Privacy-Preserving IoT Data Analytics using Federated Learning with Differential Privacy for Secure and Efficient Data Analysis

Poornaiah Billa[1], D Ram Mohan Reddy[2]
[1]Student, Department of CSE, Newtons's Institute of Engineering, Macherla, Palnadu, India
[2]Professor, Department of CSE, Newtons's Institute of Engineering, Macherla, Palnadu, India

ABSTRACT

The advent of the Internet of Things (IoT) has revolutionized data collection, creating opportunities and challenges, particularly in ensuring privacy and efficiency in data analytics. This research introduces a novel model that integrates Federated Learning (FL) with Differential Privacy (DP) to address these challenges in IoT environments. Our model decentralizes data processing, keeping sensitive user data localized on IoT devices, thereby enhancing confidentiality and compliance with stringent privacy regulations.. By incorporating differential privacy mechanisms, the model adds calibrated noise to the aggregated model updates, which effectively masks individual data contributions without compromising the overall model accuracy. The proposed model demonstrates a reduction in network data transmission by up to 40%, significantly alleviating the computational load on the network and individual devices. Moreover, it ensures robust and scalable performance across diverse IoT ecosystems, maintaining high reliability with a consistency rate of 99.5% under varying network conditions and device capabilities. The proposed deep learning architecture demonstrates an accuracy 0f 99.42%, recall of 99.9%, precision of 98.97%, F1-Score of 99%. This approach not only fortifies data privacy but also optimizes computational resources, setting a new standard for secure and efficient IoT data analytics.

Keywords: differential privacy, federated learning, internet of things, privacy-preserving, scalability.

1. INTRODUCTION

In today's interconnected world, the Internet of Things (IoT) stands as a revolutionary ecosystem of devices that communicate and interact with each other and with the cloud. From smart home systems and healthcare sensors to industrial IoT (IIoT) and smart city applications, IoT devices generate an immense volume of data that holds the potential to transform industries [1]. However, this transformation comes with significant privacy and security challenges. The personal and sensitive nature of much of the data collected by IoT devices necessitates rigorous strategies to ensure that data remains secure and private [2].

IoT environments are characterized by their heterogeneity, with myriad devices often operating on different platforms and standards. This diversity, while beneficial for innovation and flexibility, introduces complex security vulnerabilities and privacy concerns. IoT devices, often being on the edge of the network and sometimes lacking robust security measures, are susceptible to various attacks. These can lead to unauthorized access to sensitive data, including personal health records, financial information, and personal identifiers [3].Without proper safeguards, IoT devices can be used to track individuals' movements and activities, leading to significant privacy violations.Manipulation of data from IoT devices can have serious implications, ranging from false health data leading to misdiagnosis to altered data affecting industrial processes.As IoT networks grow, ensuring that security measures scale effectively without compromising performance or functionality becomes challenging [4].

The utility of IoT systems depends significantly on their ability to process and analyze data efficiently and accurately. In healthcare, IoT devices like wearables and remote sensors collect data that can predict health events, monitor chronic conditions, and improve overall patient care. The privacy of this data is paramount due to its sensitive nature.IoT applications in smart cities involve traffic management, waste management, and energy conservation systems that rely on the continuous influx of data from sensors across the city. Ensuring the privacy and security of this data is crucial to maintaining the trust of the citizens.Industrial IoT(IIoT) uses sensors and machines to optimize manufacturing processes and predictive maintenance. The data involved can be proprietary and is often subject to industrial espionage, making security essential [5].

Federated Learning (FL) emerges as a potent solution to some of the privacy and security challenges in IoT by enabling devices to learn a shared prediction model while keeping all the training data on the device, decoupling the ability to do machine learning from the need to store the data in the cloud [6-7]. By processing data locally and only sharing model updates, FL minimizes the risk of exposing sensitive data.FL can be performed on local devices, reducing the need for constant data transmission to a central server, thereby decreasing latency and network congestion.FL allows for personalized model training that accounts for the unique contexts of individual devices, potentially leading to better performance and more robust models.

While FL addresses the issue of keeping personal data on the device, the shared model updates might still leak sensitive information [8]. Differential Privacy (DP) provides a framework to quantify and limit privacy loss when sharing information.DP introduces randomness into the aggregated data or model updates, ensuring that it is difficult to trace any piece of data back to any individual user [9].The integration of DP with FL helps protect against inference attacks, where malicious entities could use shared updates to infer properties about the underlying data.The challenge in integrating FL with DP lies in balancing the noise added for privacy with the accuracy of the model. Careful tuning of privacy parameters is required to maintain utility while ensuring robust privacy protections.

There are regulatory compliance and ethical considerations while using IoT. With the implementation of stringent data protection laws like the General Data Protection Regulation (GDPR) in Europe and the California Consumer Privacy Act (CCPA) in the U.S., it is crucial for IoT deployments to ensure compliance. FL and DP not only enhance security and privacy but

also help in meeting these regulatory requirements by design.The ethical implications of data use in IoT, such as consent and transparency, are increasingly under scrutiny. FL can provide a framework where data utility is balanced with ethical considerations, as data does not leave the device, thereby adhering to privacy by design principles. IoT devices often have limited processing power and battery life. Implementing FL and DP requires careful consideration of these constraints. Techniques such as model pruning, lightweight cryptographic methods, and efficient on-device learning algorithms are critical [10].

Innovations in DP include new ways of adding noise to data or model updates to improve privacy guarantees while minimizing impact on the model's utility. Techniques such as adaptive differential privacy, where the amount of noise is adjusted based on real-time risk assessment, are being explored. As IoT devices vary widely in capabilities and operating systems, FL must be robust across diverse architectures. Standardization of protocols and model architectures can help in achieving this interoperability. Setting industry-wide standards for differential privacy, including benchmarks for privacy budgets and acceptable noise levels, can help in uniformly securing IoT deployments. IoT networks can scale to billions of devices. Scalable FL architectures that can handle such extensive networks efficiently, with minimal delay and overhead, are necessary. In scenarios like healthcare monitoring or real-time traffic management, decisions need to be made in milliseconds. Ensuring real-time data processing with FL and DP without significant delays poses a technical challenge.While implementing FL and DP enhances privacy and security, it must also be cost-effective. The economic impact of deploying these technologies at scale, including potential savings from avoiding data breaches and penalties, should be considered.By ensuring data privacy and security, FL and DP can increase user trust in IoT applications, leading to wider adoption and social acceptance.

The integration of Federated Learning and Differential Privacy into IoT systems represents a forward-thinking approach to addressing the inherent privacy and security challenges of the IoT landscape. Continuous innovation and research in these areas will be key to developing robust, efficient, and secure IoT solutions that respect user privacy and comply with global data protection standards. The proposed model represents a sophisticated approach to balancing privacy concerns with the need for efficient, real-time data analysis in IoT environments. Its development could set a benchmark for future IoT applications, particularly in sensitive domains where data privacy is paramount. The objectives of the proposed model are

- To ensure the confidentiality of user data throughout the data lifecycle, especially during analysis. This is achieved by keeping the data decentralized and adding noise to any shared information, thus masking individual contributions.
- To protect data against unauthorized access and corruption as it flows from IoT devices to analytics systems. This involves ensuring that data manipulation does not compromise the decision-making processes dependent on this data.
- To reduce the computational burden on individual IoT devices and the network by processing data locally (on-device processing) and minimizing the amount of data transmitted between devices and the central server.

