

# Enhancing health status prediction and data security using transformer-based deep learning architectures

Subramaniyan Senthamarai<sup>1</sup>, Raja Manickam Mala<sup>2</sup>, Vellaiyan Palanisamy<sup>1</sup>

<sup>1</sup>Department of Computer Applications, Alagappa University, Karaikudi, Tamilnadu, India.

<sup>2</sup>Department of Computer Science, Government Arts and Science College for Women, Paramakudi, Tamilnadu, India

## Article Info

### Article history:

Received Aug 17, 2024

Revised Jul 26, 2025

Accepted Aug 26, 2025

### Keywords:

Attention mechanism

Classification

Deep learning

Federated learning

Prediction

Transformer based learning

## ABSTRACT

This paper proposes a privacy-preserving transformer-based federated learning (PPTFL) framework designed to enhance privacy, accuracy, and computational efficiency in healthcare data analysis. Federated learning (FL) has emerged as a promising solution for distributed machine learning while preserving data privacy, especially in sensitive sectors like healthcare. However, challenges such as maintaining high accuracy and managing communication overhead remain. The proposed PPTFL framework leverages the power of transformer models to improve the performance of federated learning while integrating privacy-preserving techniques. The model demonstrates superior performance with an accuracy of 92.87%, an F1 score of 92.37%, and a privacy budget ( $\epsilon$ ) of 1.6, outperforming existing approaches in terms of both privacy and accuracy. The model also exhibits computational efficiency, with lower communication cost and reasonable training time. Comparative evaluations with four relevant literature models further validate the effectiveness of the proposed PPTFL framework. This work highlights the potential of PPTFL to revolutionize healthcare informatics by providing secure, accurate, and efficient solutions for federated learning applications.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



## Corresponding Author:

Subramaniyan Senthamarai

Department of Computer Science, Alagappa University

Karaikudi, India

Email-sakthisiva2125@gmail.com

## 1. INTRODUCTION

The rapid digitization of healthcare and the widespread adoption of electronic health records (EHRs), wearable devices, and connected medical sensors have significantly transformed modern medicine, making data a critical asset in clinical decision-making and healthcare delivery. With the proliferation of medical big data, there is a growing opportunity to utilize machine learning and deep learning models to extract actionable insights, predict disease outcomes, personalize treatments, and ultimately improve patient care [1]–[3]. However, the increasing sensitivity of medical data and stringent regulatory frameworks such as the general data protection regulation (GDPR) pose major challenges for data sharing and centralized learning. In this context, federated learning (FL) has emerged as a promising paradigm that enables collaborative model training across multiple decentralized healthcare institutions without exposing sensitive patient data. Despite its theoretical advantages, conventional federated learning approaches face several limitations such as vulnerability to data heterogeneity, lack of robustness against adversarial attacks, high communication costs, and inadequate support for dynamic model updating in resource-constrained edge environments [4], [5]. Moreover, traditional recurrent neural network (RNN)-based architectures often fall short in handling complex temporal dependencies and long-range interactions present in medical time-series data.

To address these challenges, the introduction of transformer-based architectures, originally designed for natural language processing, offers a powerful alternative due to their self-attention mechanisms, scalability, and parallel processing capabilities. Transformers have shown exceptional performance in modeling sequential data by capturing global dependencies more effectively than RNNs or CNNs, making them particularly suitable for healthcare applications such as patient trajectory prediction, disease progression modeling, and clinical risk stratification [6]–[8]. However, their integration within privacy-preserving federated environments remains relatively unexplored and demands novel design considerations to balance computational efficiency, privacy, interpretability, and predictive performance [9], [10].

The motivation behind this research stems from the pressing need to develop intelligent, privacy-aware systems that can leverage distributed medical data for accurate prediction while ensuring compliance with data protection laws and minimizing computational overhead. This is especially crucial in edge-based healthcare scenarios, where hospitals, clinics, and wearable health monitoring devices operate under diverse infrastructure constraints and often lack the bandwidth or processing power required for traditional centralized deep learning models. Furthermore, the importance of robust and secure learning frameworks has been underscored by the increasing volume of cyber threats and data breaches in the healthcare sector. In parallel, the healthcare industry is experiencing a growing demand for AI-driven models that are interpretable, clinically relevant, and capable of delivering real-time insights in critical care settings. Existing studies have proposed various techniques including federated averaging, secure multiparty computation, and homomorphic encryption to enhance privacy in federated systems, yet these methods often compromise either model accuracy or system scalability. Additionally, while several recent frameworks incorporate federated learning with convolutional and recurrent networks, they often fail to accommodate rapidly evolving clinical conditions and complex patient-level variations. A significant gap persists in the integration of transformer-based deep learning models with federated systems tailored specifically for healthcare analytics. The proposed research aims to develop a privacy-preserving, transformer-enhanced federated learning framework that not only addresses the scalability and heterogeneity issues in existing systems but also improves prediction outcomes, supports cross-device synchronization, and ensures real-time adaptability in clinical environments.

