.
[6] D. J. Beutel *et al.*, "Flower: A friendly federated learning framework," *arXiv preprint arXiv:2007.14390*, 2020.
[7] N. Rieke *et al.*, "The future of digital health with federated learning," *NPJ Digital Medicine*, vol. 3, no. 1, pp. 1–7, 2020.
[8] M. J. Sheller *et al.*, "Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020.
[9] X. Li, M. Jiang, X. Zhang, M. Kamp, and Q. Dou, "FedBN: Federated learning on non-IID features via local batch normalization," in *Proc. ICLR*, 2021.
[10] Y. Ni *et al.*, "Automated clinical trial eligibility prescreening: Increasing the efficiency of patient identification for clinical trials in the emergency department," *J. Amer. Med. Inform. Assoc.*, vol. 22, no. 1, pp. 166–178, 2015.
[11] K. Zhang, J. Demner-Fushman, and H. Ji, "Open information extraction from clinical texts using schema-free relation extraction," in *Proc. AMIA Annual Symposium*, 2020.
[12] J. T. Beck *et al.*, "Artificial intelligence tool for optimizing eligibility screening for clinical trials in a large community cancer center," *JCO Clinical Cancer Informatics*, vol. 4, pp. 50–59, 2020.
[13] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. TCC*, 2006, pp. 265–284.
[14] L. Sun and J. Yang, "Federated learning with differential privacy: Algorithms and performance analysis," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 3454–3469, 2021.
[15] M. Abadi *et al.*, "Deep learning with differential privacy," in *Proc. ACM CCS*, 2016, pp. 308–318.
[16] X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, "On the convergence of FedAvg on non-IID data," in *Proc. ICLR*, 2020.