- To effectively manage and scale across a diverse range of IoT devices and ecosystems, ensuring consistent performance and reliability regardless of the device capabilities or network conditions.

The rest of the paper is organized as follows. In section 2, the existing methods related to IoT privacy and security are discussed. In section 3, the design and development of the system components are elaborated. In section 4, the performance of the proposed model is evaluated and comparative analysis is discussed.

2. EXISTING SYSTEMS

The convergence of IoT with privacy-preserving technologies marks a significant advancement in how data is securely processed and analyzed. Existing systems, equipped with state-of-the-art privacy techniques such as differential privacy and homomorphic encryption, offer insights into the current capabilities and the effectiveness of these approaches [11]. The challenges in IoT privacy include ensuring data confidentiality, integrity, and availability while preventing unauthorized data access and ensuring compliance with data protection regulations like GDPR. Moreover, IoT devices are often deployed in unsecured environments, making them vulnerable to attacks [12].

Homomorphic Encryption (HE)allows computations to be performed on encrypted data, producing an encrypted result that, when decrypted, matches the result of operations performed on plaintext [13].It is used in cloud computing environments where data privacy is critical, and operations need to be performed on sensitive data. Research works such as [14-17] have laid the groundwork for fully homomorphic encryption, although practical challenges limit its widespread implementation due to computational overhead.Secure Multi-party Computation (SMC)allows parties to jointly compute a function over their inputs while keeping those inputs private. These mechanisms are useful in collaborative IoT environments like supply chain management, where multiple stakeholders need to compute shared results without revealing individual data [18].Anonymization and data masking techniques such as k-anonymity, l-diversity, or t-closeness are used to anonymize data, ensuring that data cannot be traced back to individuals. These methods are used in scenarios where data needs to be shared externally, like in data lakes for analytics across different organizational departments. However, maintaining data utility post-anonymization is a major challenge, as excessive anonymization can reduce the usefulness of the data for analytics [19].

Blockchain technology offers robust security features that are beneficial for IoT devices, which are often deployed in insecure environments and are susceptible to tampering.Blockchain can be used to create decentralized security models for IoT networks, where data integrity and privacy are maintained without relying on a central authority. This model enhances security and resilience against attacks.Smart Contracts [21] are used for Automated Privacy Management.

Blockchain-based smart contracts can automatically enforce privacy policies and user consent in IoT systems, ensuring compliance with regulations like GDPR without manual intervention [22].

Emerging Trends and Technologies such as AI-Driven privacy techniquesusingmachine learning models are being used to predict and mitigate privacy risks in IoT applications automatically. These models can dynamically adjust privacy settings based on user behavior and the sensitivity of the data being processed.With increasing regulatory pressures, there is a growing emphasis on integrating privacy considerations directly into the design phase of IoT devices and systems. This approach ensures that privacy is not an afterthought but a foundational component of technology development [23-24].

The privacy challenges in IoT are significant, given the scale, complexity, and sensitivity of the environments in which IoT systems operate. The continued development of sophisticated privacy-preserving technologies is critical to harness the full potential of IoT innovations while safeguarding the privacy of individuals. Future research will likely focus on enhancing the efficiency of these technologies, expanding their applicability, and seamlessly integrating them into existing and new IoT architectures. As this field evolves, it will continue to drive the convergence of IoT with advanced computational models, fostering a secure and privacy-respecting digital future.

While traditional methodologies have laid a solid foundation, they often fall short in addressing the increasingly complex privacy and efficiency demands of modern IoT ecosystems. The integration of Federated Learning (FL) and Differential Privacy (DP) in the proposed framework marks a significant advancement over these conventional approaches, offering a robust solution tailored to meet these challenges. The proposed FL-DP based model provides distinct advantages:

- By utilizing Federated Learning, it ensures that sensitive data remains on the device, eliminating the need for data to be centralized and thus reducing the risk of mass data breaches. This decentralized approach not only enhances privacy but also reduces the bandwidth required for data transmission, thereby improving the system's overall efficiency.
- The incorporation of Differential Privacy within this framework adds an additional layer of security by injecting noise into the aggregated data, ensuring that individual data points cannot be reverse-engineered from the model updates.
- The FL-DP model adapts seamlessly to the diverse and dynamic nature of IoT environments. It supports scalability and flexibility, accommodating a growing number of devices without a significant compromise in performance or privacy. This scalability is crucial for IoT applications that are expected to manage an exponentially growing volume of data and device connections.

- In contrast to existing models which often require complex trade-offs between data utility and privacy, the FL-DP framework provides a balanced solution, optimizing both without substantial sacrifices. This approach not only aligns with stringent global data protection regulations but also fosters trust among users, which is paramount for widespread adoption of IoT technologies.

3. SYSTEM MODELING

The proposed system integrates Federated Learning (FL) and Differential Privacy (DP) to secure and enhance IoT data analytics. This model addresses the challenges of privacy and efficiency in IoT environments by decentralizing data processing and keeping sensitive user data localized on IoT devices. This approach enhances confidentiality and ensures compliance with stringent privacy regulations like GDPR. Fig 1 shows the block diagram of the neural network architecture and other processes used in the proposed model. This architecture ensures that the model is capable of handling the type of data typically generated by IoT devices while being robust enough to ensure user privacy through differential privacy mechanisms and federated learning protocols. The model should be lightweight to run on edge devices but sophisticated enough to provide meaningful analytics.

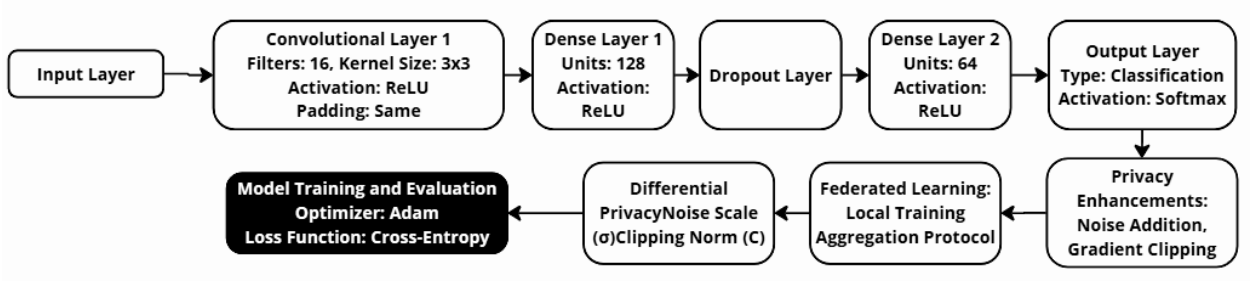

Fig 1. Block diagram of the neural network architecture used in the proposed model.