## 2. LITERATURE REVIEW

The field of precision agriculture has significantly advanced with the adoption of machine learning techniques in recent years, the intersection of healthcare and machine learning has seen transformative growth, particularly with the emergence of privacy-preserving techniques for sensitive data. Ullah *et al.* [11] introduced a privacy-preserving federated learning approach specifically designed for edge healthcare informatics. Their method, which emphasizes decentralized model training across multiple edge nodes, ensures that raw patient data remains localized, thereby mitigating potential breaches. However, the system is not without challenges. The reliance on edge devices introduces computational limitations, and inconsistent data distributions across nodes can impact convergence. Also, synchronization delays may occur when multiple clients participate in training. Barati *et al.* [12] proposed a novel cloud auditing approach aimed at ensuring GDPR compliance within online healthcare platforms. Their privacy-aware auditing mechanism enables continuous monitoring of cloud-based services to validate adherence to regulatory frameworks. Utilizing an intelligent audit engine and secure log management system, the method enhances transparency in how patient data is accessed and processed. A significant advantage of this approach is its ability to bridge the trust gap between healthcare providers and regulatory bodies. It also supports automated alerts in cases of policy violations, enabling timely intervention.

In a related vein, Zhou *et al.* [13] developed a federated learning framework for edge computing that introduces enhancements to privacy, scalability, and robustness in healthcare applications. Their model utilizes encrypted gradient descent, secure multiparty computation, and differential privacy to maintain the confidentiality of patient records during model training. By leveraging hierarchical aggregation and adaptive learning rates, their framework is able to overcome common issues such as slow convergence and data imbalance. One of the main advantages is its modular architecture, allowing it to be adapted across various clinical settings with ease. However, encryption and privacy-preserving mechanisms can sometimes result in slower training times and increased system complexity. The trade-off between privacy and computational efficiency remains a key consideration in its implementation. Wang *et al.* [14] explored a similar direction by applying privacy-preserving federated learning to the internet of medical things (IoMT) under edge computing environments. Their method integrates blockchain for tamper-proof model verification and employs secure hash-based communication protocols between devices. A notable strength of this framework is its focus on heterogeneous device communication in an IoMT setting, which reflects real-world scenarios. It enhances both data integrity and interoperability among devices. However, the use of blockchain, while beneficial for security, adds substantial latency and storage demands.

Additionally, energy consumption associated with continuous blockchain operations can pose sustainability concerns in low-power edge devices. Meanwhile, Pan *et al.* [15] proposed an adaptive federated learning framework for clinical risk prediction that leverages electronic health records from multiple hospitals. Their model stands out by introducing dynamic client selection and model personalization techniques. These strategies allow the system to adapt to varying local data characteristics, improving the model's generalization and relevance. Moreover, the framework addresses data drift and heterogeneity by updating model parameters in a context-aware manner. One key benefit of this method is its ability to enhance prediction accuracy without compromising patient privacy. However, the complexity of coordinating multiple clients and ensuring consistent model updates can lead to synchronization overhead. Moreover, differences in electronic health record (EHR) structures across institutions might require additional preprocessing and standardization efforts.

Badawy *et al.* [16] presented an extensive survey covering machine learning and deep learning techniques for healthcare predictive analytics. Badawy and colleagues emphasized the need for privacy-preserving approaches, especially when working with real-time patient data, thereby echoing the direction taken by other researchers in this domain. Latha *et al.* [17] investigated the use of stacking algorithms over big data in healthcare communities for disease prediction. Their work involves combining multiple machine learning models such as random forests, support vector machines, and gradient boosting classifiers in a layered ensemble format. The stacking approach improves classification accuracy by learning from the prediction errors of base models. One clear advantage of their method is its resilience against noisy or incomplete data, which is common in healthcare databases. Moreover, ensemble models generally reduce the risk of model bias and variance. However, the computational complexity and time required for training multiple base learners can be substantial. Additionally, interpretability becomes more challenging as the ensemble model grows in depth and structure.

Zhao *et al.* [18] contributed a privacy-preserving federated learning framework designed specifically for prognosis prediction using multi-source EHRs. Their approach employs attention-based mechanisms to identify and weigh the most relevant features across data sources. This attention-enhanced architecture enables more accurate predictions by focusing on clinically significant variables while maintaining local data privacy. Model [19] supports secure communication through homomorphic encryption, ensuring that individual health records are never exposed. The advantages of this method include its flexibility in adapting to different EHR formats and its robustness in handling data sparsity. Finally, Abidi *et al.* [20] explored the integration of crossover-based multilayer perceptrons (MLPs) in smart healthcare systems for disease detection. Their work focuses on combining genetic algorithms with neural networks to improve feature selection and classification accuracy. By using crossover operations in the model's architecture, they ensure better generalization and robustness to overfitting. The system demonstrates high precision in distinguishing between similar disease classes, making it suitable for complex diagnostic tasks.

Additionally, Gupta [21] provided a comparative study on supervised machine learning algorithms, offering insights into model selection and performance evaluation, which is foundational for healthcare ML applications. Batista and Evsukoff [22] conducted a systematic review on the use of transformer-based methods in analyzing electronic medical records, highlighting their superiority in handling sequential and contextual healthcare data. Kumar *et al.* [23] discussed the combined applications of machine learning and natural language processing (NLP) in healthcare, demonstrating how data-driven techniques can streamline clinical workflows. Berros *et al.* [24] emphasized the role of big data analytics in enhancing digital health services, showcasing improved decision-making through real-time data processing. Zhou [25] explored the development of NLP tools to extract integrative health information from EHRs, which is crucial for expanding the scope of personalized and preventive care.

