WCRG Global Research
WORLD COMET RESEARCH GROUP

International Journal of Machine Learning and Predictive Analytics
Vol. 8, No. 1, 2025
Doi: 10.67228/3142788X/IJMLPA-2025PI6L4G
PP. 01-15

Original Article

Federated Predictive Learning with Privacy-Aware Model Aggregation for Distributed Analytics

H. N. Mahabala[1], N. Seshagiri[2]

[1]*Computer Scientist, Tata Institute of Fundamental Research, India.*

[2]*Information Technology Pioneer, National Informatics Centre, India.*

Received: 12-03-2024 Revised: 12-25-2024 Accepted: 01-01-2025 Published: 01-03-2025

ABSTRACT

The rapid growth of IoT, edge computing, healthcare, and cyber-physical systems has increased the need for privacy-preserving distributed analytics. Traditional centralized machine learning requires sharing raw data, creating privacy, regulatory, and communication challenges. Federated Learning (FL) enables collaborative model training without exchanging sensitive data but faces limitations such as Non-IID data, communication overhead, privacy leakage, malicious updates, and inefficient aggregation. To address these issues, this study proposes a Federated Predictive Learning with Privacy-Aware Model Aggregation (FPL-PAMA) framework. The framework combines secure local training, privacy-aware weighted aggregation, client trust evaluation, and adaptive optimization to improve prediction accuracy, privacy protection, and communication efficiency. It is suitable for applications including healthcare, IoT, smart manufacturing, transportation, and financial fraud detection, providing a secure and scalable solution for next-generation distributed intelligent systems.

KEYWORDS

Federated Learning, Distributed Analytics, Privacy Preservation, Predictive Learning, Model Aggregation, Secure Machine Learning, Edge Computing, Artificial Intelligence, Distributed Optimization, Federated Averaging, Privacy-Aware Aggregation, Machine Learning Security.

WCRG
Global Research
ISSN: 3142-788X
PP: 01-15.

1. INTRODUCTION

1.1. Background

The proliferation of cloud computing, edge computing and the Internet of Things (IoT) has revolutionized the way data is created, computed and used to make intelligent decisions. Today, enormous amounts of structured and unstructured data are generated across geographically distributed devices and organizations in continuous, real-time streams in modern healthcare, financial, industrial automation, smart city, transportation and mobile computing applications. These datasets are traditionally gathered and stored in a central cloud server to start training models with the centralized machine learning techniques. While this technique allows for large scale data analysis, it comes with a number of disadvantages such as high communication overhead, latency, storage requirements, and high computational costs. Perhaps more importantly, sharing information like medical data, financial transactions, and personal user data with a central server introduces significant privacy and security issues and complicates adherence to data protection laws. This has led to the research and development of distributed learning frameworks capable of collaborative model training without compromising data privacy and eliminating reliance on centralized data storage.

1.2. Needs of Federated Predictive Learning with Privacy-Aware

Distributed computing environments have become more prevalent, and there is rising demand for machine learning methods that can make accurate predictions while maintaining the privacy and security of the data. Traditional centralized approaches also involve uploading raw data to a central server, exposing a potential data leakage risk, data access issues, and potentially various legal and regulatory compliance concerns. The Federated Predictive Learning with Privacy-Aware Model Aggregation (FPL-PAMA) framework aims to solve these problems by allowing for model training without sharing and without providing privacy guarantees. Several practical needs exist to motivate the need for such a framework in modern distributed applications.

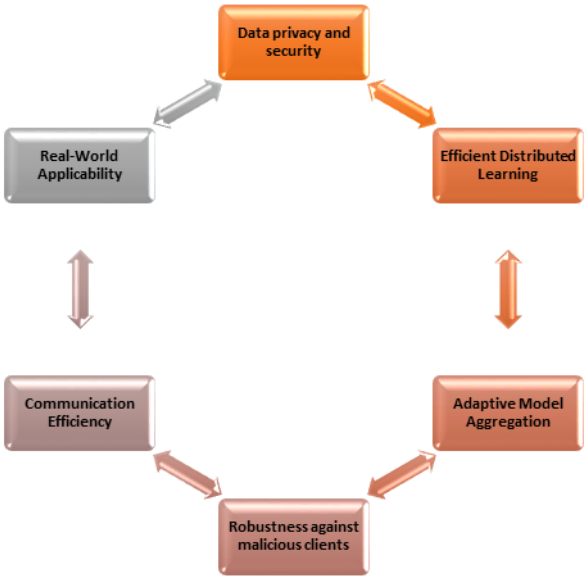

Fig 1 - Needs of Federated Predictive Learning with Privacy-Aware

ISSN: 3142-788X
PP: 01-15.