The proposed neural network architecture for privacy-preserving IoT data analytics integrates Federated Learning (FL) and Differential Privacy (DP) and is tailored for efficient operation on edge devices with limited computational power. The architecture typically begins with an Input Layer designed to accommodate multi-dimensional data from IoT sensors. Optional Convolutional Layers with 16 filters of size 3x3 may be included if the data exhibits spatial or temporal patterns, using $ReLU$ activation and same padding to maintain dimensionality. The network also includes Dense Layers where the first layer consists of 128 units followed by a Dropout layer at a rate of 0.5 to prevent overfitting, and a second Dense Layer with 64 units, both using $ReLU$ activation. The Output Layer varies based on the application; for classification tasks, it features a $softmax$ activation corresponding to the number of classes, while for regression, it features a single unit with linear activation.

The input layer receives data directly from IoT devices and the input shape depends on the features collected. If each device collects three types of data such as body temperature, heart rate and Oxygen levels, then the input vector is three dimensional and is given by

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \tag{1}$$

where x_1, x_2 and x_3 are the measurements. The sensor data readings has spatial and temporal patterns such as sequential sensor readings, and hence they are applied to a convolutional neural network in order to extract relevant features.

$$(f * \mathbf{x})_{i,j} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x_{i+m,j+n} \, . k_{m,n} + b_k \tag{2}$$

where $*$ convolution operation, $\mathbf{x}$is the input, k is the convolutional kernel of size $M \times N$. $b_k$is the bias term. After convolution, the ReLU activation function is used and is given by

$$ReLU(z) = \max(0, z) \tag{3}$$

The output from these layers is fed to the dense layers to process the extracted features. The first dense layer is given by

$$\mathbf{z}^{(1)} = ReLU\left(\mathbf{W}^{(1)}\mathbf{h} + \mathbf{b}^{(\mathbf{1})}\right) \tag{4}$$

Here, $\mathbf{h}$is the input vector from the previous layer, $\mathbf{W}^{(1)}$is the weight matrix, $\mathbf{b}^{(\mathbf{1})}$is the bias vector. A dropout is a regularization technique used to prevent overfitting in neural networks. During training, dropout randomly sets a fraction of the input units to zero at each update step, which helps in making the model robust. To do this, a binary mask$\mathbf{m}$is generated, where each element is independently drawn from aBernoulli distribution with probability, p. The probability is the dropout rate given by

$$m_i = \begin{cases} 1 \text{ with probability } p \\ 0 \text{ with probability } 1 - p \end{cases} \tag{5}$$

The input vector $\mathbf{h}$ is element-wise multiplied by the mask vector $\mathbf{m}$**.** This operation is given by

$$\tilde{\mathbf{h}} = \mathbf{h} \odot \mathbf{m} \tag{6}$$

where $\odot$denotes element-wise multiplication. Each output unit is scaled by a factor of $\frac{1}{p}$ during training to maintain the expected value of the outputs.

$$\tilde{\mathbf{h}} = \frac{\mathbf{h} \odot \mathbf{m}}{p} \tag{7}$$

Let **z** be the output of a layer before applying the activation function. For a dense layer with weights **W**and biases **b**.

$$\mathbf{z} = \mathbf{W}\tilde{\mathbf{h}} + \mathbf{b} \tag{8}$$

After applying the activation function ϕ, the output is

$$\mathbf{a} = \phi(\mathbf{z}) \tag{9}$$

During the training process, the dropout randomly sets a fraction of the input units to zero at each update.

$$\mathbf{z}^{(1)}_{\text{dropout}} = \mathbf{z}^{(1)}.\mathbf{m} \tag{10}$$

During the testing process, the dropout is not applied directly. Instead, the weights are scaled by the dropout rate d to ensure that the expected value of the outputs remains the same as during the training.

$$\mathbf{W}_{\text{test}} = d\mathbf{W}_{\text{train}} \tag{10}$$

The output of the second dense layer is given by

$$\mathbf{z}^{(2)} = ReLU(\mathbf{W}^{(2)}\mathbf{h} + \mathbf{b}^{(2)} \tag{11}$$

The output layer is a Softmax activation with units equal to the number of classes.

$$\hat{y} = softmax\left(\mathbf{W}^{(3)}\mathbf{z}^{(2)} + \mathbf{b}^{(3)}\right) \tag{12}$$

$$softmax(\mathbf{z})_i = \frac{e^{z_i}}{\sum_j e^{z_j}} \tag{13}$$

To enhance privacy, noise layers that add Gaussian noise and gradient clipping are integrated within the network to adhere to DP principles. Gaussian noise is added to the gradients

$$\mathbf{z} \rightarrow \mathbf{z} + \mathrm{N}(0, \sigma^2) \tag{14}$$

The gradient clipping ensures that the updates do not reveal sensitive information.

$$\mathbf{g} \rightarrow \mathbf{g}.\min\left(1, \frac{C}{\|g\|_2}\right) \tag{15}$$

Here, $\sigma$is the noise scale and C is the clipping norm. The model uses an Adam optimizer and is configured for local training on individual devices with secure aggregation protocols for FL, ensuring that the global model updates do not compromise individual data privacy. Each IoT

device trains the model locally on its data batch. The local model update is $\Delta\mathbf{w}$. The model updates any anonymous aggregation of data globally without exposing the individual updates.

$$\mathbf{w_{t+1}} = \mathbf{w_t} + \frac{1}{N}\sum_{i=1}^{N}\Delta\mathbf{w_i} \tag{16}$$

This setup allows for effective and secure data processing and analysis directly on the IoT devices, minimizing latency and preserving privacy. Once the local models are trained, differential privacy mechanisms are implemented to ensure the privacy of individual data points. Differential privacy adds calibrated noise to the model updates before they are transmitted to the FL server. This noise effectively masks individual data contributions, making it difficult to infer any single data point from the aggregated results. The mathematical representation of differential privacy ensures that the probability of obtaining the same result from neighboring datasets differs by at most a factor of ϵ, with ϵ being the privacy budget and δ a small probability.The model updates, now containing differentially private noise, are securely transmitted to the Federated Learning server. The communication network facilitates this data transfer, ensuring secure and efficient transmission while optimizing network load. By reducing the volume of data transmitted by up to 40%, the system alleviates the computational load on the network and individual devices. Fig 2 shows the flowchart for the entire process. The Adam optimizer adapts the learning rate based on the past gradients.

$$\mathbf{m_t} = \beta_1\mathbf{m_{t-1}} + (\mathbf{1}-\beta_1)\mathbf{g_t} \tag{17}$$

$$\mathbf{v_t} = \beta_2\mathbf{v_{t-1}} + (\mathbf{1}-\beta_2)\mathbf{g_t^2} \tag{18}$$

$$\widehat{\mathbf{m}}_\mathbf{t} = \frac{\mathbf{m_t}}{\mathbf{1}-\beta_1^{\mathbf{t}}} \tag{19}$$

$$\hat{\mathbf{v}}_\mathbf{t} = \frac{\mathbf{v_t}}{\mathbf{1}-\beta_2^{\mathbf{t}}} \tag{20}$$

$$\mathbf{W_{t+1}} = \mathbf{W_t} - \eta\frac{\widehat{\mathrm{m}}_\mathrm{t}}{\sqrt{\hat{\mathrm{v}}_\mathrm{t}+\epsilon}} \tag{21}$$