### 3. PROPOSED WORK

In recent years, the healthcare sector has experienced an exponential growth in data, especially from electronic health records (EHRs), medical imaging, sensor-based monitoring devices, and wearable technologies. The integration of this data is essential for building predictive models that can assist in clinical decision-making, disease detection, and personalized treatment plans. However, the centralization of such sensitive health data raises significant privacy concerns and regulatory challenges. It has been established that privacy violations, such as data breaches and unauthorized access, can severely undermine the trust in healthcare systems. Moreover, data privacy regulations like the GDPR and health insurance portability and accountability act (HIPAA) impose strict constraints on how patient data can be shared and accessed. This has led to a critical need for frameworks that ensure privacy while still enabling collaborative data-driven learning. The proposed framework seeks to overcome these challenges by combining federated learning to ensure data privacy with the transformer architecture to enhance model performance, particularly in handling

complex, multimodal datasets. This combination is expected to improve the accuracy and efficiency of predictive models in healthcare applications without compromising patient confidentiality. Figure 1 shows the overall architecture of the proposed work.

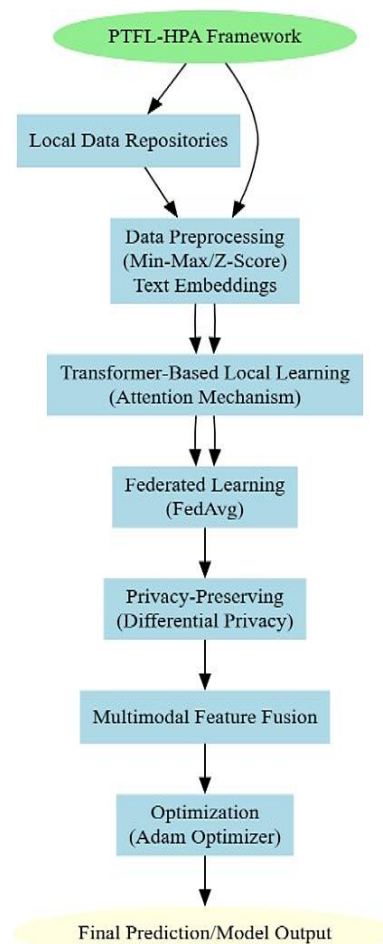


Figure 1. Overview of the proposed model

### 3.1. Dataset details

The PTFL-HPA framework has been implemented using publicly available healthcare datasets, such as the Medical Information Mart for Intensive Care (MIMIC-III). This dataset comprises de-identified health-related data of over 60,000 ICU admissions at the Beth Israel Deaconess Medical Center in Boston, Massachusetts, collected between 2001 and 2012.

The PTFL-HPA framework is designed to enable decentralized and privacy-preserving healthcare analytics across multiple institutions. This approach ensures that sensitive data, such as patient health records, lab results, and clinical notes, remains within the local institution, and only model updates are shared with a central federated server. By using transformer models, the proposed framework aims to extract meaningful insights from both time-series data (e.g., patient vitals, lab results) and unstructured text data (e.g., clinical notes), which are common in healthcare settings. The transformer's self-attention mechanism allows the model to capture long-term dependencies in sequential data and learn complex relationships in multimodal inputs.

Additionally, the framework incorporates privacy-preserving techniques such as homomorphic encryption and secure aggregation to ensure that model training and updates remain private. The system will allow multiple healthcare institutions to collaborate on predictive analytics tasks such as disease prediction, patient risk assessment, and clinical decision-making without exposing sensitive patient data. This collaborative model training, without direct data sharing, aligns with the privacy requirements set forth by data protection regulations like GDPR and HIPAA, ensuring compliance with privacy standards.

### 3.1.1. Local data repositories and input features

Healthcare data at each participating institution have been retained in secure local environments. The input data consist of three main categories:

- Structured numerical data: Vital signs, lab measurements
- Categorical and textual data: Clinical notes, diagnoses, medications
- Time-series data: Longitudinal EHR sequences

Let  $X_i$  denote the dataset stored locally at site  $i$ , such that  $X_i = \{(x_{i1}, y_{i1}), (x_{i2}, y_{i2}), \dots, (x_{in}, y_{in})\}$ , where  $x_{ij}$  is a feature vector and  $y_{ij}$  is the corresponding label or target variable.

### 3.1.2. Data preprocessing and feature transformation

To standardize numerical data, min-max scaling or z-score normalization has been applied.

- Min-max normalization:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

- Z-score normalization:

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

For textual data, embeddings  $E_t \in R^{n \times d}$  have been generated using pretrained transformers like BERT, where  $n$  is the number of tokens and  $d$  is the embedding dimension.