1.2.1. Data privacy and security

Certain institutions like hospitals, banks, and government bodies have access to sensitive data that cannot be released owing to privacy laws and safety risks. A privacy preserving federated learning scheme guarantees that model parameters are encrypted when passed between clients and that the raw data stays local. This greatly lowers the risk of data breaches and guarantees the privacy of confidential data during the learning process.

1.2.2. Efficient Distributed Learning

Distributed systems produce enormous amounts of data at various locations, which is costly and inefficient to collect data from a single location. Federated predictive learning allows each client to learn models locally from its own data and collectively upload the updated parameters to a global model without compromising client data privacy. This is distributed learning strategy reduces communication costs and efficiently utilise the local computational resources.

1.2.3. Adaptive Model Aggregation

In traditional federated learning algorithms, all participating clients are treated equally, even if they have varying data quality or reliability. The actual environment in which clients run is different in terms of data distribution, computational power, communication quality and trustworthiness. They then need to develop an adaptive privacy-aware aggregation mechanism that will assess the contribution of each client and weight them to generate more accurate and reliable global models.

1.2.4. Robustness against malicious clients

Malicious agents can send false or altered model updates to reduce the global model accuracy in federated learning systems. A privacy-preserving framework also includes trust evaluation and reliability assessment to detect suspicious client actions and minimise the impact of malicious updates. This enhances the robustness, security and stability of the collaborative learning process.

1.2.5. Communication Efficiency

The exchange of frequent messages between the clients and the aggregation server may lead to higher network traffic and slow down the model convergence. Federated predictive learning aims to optimize the communication in a privacy-aware way by focusing on accurate model updates and minimizing redundant transmissions. This reduces the communication overhead and accelerates the convergence, making this a framework ideal for large-scale distributed environments.

1.2.6. Real-World Applicability

Collaborative intelligence without sacrificing user privacy is vital in many real-world applications such as healthcare, financial services, industrial Internet of Things (IIoT), smart transportation, and mobile edge computing. The challenge of achieving prediction accuracy, privacy protection, computational efficiency, and communication optimization in these domains is addressed by a privacy-aware federated predictive learning framework that is scalable and practical. Therefore, it facilitates secure and trustworthy AI in contemporary distributed systems.

ISSN: 3142-788X
PP: 01-15.

1.3. Problem Statement

Federated Learning (FL) is a distributed machine learning paradigm with the potential of avoiding the sharing of data that offers effective solutions to some of these challenges, but it also presents a number of problems that pose the need for continued research. Currently most federated-learning algorithms, such as Federated Averaging (FedAvg), require all participating clients to contribute equally to the global learning process and they have the same data distribution, computational power, and communication capabilities. This, however, is not always true in real distributed systems. For the real world applications like healthcare, financial, IoT in industry, mobile edge computing, and smart transportation, clients often have heterogeneous datasets, with different processing power, network conditions, privacy concerns etc. This makes it difficult to obtain optimal global models using traditional aggregate approaches since they fail to differentiate between reliable and unreliable updates from clients. Moreover, there is a possibility of malicious clients or compromised clients that can supply incorrect or manipulated model parameters, causing model poisoning attacks that reduce prediction accuracy and impact system reliability. Current aggregation mechanisms also show little sensitivity to trustworthiness of the clients, communication quality, historical learning performance, and privacy sensitivity during the aggregation process. While some privacy-preserving methods include Differential Privacy, Secure Multi-Party Computation and Homomorphic Encryption, they often require extra computational resources, communication latency and added resources, thus posing challenges for large-scale distributed deployment. Furthermore, the majority of current federated learning methods have been tested in theoretical experiments and failed to account for the dynamic behaviors of heterogeneous devices in real distributed systems. The restrictions limit scalability, hinder model convergence, raise the communication expenses, and diminish the robustness of collaborative learning. Hence, the necessity of a smart federated learning system considering both client privacy and trustworthiness, communication reliability and model quality during aggregation. It should be able to allocate the adaptive aggregation weights according to a detailed evaluation of the clients and increase the prediction accuracy, preserve privacy, reduce the communication overhead, speed up the convergence and increase robustness to malicious clients. The proposed Federated Predictive Learning with Privacy-Aware Model Aggregation (FPL-PAMA) framework is motivated to address these challenges to enable next generation distributed predictive analytics to be secure, scalable, and efficient.

2. LITERATURE SURVEY

The recent progress in distributed AI has greatly accelerated research efforts in federated learning and privacy-preserving predictive analytics. As more and more organizations seek to collaborate in the development of machine learning systems that do not leak sensitive data, researchers have begun to work on decentralized methods to train intelligent models together without sharing sensitive data. Much work has been done for decentralized optimization, secure aggregation, communication-efficient learning and trustworthy AI. Although these advances have been made, problems with data heterogeneity, privacy protection, communication cost, and adaptive model aggregation are all still unsolved. In this section, the main achievements on Federated

ISSN: 3142-788X
PP: 01-15.