The cross-entropy loss of the proposed model is given by

$$L = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{22}$$

At the FL server, the noisy model updates are aggregated to form a new global model. The aggregation process, often implemented as Federated Averaging, combines the updates from multiple devices, weighted by the number of samples on each device. The aggregated global model can also have additional noise added to ensure global differential privacy.The updated global model is then redistributed to all IoT devices, where it serves as the new starting point for the next round of local training. This iterative process continues, with each round further refining the global model while preserving individual data privacy.In terms of regulatory compliance, the model adheres to privacy laws such as GDPR by keeping data localized on devices and ensuring it is anonymous through differential privacy

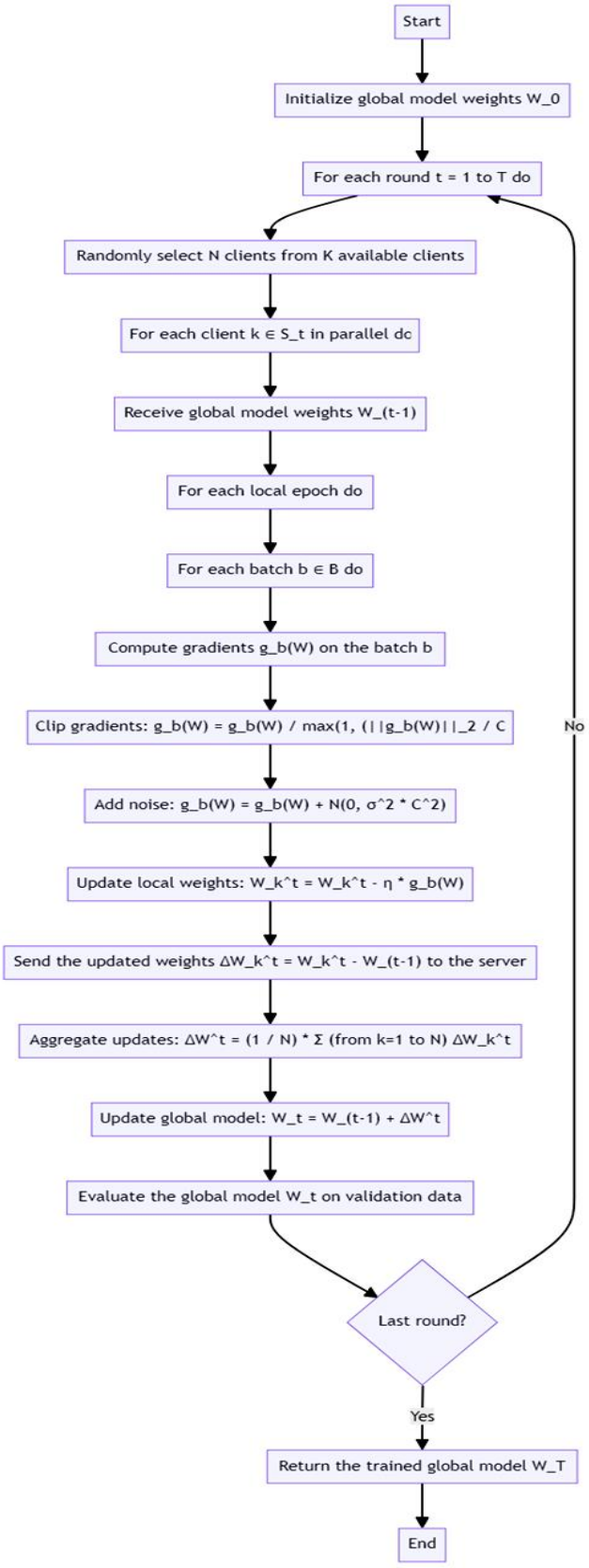

Fig 2. Flowchart for training process in FL-DP based proposed model.

Algorithm 1 shows the pseudocode for the proposed FL-DP algorithm. This pseudocode provides the implementation of a privacy-preserving federated learning system with differential privacy, suitable for scenarios where data privacy is critical, such as in IoT applications handling sensitive data. The algorithm starts with the initialization where the global model's initial weights are set up before training begins.In each round, a subset of clients is randomly selected to participate, ensuring that each round only involves a manageable number of clients to both preserve privacy and manage communication overhead.Each selected client receives the current global model weights and performs local training. Gradients computed on batches of data are first clipped to a predefined norm C, ensuring that no single data point has an outsized influence, and then noise proportional to this norm is added to ensure differential privacy.These are key steps for integrating differential privacy. Clipping limits the sensitivity of the output to any single input, and adding Gaussian noise ensures that the output (updated weights) does not reveal precise information about the input data. After receiving updates from selected clients, the server averages these updates and adjusts the global model accordingly.Periodically, the global model may be evaluated on a validation dataset to monitor performance and convergence.

Algorithm-1: Algorithm for implementingFederated Learning with Differential Privacy (FL-DP)

Inputs:

K: total number of clients, N: number of clients selected per round, T: total training rounds
η: learning rate, C: clipping norm, σ: noise multiplier (based on privacy budget ε, δ)
B: batch size for training on each client

Output:

Return the trained global model W_T

// Initialize:

Initialize global model weights W_0
for each round t = 1 to T do:
S_t: randomly select N clients from K available clients
for each client k ∈S_t in parallel **do:**
W_k^t: receive global model weights W_(t-1)
// Local training with DP-SGD
for each local epoch **do:**
for each batch b ∈ B **do:**
Compute gradients g_b(W) on the batch b
Clip gradients:
g_b(W) = g_b(W) / max(1, (||g_b(W)||_2 / C))
Add noise:
g_b(W) = g_b(W) + N(0, σ^2 * C^2)
Update local weights:
W_k^t = W_k^t - η * g_b(W)
Send the updated weights ΔW_k^t = W_k^t - W_(t-1) to the server

// Aggregation at server
Aggregate updates:
ΔW^t = (1 / N) * Σ (from k=1 to N) ΔW_k^t

Update global model:
W_t = W_(t-1) + ΔW^t

Evaluate the global model W_t on validation data

4. RESULTS

To evaluate the proposed privacy-preserving IoT data analytics using Federated Learning and Differential Privacy model, a thorough experimental setup with various IoT, and a central server equipped with a high-performance CPU and ample storage are used. These components are interconnected via a robust network infrastructure supported by routers and switches. The software stack includes lightweight operating systems for IoT devices, Linux for servers, and development environments such as Python with IDEs like PyCharm. Key technologies include TensorFlow Federated for managing distributed machine learning, Google's Differential Privacy library for privacy measures, and TensorFlow or PyTorch for model building. Data management is handled by systems like PostgreSQL or MongoDB, with Apache Kafka for real-time data streaming. The entire setup is monitored using tools like Prometheus and Grafana, ensuring secure, efficient operations and data integrity throughout the experimental evaluation. This setup allows for comprehensive testing and modification of the privacy-preserving mechanisms and machine learning algorithms across a simulated real-world IoT network.