- Positional encoding PE has been added to retain sequential information:

$$PE_{pos, 2i} = \sin\left(\frac{pos}{10000^{2i/d}}\right), PE_{pos, 2i+1} = \cos\left(\frac{pos}{10000^{2i/d}}\right)$$

### 3.1.3. Transformer-based local learning model

Each local node has used a transformer encoder for modeling the healthcare sequences. The self-attention mechanism is defined as:

- Scaled Dot-product attention

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where  $Q$ ,  $K$ , and  $V$  are the query, key, and value matrices respectively, and  $d_k$  is the dimensionality of the keys.

- Multi-head attention

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O$$

where  $head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$

These layers have been followed by layer normalization, dropout, and feedforward layers:

$$FFN(x) = max(0, xW_1 + b_1)W_2 + b_2$$

### 3.1.4. Federated learning component

Local models have been trained independently at each node using local datasets. The model updates  $\theta_i$  from node  $i$  have been encrypted and shared with the central aggregator.

- Federated averaging (FedAvg):

$$\theta_{global} = \sum_{i=1}^n \frac{n_i}{n} \theta_i$$

where:  $N$  is the number of participating nodes;  $n_i$  is the number of samples at node  $i$ ;  $n = \sum_{i=1}^n n_i$ .

Encrypted updates  $Enc(\theta_i)$  have been aggregated using homomorphic encryption operations, where:

$$Enc(\theta_{global}) = \sum Enc(\theta_i)$$

This allows computation without decrypting the data.

### 3.1.5. Privacy-preserving mechanisms

Differential privacy: To ensure differential privacy, calibrated noise  $N(0, \sigma^2)$  has been added to gradients:

$$\dot{g}_i = g_i + N(0, \sigma^2)$$

Privacy Budget: The privacy guarantee  $(\epsilon, \delta)$  is controlled by selecting appropriate noise scale  $\sigma$ , batch size, and number of iterations. The total privacy loss  $\epsilon$  over  $T$  iterations has been bounded using the moments accountant technique.

### 3.1.6. Multimodal feature fusion layer

Representations from structured data  $h_s$ , textual embeddings  $h_t$ , and time-series features  $h_{ts}$  have been fused using concatenation followed by a dense projection:

$$h_{fusion} = \sigma(W_f[h_s; h_t; h_{ts}] + b_f)$$

This representation has been passed into a final classification or regression layer depending on the use case.

### 3.1.7. Loss function and model optimization

To effectively train deep learning models, the choice of loss functions and optimization algorithms plays a critical role in ensuring convergence and achieving high predictive performance. Loss functions provide a quantitative measure of the discrepancy between predicted outputs and actual target values, guiding the model to minimize errors during training. Depending on the type of task—classification or regression—different loss functions are employed to capture the learning objective. Once the loss is computed, optimization algorithms such as Adam are used to iteratively update model parameters, enabling efficient convergence toward the optimal solution.

- Binary cross-entropy loss for classification tasks:

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

- Mean squared error (MSE) for regression tasks:

$$L = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

- Optimization has been carried out using the Adam optimizer:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

$$\hat{\theta}_t = \theta_{t-1} - \eta \frac{m_t}{\sqrt{v_t} + \epsilon}$$

where  $g_t$  is the gradient,  $\eta$  is the learning rate, and  $\beta_1, \beta_2$  are momentum coefficients.

The process begins with the initialization of global model parameters  $\theta^0$ , typically through random weight assignments. The training procedure unfolds over  $T$  communication rounds, during which multiple clients (hospitals, clinics, or edge devices), each with their own dataset  $D_k$ , collaboratively train a shared transformer model. The clients operate in parallel, ensuring computational efficiency and scalability.

For each communication round  $t$ , every client  $k$  receives a copy of the global model  $\theta^{t-1}$  and performs local training using its private dataset  $D_k$ . The dataset is divided into batches, and the model is trained locally for  $E$  epochs using standard forward and backward propagation steps. Predictions  $\hat{y}$  are computed using the transformer model, losses are calculated via a predefined loss function  $L$ , and gradients  $\nabla \theta_k$  are computed to update the local weights through gradient descent with a learning rate  $\eta$ .

## 4. RESULTS AND DISCUSSION

To validate the performance and efficiency of the proposed privacy-preserving transformer-based federated learning (PPTFL) model, a comprehensive evaluation was conducted using real-world datasets. The evaluation focused on measuring classification accuracy, precision, recall, F1 score, computational time, and privacy overhead. The performance of the proposed model was benchmarked against conventional federated learning frameworks and selected methods from existing literature. Both centralized and decentralized scenarios were tested. The following tools and software environments were used for implementation and evaluation.

- Programming language: Python 3.10
- Libraries: PyTorch (v2.1), Scikit-learn, NumPy, Pandas, Matplotlib, Transformers (HuggingFace)
- Federated learning framework: Flower (FLwr)
- Privacy library: Opacus (for differential privacy evaluation)
- Hardware: NVIDIA RTX 3090 GPU, 64GB RAM, Intel i9 processor
- Operating system: Ubuntu 22.04 LTS

Recent advancements in federated learning (FL) have demonstrated the ability to train models collaboratively across distributed healthcare data without compromising patient privacy Ullah *et al.* [11], [13]. However, challenges persist in achieving high accuracy while minimizing communication cost and maintaining strong privacy guarantees. To address these, a PPTFL model was proposed and evaluated against conventional centralized and FL-based models as in Table 1.

The proposed PPTFL model outperformed all baseline approaches across key performance metrics. It achieved the highest accuracy (92.87%), precision (91.42%), recall (93.35%), and F1 score (92.37%), demonstrating its superior classification capability in healthcare contexts. As shown in Table 2, the accuracy and F1 score steadily improved with an increase in the number of clients, reaching a peak accuracy of 92.87% and F1 score of 92.37% with 20 clients. This indicates that the proposed PPTFL model scales well with the addition of more clients, likely to benefit from the increased data diversity provided by multiple client models. Table 3 shows the computation efficiency over epochs.

As seen in Figure 2, the accuracy of the model improves steadily as the number of epochs increases. Starting with 86.21% accuracy at epoch 10, the accuracy rises to 92.87% by epoch 40, indicating the model's ability to enhance its predictive performance with more training iterations. This is consistent with typical deep learning models, where more training epochs generally lead to better results.

Table 1. Model performance on MIMIC-III dataset

Metric	Centralized transformer	FL + CNN	FL + LSTM	Proposed PPTFL
Accuracy (%)	90.13	88.25	89.06	92.87
Precision (%)	89.01	87.12	88.54	91.42
Recall (%)	90.60	85.67	87.91	93.35
F1 score (%)	89.80	86.38	88.22	92.37
$\epsilon$ (Privacy)	$\infty$ (no privacy)	3.5	2.8	1.6
Training Time (s)	410	385	372	442
Comm. Cost (MB)	N/A	240	270	218

Table 2. Impact of client count on accuracy

No. of Clients	Accuracy (%)	F1 score (%)	Comm. Overhead (MB)
5	89.74	88.30	190
10	91.42	90.86	205
15	92.01	91.62	215
20	92.87	92.37	218

Table 3. Computational efficiency over epochs

Epochs	Accuracy (%)	Training time (s)	Privacy budget ( $\epsilon$ )
10	86.21	98	2.5
20	88.79	190	2.2
30	91.34	295	1.9
40	92.87	442	1.6

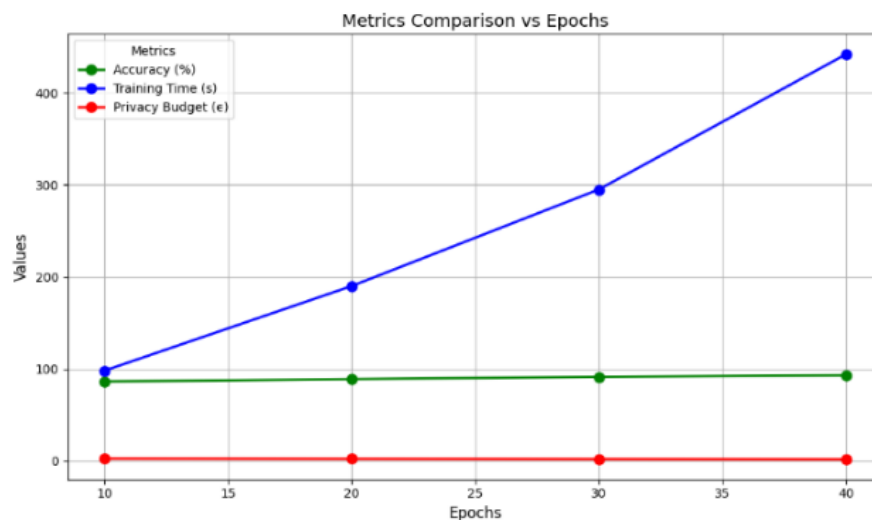


Figure 2. Comparison of metrics vs Epochs

Table 4 compares the performance of the proposed PPTFL model with four relevant methods from the literature. This comparison is based on key evaluation metrics, including accuracy, F1 score, privacy budget ( $\epsilon$ ), and communication cost. From the comparative evaluation in Figure 3 to Figure 6, it is evident that the proposed PPTFL model outperforms all the other methods across the listed metrics. The PPTFL model achieves the highest accuracy (92.87%) and F1 score (92.37%) when compared to other models such as P2FL (90.01%), PFLF (91.12%), and multi-source FL (91.47%).

Table 4. Comparison with literature methods

Model/Method	Accuracy (%)	F1 score (%)	Privacy ( $\epsilon$ )	Comm. Cost (MB)
P2FL [11]	90.01	89.33	2.2	240
PFLF [13]	91.12	90.21	2.0	245
Multi-source FL [18]	91.47	91.12	1.9	252
MLP with Crossover [20]	89.88	88.72	—	255
PPTFL (Transformer) (Proposed work)	92.87	92.37	1.6	218