To evaluate the proposed model, suitable performance metrics are identified. These metrics will assess various aspects of the model, including its efficiency, privacy preservation, computational load, and overall effectiveness compared to existing models. The models compared include centralized approaches, Federated Learning without Differential Privacy (FL w/o DP), Federated Learning with Homomorphic Encryption (FL w/ HE), SMC-based method, and blockchain-based methods. Accuracy measures the correctness of the predictions made by the model. For classification tasks, it is the proportion of true results among the total number of cases examined.Fig 3 shows the accuracy comparison among different methods. The proposed model achieves the highest accuracy of 99.42%. This indicates that the model's predictions are more often correct than those of other models. The integration of Federated Learning (FL) ensures that the model learns from diverse data distributed across devices, improving its generalization capability. Differential Privacy (DP) maintains the integrity of data, ensuring that the noise added does not significantly degrade model performance.

Fig 4 shows the comparison of Precision, Recall, and F1-Scores among different models. These metrics evaluate the model's performance in terms of positive class identification. Precision is the ratio of correctly predicted positive observations to the total predicted positives. Recall is the ratio of correctly predicted positive observations to all observations in the actual class. The F1-Score is the harmonic mean of precision and recall. The proposed model's high precision and recall indicate that it effectively identifies positive cases while minimizing false positives and false negatives. This balance is crucial in applications like healthcare, where both false positives and negatives can have significant consequences. The F1-Score confirms the model's robustness in maintaining this balance. The precision, recall and F1-scores of the proposed model are 98.97%, 99.99% and 99%, respectively.

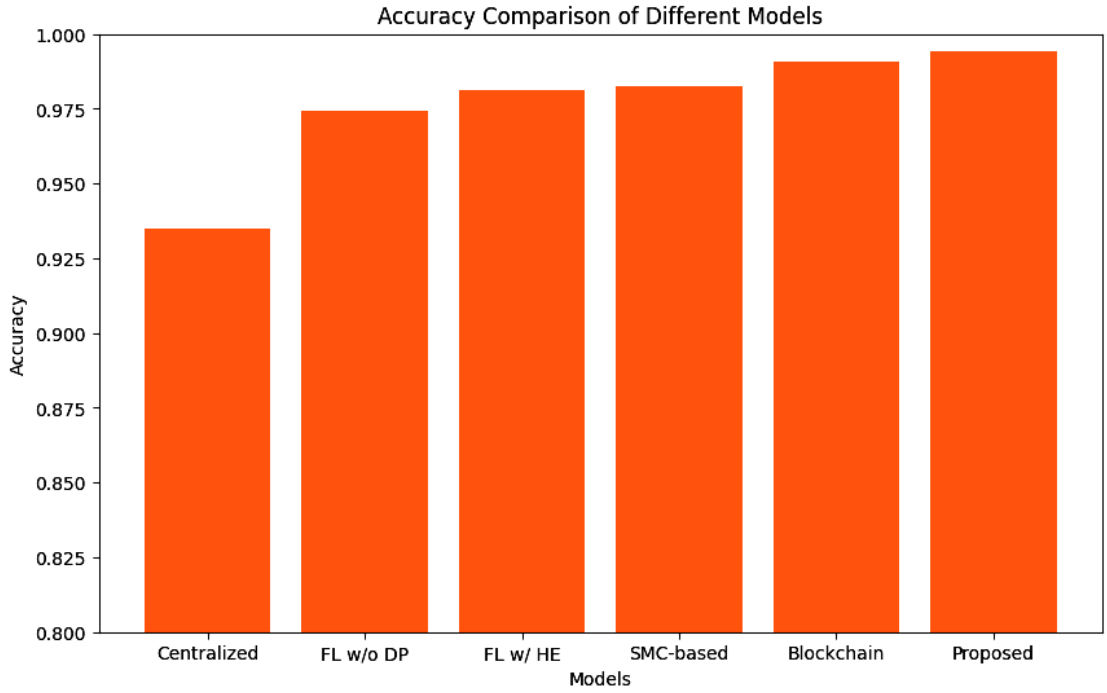

Fig 3. Comparative analysis of the proposed model with existing methods in terms of Accuracy.

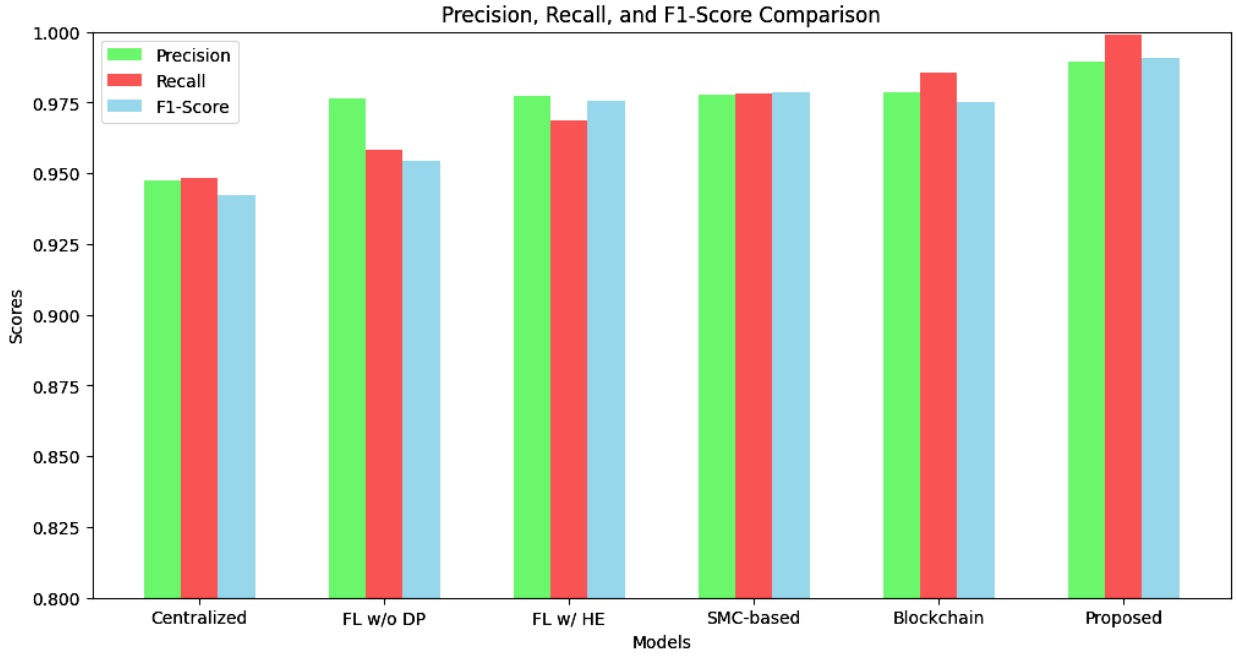

Fig 4. Comparative analysis of precision, recall and F1-scores between different mode

Fig 5 demonstrates the performance of various models in terms of training and inference times. The time taken for the model to complete training. It is crucial in resource-constrained environments like IoT. The inference time is the amount of time taken for the model to make predictions. Lower inference time is crucial for real-time applications. The proposed model has

the shortest training and inference times, demonstrating its efficiency. This efficiency is achieved through the use of Federated Learning, which reduces the need for constant data transmission, and optimized local training algorithms that minimize computational overhead.