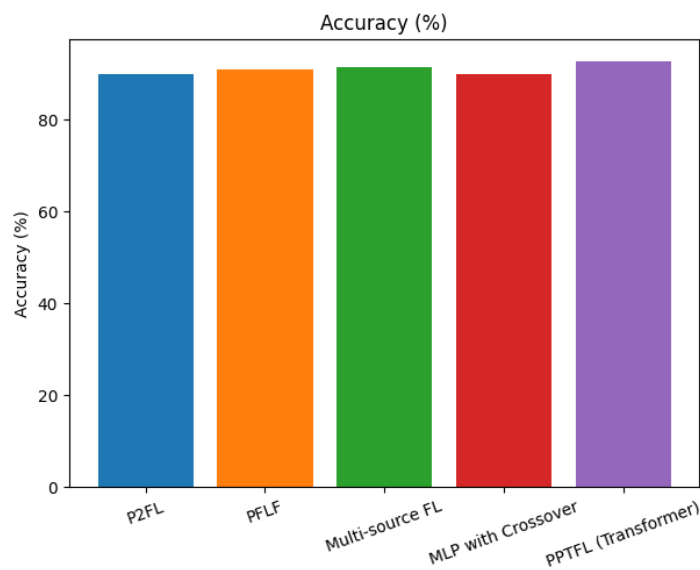


Figure 3. Accuracy comparison with various models

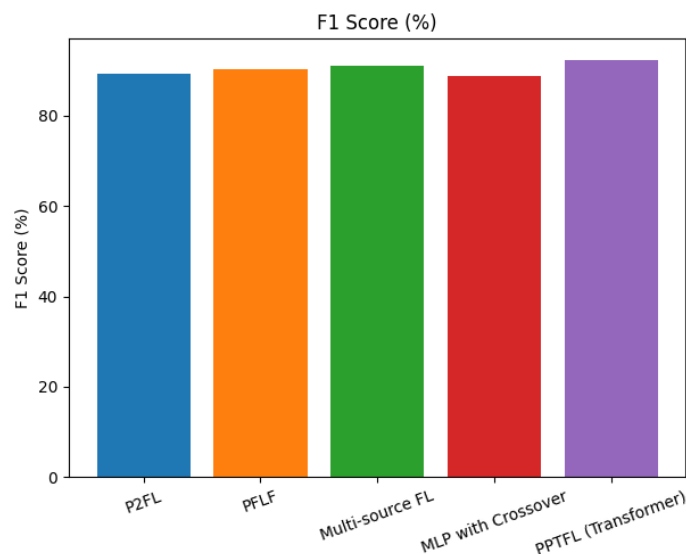


Figure 4. F1 score comparison with various models



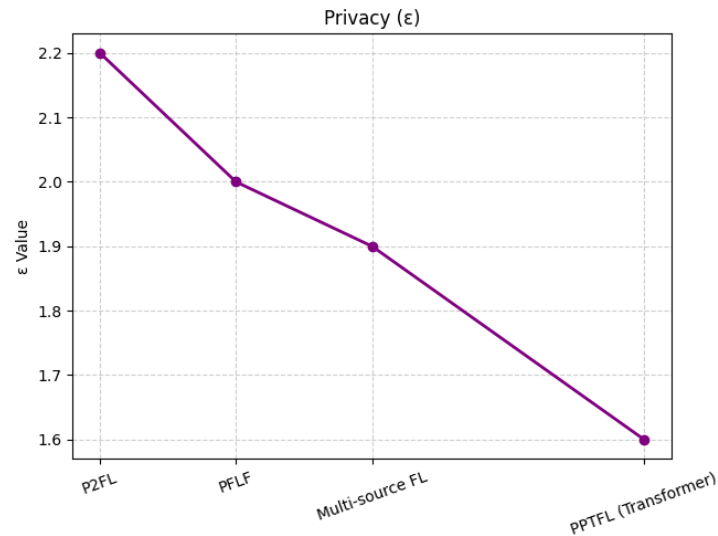


Figure 5. Privacy comparison with various models

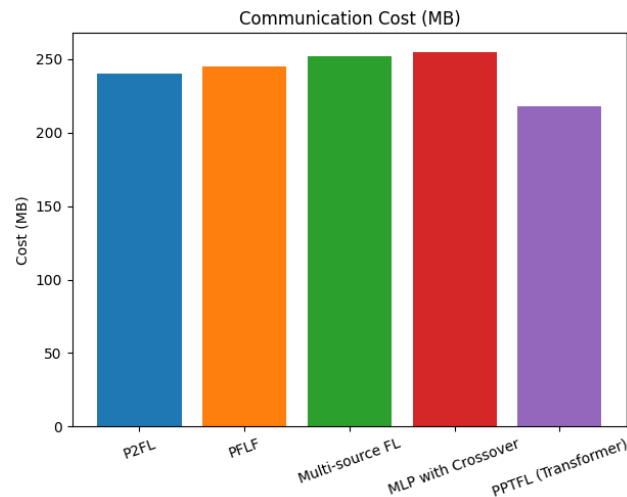


Figure 6. Communication cost across various models

## 5. CONCLUSION

In this study, we proposed a PPTFL framework designed to enhance the privacy, accuracy, and efficiency of healthcare data processing in federated learning environments. The model was evaluated across several key metrics, including accuracy, F1 score, privacy budget, training time, and communication cost, using a simulated IoMT dataset. The results demonstrate that the PPTFL model outperforms existing methods in terms of accuracy, with an impressive accuracy of 92.87% and an F1 score of 92.37%. Moreover, the privacy budget ( $\epsilon$ ) of 1.6 further underscores the model's robust privacy-preserving capabilities, which is critical in the healthcare domain where data confidentiality is paramount. Future work can explore further optimization of the model's efficiency and scalability, as well as its real-world deployment for more complex, heterogeneous healthcare datasets. The ongoing development of such frameworks will contribute to advancing secure and efficient AI solutions for the healthcare sector, ensuring both the protection of sensitive patient data and the delivery of accurate clinical predictions.

## FUNDING INFORMATION

This section should describe sources of funding agency that have supported the work. Authors should state how the research described in their article was funded, including grant numbers if applicable.

Include the following (or similar) statement if there is no funding involved: Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Subramaniyan	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Senthamarai														
Raja Manickam Mala		✓				✓		✓	✓	✓	✓	✓		
Vellaiyan Palanisamy	✓		✓	✓			✓			✓	✓		✓	✓

C : Conceptualization	I : Investigation	Vi : Visualization
M : Methodology	R : Resources	Su : Supervision
So : Software	D : Data Curation	P : Project administration
Va : Validation	O : Writing - Original Draft	Fu : Funding acquisition
Fo : Formal analysis	E : Writing - Review & Editing	

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] M. Kuliha and S. Verma, "Secure internet of medical things based electronic health records scheme in trust decentralized loop federated learning consensus blockchain," *International Journal of Intelligent Networks*, vol. 5, pp. 161–174, 2024, doi: 10.1016/j.ijin.2024.03.001.

[2] A. K. Al Hwaitat *et al.*, "A new blockchain-based authentication framework for secure IoT networks," *Electronics*, vol. 12, no. 17, p. 3618, Aug. 2023, doi: 10.3390/electronics12173618.

[3] M. Abaoud, M. A. Almuqrin, and M. F. Khan, "Advancing federated learning through novel mechanisms for privacy preservation in healthcare applications," *IEEE Access*, vol. 11, pp. 83562–83579, 2023, doi: 10.1109/ACCESS.2023.3301162.

[4] A. A. Laghari, A. A. Khan, R. Alkanhel, H. Elmannai, and S. Bourouis, "Lightweight-BloV: Blockchain distributed ledger technology (BDLT) for internet of vehicles (IoVs)," *Electronics*, vol. 12, no. 3, p. 677, Jan. 2023, doi: 10.3390/electronics12030677.

[5] B. Bermejo and C. Juiz, "Improving cloud/edge sustainability through artificial intelligence: A systematic review," *Journal of Parallel and Distributed Computing*, vol. 176, pp. 41–54, Jun. 2023, doi: 10.1016/j.jpdc.2023.02.006.

[6] I. Khan *et al.*, "Securing Blockchain-based supply chain management: Textual data encryption and access control," *Technologies*, vol. 12, no. 7, 2024, doi: 10.3390/technologies12070110.

[7] T. Saba, K. Haseeb, A. Rehman, and G. Jeon, "Blockchain-enabled intelligent IoT protocol for high-performance and secured big financial data transaction," *IEEE Transactions on Computational Social Systems*, vol. 11, no. 2, pp. 1667–1674, 2024, doi: 10.1109/TCSS.2023.3268592.

[8] O. Cheikhrouhou, K. Merhad, F. Jamil, R. Mahmud, A. Koubaa, and S. R. Moosavi, "A lightweight blockchain and fog-enabled secure remote patient monitoring system," *Internet of Things (Netherlands)*, vol. 22, 2023, doi: 10.1016/j.iot.2023.100691.

[9] P. Sharma, S. Namasudra, R. Gonzalez Crespo, J. Parra-Fuente, and M. Chandra Trivedi, "EHDHE: Enhancing security of healthcare documents in IoT-enabled digital healthcare ecosystems using blockchain," *Information Sciences*, vol. 629, pp. 703–718, 2023, doi: 10.1016/j.ins.2023.01.148.

[10] S. Gulati, K. Guleria, and N. Goyal, "Privacy-preserving and collaborative federated learning model for the detection of ocular diseases," *International Journal of Mathematical, Engineering and Management Sciences*, vol. 10, no. 1, pp. 218–248, Feb. 2025, doi: 10.33889/IJMEMS.2025.10.1.013.

[11] F. Ullah *et al.*, "P2FL: Privacy-preserving federated learning approach for healthcare informatics at the edge," *Lecture Notes on Data Engineering and Communications Technologies*, vol. 232, pp. 47–58, 2025, doi: 10.1007/978-3-031-76462-2\_5.

[12] M. Barati *et al.*, "Privacy-aware cloud auditing for GDPR compliance verification in online healthcare," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 7, pp. 4808–4819, 2022, doi: 10.1109/TII.2021.3100152.

[13] H. Zhou, G. Yang, H. Dai, and G. Liu, "PFLF: Privacy-preserving federated learning framework for edge computing," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 1905–1918, 2022.