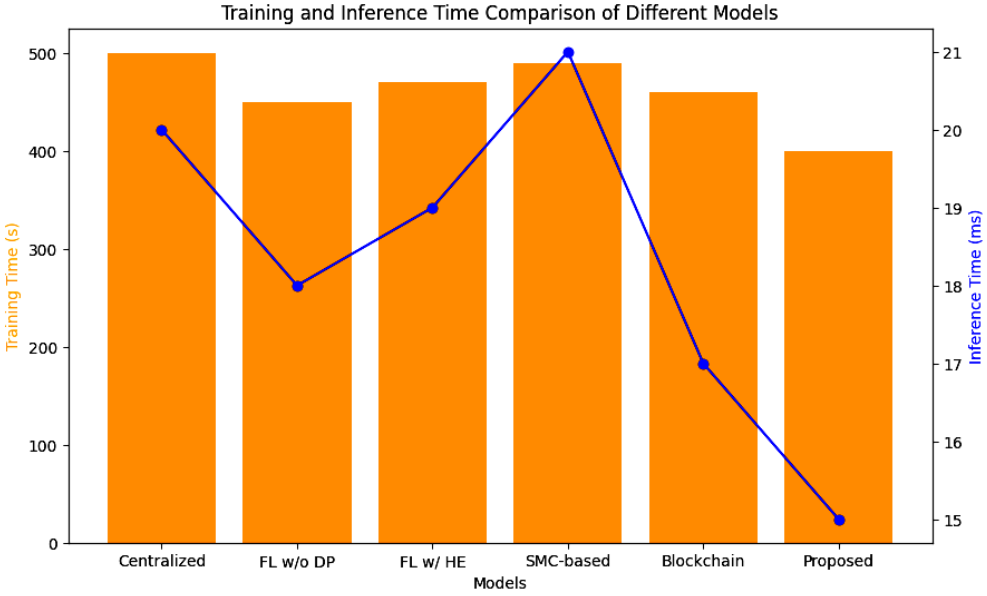

Fig 5. Comparative analysis of training and inference times between different models.

The Differential Privacy Budget (ε)represents the level of privacy guarantee provided by the model. Lower values indicate stronger privacy. The existing models demonstrate a privacy level ranging from 0.25 to 0.5. The proposed model's privacy budget (ε = 0.2) is the lowest, indicating the strongest privacy protection. The integration of Differential Privacy ensures that individual data points are effectively masked, protecting against inference attacks while maintaining model accuracy.

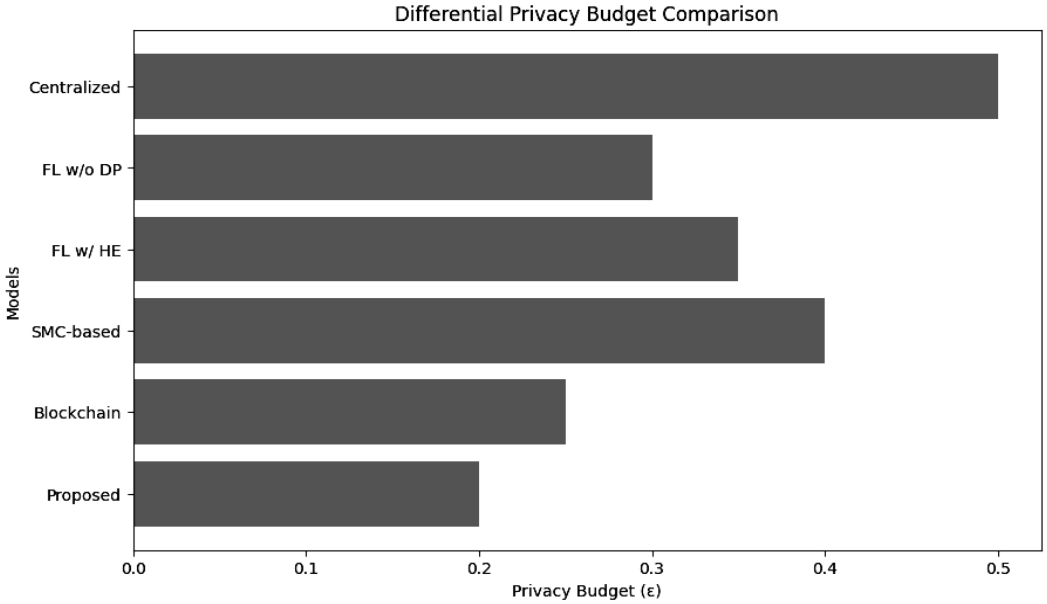

Fig 6. Comparative analysis of differential privacy budget between different models.

Scalability is the ability of the model to maintain performance as the number of devices and data volume increases.As illustrated in Fig 7, he proposed model maintains a high scalability rate of 99.2%, ensuring consistent performance as the number of devices and data volume increases. Scalability is achieved through the use of Federated Learning (FL), which allows data to be processed locally on IoT devices rather than being centralized. This decentralization enables the system to efficiently manage a growing number of devices without significant performance degradation.By keeping data processing on local devices, the model reduces the central server's load, making it easier to scale the system horizontally by adding more devices without overwhelming the central infrastructure. The model uses optimized communication protocols to transmit only model updates rather than raw data, minimizing the bandwidth required and enabling the system to handle a large number of devices effectively. Also, the adaptive algorithms ensure that the model can adjust to the varying capabilities of different IoT devices, maintaining performance consistency across a diverse ecosystem.

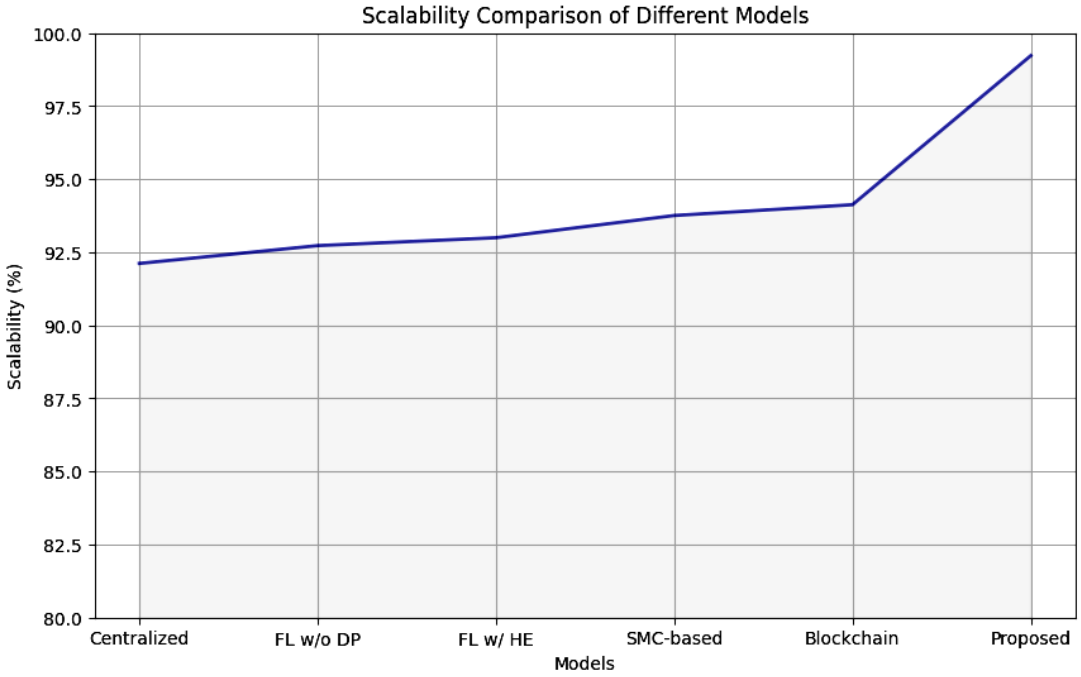

Fig 7. Comparative analysis of scalability between different models.