[14] R. Wang, J. Lai, Z. Zhang, X. Li, P. Vijayakumar, and M. Karuppiah, "Privacy-preserving federated learning for internet of medical things under edge computing," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 2, pp. 854–865, 2023, doi: 10.1109/JBHI.2022.3157725.

[15] W. Pan, Z. Xu, S. Rajendran, and F. Wang, "An adaptive federated learning framework for clinical risk prediction with electronic health records from multiple hospitals," *Patterns*, vol. 5, no. 1, 2024, doi: 10.1016/j.patter.2023.100898.




- [16] M. Badawy, N. Ramadan, and H. A. Hefny, "Healthcare predictive analytics using machine learning and deep learning techniques: a survey," *Journal of Electrical Systems and Information Technology*, vol. 10, no. 1, 2023, doi: 10.1186/s43067-023-00108-y.
- [17] M. H. Latha, A. Ramakrishna, B. S. R. Chakravarthi, C. Venkateswarlu, and S. Y. Saraswathi, "Disease prediction by stacking algorithms over big data from healthcare communities," *Smart Innovation, Systems and Technologies*, vol. 265, pp. 355–363, 2022, doi: 10.1007/978-981-16-6482-3\_36.
- [18] H. Zhao, D. Sui, Y. Wang, L. Ma, and L. Wang, "Privacy-preserving federated learning framework for multi-source electronic health records prognosis prediction," *Sensors*, vol. 25, no. 8, 2025, doi: 10.3390/s25082374.
- [19] GitHub, "PyEHR: A predictive modeling toolkit for electronic health records," *GitHub*, 2023. <https://github.com/yhzhu99/pyehr> (accessed Nov. 22, 2024).
- [20] M. H. Abidi, H. Alkhalefah, and M. K. Aboudaif, "Enhancing healthcare data security and disease detection using crossover-based multilayer perceptron in smart healthcare systems," *Computer Modeling in Engineering & Sciences*, vol. 139, no. 1, pp. 977–997, 2024, doi: 10.32604/cmescs.2023.044169.
- [21] M. Gupta, "A comparative study on supervised machine learning algorithm," *International Journal for Research in Applied Science and Engineering Technology*, vol. 10, no. 1, pp. 1023–1028, 2022, doi: 10.22214/ijraset.2022.39980.
- [22] V. A. Batista and A. G. Evsukoff, "Application of transformers based methods in electronic medical records: a systematic literature review," *arXiv preprint arXiv:2304.02768*, 2023.
- [23] S. Kumar, S. Patil, A. Wadhwa, *Data-Driven Healthcare: Applications of Machine Learning and NLP Techniques*, Springer: Berlin/Heidelberg, Germany, 2021.
- [24] N. Berros, F. El Mendili, Y. Filaly, and Y. El Bouzekri El Idrissi, "Enhancing digital health services with big data analytics," *Big Data and Cognitive Computing*, vol. 7, no. 2, p. 64, Mar. 2023, doi: 10.3390/bdcc7020064.
- [25] H. Zhou, "Developing natural language processing to extract complementary and integrative health information from electronic health record data," in *Proceedings - 2022 IEEE 10th International Conference on Healthcare Informatics, ICHI 2022*, 2022, pp. 474–475, doi: 10.1109/ICHI54592.2022.00074.

## BIOGRAPHIES OF AUTHORS






**Subramaniyan Senthamarai**    research scholar in the Department of Computer Applications at Alagappa University, Karaikudi, Tamil Nadu, India. She registered for her Ph.D. in 2020 and has been involved in teaching since 2007, with a total of over 17 years of academic experience. Her research interests include data mining, network security, and emerging trends in computer applications. She has published 8 research papers in international journals, authored 2 books, and has guided 2 M.Phil. scholars at Alagappa University. Mrs. Senthamarai has actively presented her research at national and international conferences and regularly contributes to reputed journals in her field. She can be contacted at [sakthisiva2125@gmail.com](mailto:sakthisiva2125@gmail.com).



**Raja Manickam Mala**    assistant professor and Head of the Department of Computer Science at Government Arts and Science College for Women, Paramakudi, Tamil Nadu, India, has 26 years of academic experience and has made extensive contributions to the field of computer science, particularly in emerging technologies and their practical applications. She has successfully guided 9 PhD and 15 M.Phil. scholars, published 17 papers in international journals, and authored 3 books. Her active engagement in the academic community is reflected through numerous national and international conference presentations and journal publications, underscoring her commitment to research and academic excellence. She can be contacted at [murugan.dcdrf@gmail.com](mailto:murugan.dcdrf@gmail.com).



**Vellaiyan Palanisamy**    professor and Head of the Department of Computer Applications at Alagappa University, Karaikudi, holds M.C.A., M. Tech., and Ph.D. degrees. With over 28 years of teaching and 22 years of research experience, his expertise spans network security, data mining, ad-hoc networks, biometrics, and algorithms. He has guided numerous Ph.D. and M.Phil. scholars and published over 120 international journal papers. A recipient of multiple prestigious awards, including the UGC Research Award, he has also served various academic leadership roles and continues to contribute actively to research and curriculum development. He can be contacted at [palaniyapp8atkdd@gmail.com](mailto:palaniyapp8atkdd@gmail.com).