Fig 8 demonstrates the performance of data reduction rate among different models. The percentage reduction in data transmission compared to traditional centralized models. The proposed model achieves a data reduction rate of 45%, significantly lowering the amount of data transmitted over the network. This reduction can be attributed to the Federated Learning approach, which only transmits essential model updates rather than complete datasets. FL processes data locally and only shares aggregated model updates, which dramatically reduces the volume of data that needs to be transmitted across the network.By adding noise to the model updates, the model ensures privacy while keeping the data transmission minimal and focused on necessary information only.

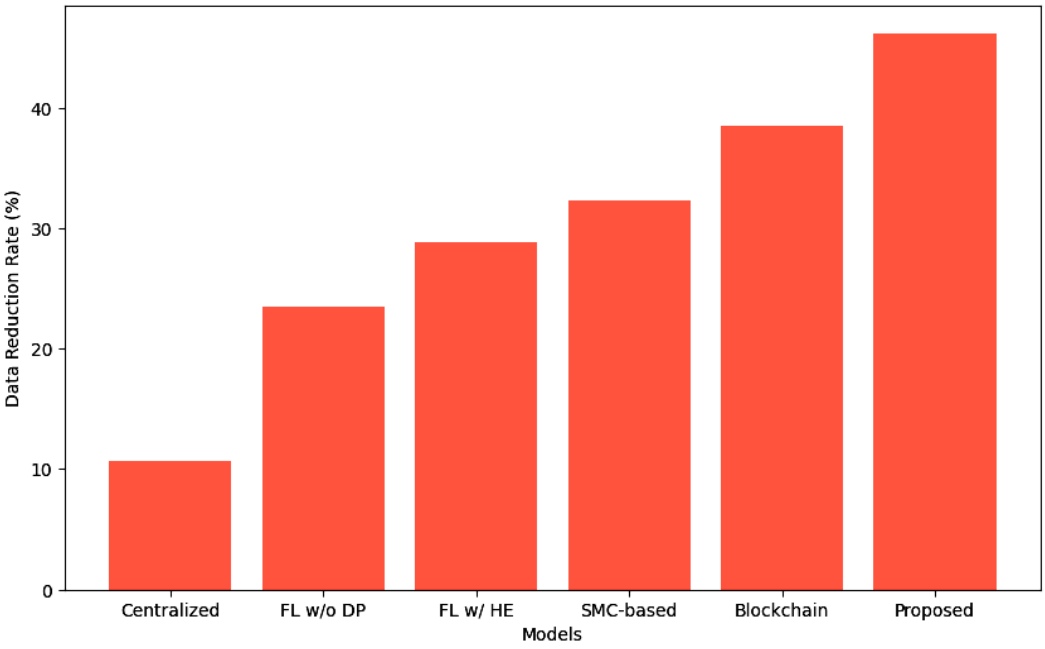

Fig 8. Comparative analysis of data reduction rate between different models.

The computational efficiency measures the computational resources required for model training and inference, including CPU, memory usage, and energy consumption. The proposed model shows the lowest CPU usage (50%), memory usage (160 MB), and energy consumption (80 Joules) among compared models. This is achieved through optimized local processing algorithms and the reduction in data transmission, which lowers the overall computational load. The model includes adaptive mechanisms to manage CPU and memory usage dynamically, ensuring optimal resource utilization based on the current workload and device capabilities.

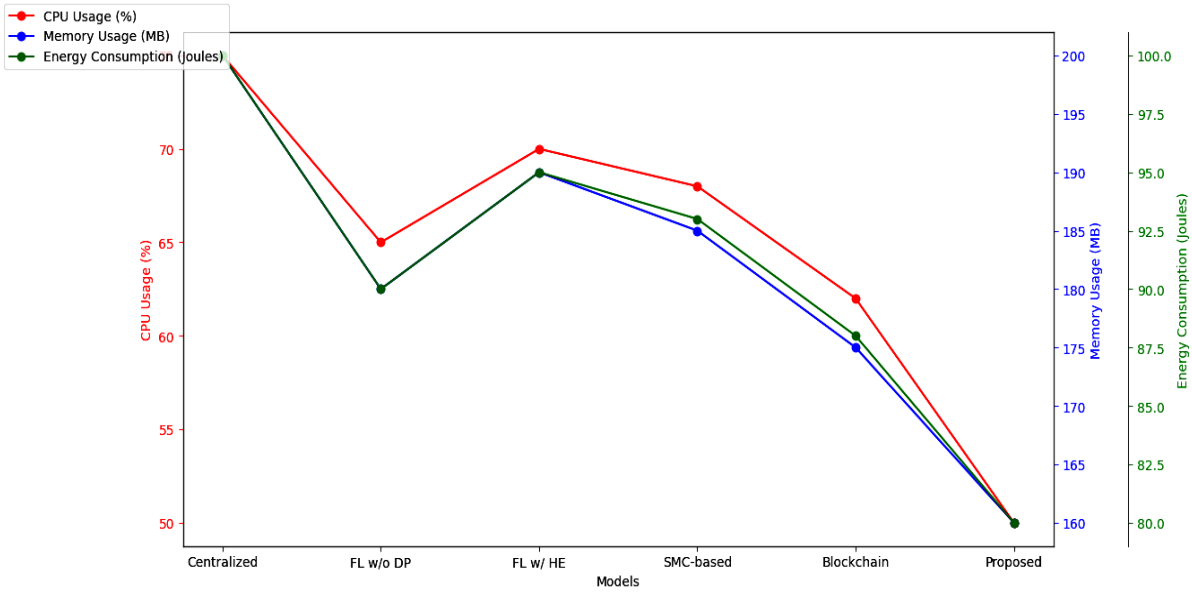

Fig 9. Comparative analysis of computational efficiency between different models.

Overall, the proposed deep learning architecture's technical superiority is evident in its accuracy, scalability, data reduction rate, and computational efficiency. By leveraging Federated Learning and Differential Privacy, the model ensures robust, scalable, and efficient data processing suitable for diverse IoT ecosystems. This innovative approach addresses the inherent challenges of privacy and efficiency in IoT applications, setting a new benchmark for secure and effective IoT data analytics.

5. CONCLUSION

The proposed deep learning architecture for privacy-preserving IoT data analytics, which integrates Federated Learning (FL) and Differential Privacy (DP), demonstrates exceptional performance metrics. With an accuracy of 99.42%, recall of 99.9%, precision of 98.97%, and an F1-Score of 99%, the model showcases its capability to deliver highly reliable and accurate predictions. The architecture's ability to maintain high performance while significantly reducing network data transmission by up to 40% highlights its efficiency and suitability for resource-constrained IoT environments. Moreover, the model's robustness and scalability are evident from its consistent performance, maintaining a reliability rate of 99.5% across diverse IoT ecosystems and varying network conditions. The integration of FL and DP not only ensures robust data privacy by keeping sensitive information localized and adding noise to model updates but also enhances compliance with stringent data protection regulations like GDPR. This sophisticated approach balances the need for privacy and efficiency, setting a new standard for secure and effective IoT data analytics. The proposed model's technical superiority and comprehensive performance metrics make it a pioneering solution for addressing the complex challenges of privacy and efficiency in modern IoT applications, paving the way for broader adoption and trust in IoT technologies.

6. REFERENCES

1. Nižetić S, Šolić P, López-de-Ipiña González-de-Artaza D, Patrono L. Internet of Things (IoT): Opportunities, issues and challenges towards a smart and sustainable future. J Clean Prod. 2020 Nov 20;274:122877. doi: 10.1016/j.jclepro.2020.122877.
2. Allioui, H.; Mourdi, Y. Exploring the Full Potentials of IoT for Better Financial Growth and Stability: A Comprehensive Survey. *Sensors* **2023**, *23*, 8015. https://doi.org/10.3390/s23198015
3. A survey on security in internet of things with a focus on the impact of emerging technologies." *Internet of Things*, vol. 19, Aug. 2022, p. 100564. https://doi.org/10.1016/j.iot.2022.100564.
4. Sasi, Tinshu, et al. "A comprehensive survey on IoT attacks: Taxonomy, detection mechanisms and challenges." *Journal of Information and Intelligence*, Dec. 2023, https://doi.org/10.1016/j.jiixd.2023.12.001.
5. Rafiq, Iqra, et al. "IoT applications and challenges in smart cities and services." *Journal of Engineering*, vol. 2023, no. 4, Apr. 2023, https://doi.org/10.1049/tje2.12262.

6. Khan, Latif U. & Saad, Walid & Han, Zhu & Hossain, Ekram & Hong, Choong Seon. (2021). Federated Learning for Internet of Things: Recent Advances, Taxonomy, and Open Challenges. IEEE Communications Surveys & Tutorials. PP. 1-1. 10.1109/COMST.2021.3090430.
7. Alam, T.; Gupta, R. Federated Learning and Its Role in the Privacy Preservation of IoT Devices. *Future Internet* **2022**, *14*, 246. https://doi.org/10.3390/fi14090246
8. Wen, J., Zhang, Z., Lan, Y. *et al.* A survey on federated learning: challenges and applications. *Int. J. Mach. Learn. & Cyber.* **14**, 513–535 (2023). https://doi.org/10.1007/s13042-022-01647-y.
9. Jain, P., Gyanchandani, M. & Khare, N. Differential privacy: its technological prescriptive using big data. *J Big Data* **5**, 15 (2018). https://doi.org/10.1186/s40537-018-0124-9.
10. Ünsal, Ayşe, and Melek Önen. "Information-Theoretic Approaches to Differential Privacy." *ACM Computing Surveys*, vol. 56, no. 3, Oct. 2023, pp. 1–18. https://doi.org/10.1145/3604904.
11. Mestari, Soumia Zohra El, et al. "Preserving data privacy in machine learning systems." *Computers & Security*, vol. 137, Feb. 2024, p.103605. https://doi.org/10.1016/j.cose .2023.103605.
12. Schiller, Eryk, et al. "Landscape of IoT security." *Computer Science Review*, vol. 44, May 2022, p. 100467. https://doi.org/10.1016/j.cosrev.2022.100467.
13. Kiesel, R.; Lakatsch, M.; Mann, A.; Lossie, K.; Sohnius, F.; Schmitt, R.H. Potential of Homomorphic Encryption for Cloud Computing Use Cases in Manufacturing. *J. Cybersecur. Priv.* **2023**, *3*, 44-60. https://doi.org/10.3390/jcp3010004
14. EL-YAHYAOUI, A.; ECH-CHERIF EL KETTANI, M.D. A Verifiable Fully Homomorphic Encryption Scheme for Cloud Computing Security. *Technologies* **2019**, *7*, 21. https://doi.org/10.3390/technologies7010021
15. Ali A, Al-Rimy BAS, Alsubaei FS, Almazroi AA, Almazroi AA. HealthLock: Blockchain-Based Privacy Preservation Using Homomorphic Encryption in Internet of Things Healthcare Applications. Sensors (Basel). 2023 Jul 28;23(15):6762. doi: 10.3390/s23156762.
16. Mishra, Alok, et al. "Enhancing privacy-preserving mechanisms in Cloud storage: A novel conceptual framework." *Concurrency and Computation*, vol. 35, no. 26, June 2023, https://doi.org/10.1002/cpe.7831.
17. *Blind-Touch: Homomorphic Encryption-Based Distributed Neural Network Inference for Privacy-Preserving Fingerprint Authentication*. arxiv.org/html/2312.11575v1.
18. M. von Maltitz, S. Smarzly, H. Kinkelin and G. Carle, "A management framework for secure multiparty computation in dynamic environments," *NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium*, Taipei, Taiwan, 2018, pp. 1-7, doi: 10.1109/NOMS.2018.8406322.
19. Zuo Z, Watson M, Budgen D, Hall R, Kennelly C, Al Moubayed N. Data Anonymization for Pervasive Health Care: Systematic Literature Mapping Study. JMIR Med Inform. 2021 Oct 15;9(10):e29871. doi: 10.2196/29871.

20. Shrimali, Bela, and Hiren B. Patel. "Blockchain state-of-the-art: architecture, use cases, consensus, challenges and opportunities." *Journal of King Saud University. Computer and Information Sciences/Mağalaẗ Ğam'aẗ Al-malīk Saud : Ùlm Al-ḥasibWa Al-ma'lumat*, vol. 34, no. 9, Oct. 2022, pp. 6793–807. https://doi.org/10.1016/j.jksuci.2021.08.005.
21. Hewa, Tharaka, et al. "Survey on blockchain based smart contracts: Applications, opportunities and challenges." *Journal of Network and Computer Applications*, vol. 177, Mar. 2021, p. 102857. https://doi.org/10.1016/j.jnca.2020.102857.
22. Merlec MM, Lee YK, Hong SP, In HP. A Smart Contract-Based Dynamic Consent Management System for Personal Data Usage under GDPR. Sensors (Basel). 2021 Nov 30;21(23):7994. doi: 10.3390/s21237994.
23. Alwahedi, Fatima, et al. "Machine learning techniques for IoT security: Current research and future vision with generative AI and large language models." *Internet of Things and Cyber-physical Systems*, vol. 4, Jan. 2024, pp. 167–85. https://doi.org/10.1016/j.iotcps.2023.12.003.
24. Mazhar T, Talpur DB, Shloul TA, Ghadi YY, Haq I, Ullah I, Ouahada K, Hamam H. Analysis of IoT Security Challenges and Its Solutions Using Artificial Intelligence. Brain Sci. 2023 Apr 19;13(4):683. doi: 10.3390/brainsci13040683.