Learning, privacy-preserving machine learning, model aggregation methods and the gaps that inspired the proposed framework are reviewed.

2.1. Federated Learning

Federated learning (FL), conceived by Google, is a machine learning method that allows multiple clients to train a common global model using their local data without having to share it with one another. Each client who participates in the training process trains the model locally, and only sends new parameters or gradients to a central aggregation server. These updates are then aggregated by the server to produce a better global model, and then sent back to all the contributing clients for the next training round. This is a learning process, and it can greatly mitigate the privacy risks and facilitate collaborative intelligence in distributed environments. In the past few years, a number of high-level federated learning algorithms have been proposed to enhance the convergence speed, communication overhead, and the problem of statistical heterogeneity among clients, such as Federated Averaging (FedAvg), FedProx, FedNova, FedDyn and SCAFFOLD. In the Healthcare industry, Federated learning has been successfully used for diagnosis, in Financial Fraud detection, in Intelligent Transportation Systems, in Recommendation Systems, in Mobile Edge Computing, in Smart Manufacturing, and in Industrial Internet of Things (IIoT) Applications. However, there are still several challenges to be addressed for practical deployment, such as the Non-IID data distribution, the different computational capacities of clients, the intermittent participation of clients, and the limitations in communication between them and the server. These factors also influence the performance and scalability of models.

2.2. Privacy-Preserving Machine Learning

One of the most crucial goals of federated learning is to maintain privacy, as the data stored on different clients is not brought together into a single central location. Research has shown that even if raw data are not transmitted, model updates transmitted via the gradient may still leak confidential information via model reconstruction and gradient inversion attacks. Some privacy preserving methods have been added to federated learning frameworks to address these security issues. Differential Privacy introduces carefully designed statistical noise into the updates of the model before they are transmitted to preserve privacy of individual data records, minimizing the risk of information leakage. Secure Multi-Party Computation (SMPC) provides a way for multiple parties to jointly execute computations without exposing any private data and Homomorphic Encryption is a way to perform mathematical operations directly on encrypted model parameters without having to decrypt them. Additionally, trusted execution environments (TEEs) offer hardware-level separation for critical computations, increasing security. While these methods significantly enhance data confidentiality and regulatory adherence, they may also introduce challenges in terms of computational complexity, communication latency, memory usage, and energy consumption. Thus, the design of privacy-preserving machine learning systems that preserve high accuracy in prediction with the minimum computation cost remains an important research challenge.

ISSN: 3142-788X
PP: 01-15.

2.3. Model Aggregation Techniques

The main challenge of federated learning is model aggregation, which involves updating the global model with local updates from various clients to achieve an accurate and robust global model. Federated Averaging (FedAvg) is the most popular aggregation algorithm, which takes a weighted average of the parameters of the local models based on the size of the dataset on the client. But FedAvg requires relatively homogeneous client data and equal participation, which is not optimal in scenarios with Non-IID data distributions, intermittent network connectivity, and varying computational power. To address these limitations, several advanced aggregation strategies have been proposed to enhance the optimization in heterogeneous conditions, including FedProx, FedNova, FedDyn, and SCAFFOLD. Furthermore, adaptive aggregation methods also include reinforcement learning, historical model performance, communication quality, and gradient similarity, as well as the importance of clients that are changing dynamically based on each client's scores. Further, strong aggregation functions such as Krum, Trimmed Mean Aggregation, Median Aggregation and Byzantine-resilient optimization algorithms have been developed to safeguard against malicious or compromised clients trying to "poison" the global model. Recently, blockchain technology, attention mechanisms, and artificial intelligence-driven trust evaluation have been woven into an aggregation framework to enhance transparency, reliability, and decentralized decision-making. In spite of these developments, many existing aggregation techniques are mainly concerned with prediction accuracy and underestimate the privacy sensitivity of the clients, as well as the dynamic behavior of certain clients or a complete trust evaluation.

2.4. Research Gap

Current federated learning systems have been noted for various key weaknesses in real-world distributed analytics that limit their effectiveness. The current aggregation algorithms focus mostly on prediction accuracy and offer only minimal support for privacy awareness, adaptive trust evaluation, communication efficiency and client reliability in a single framework. In many studies the aggregation policies are static, that is, the importance of the participating clients is fixed in time without considering time-variations in client behavior, network quality, computational capability and data reliability. In addition, while these privacy-preserving approaches – including Differential Privacy, Secure Multi-Party Computation, and Homomorphic Encryption – boost data security, they often come with extra computational expenses and communication delays, which diminish the system's efficiency. The current research also tends to test federated learning in controlled experiments that typically feature ideal communication conditions, but not in realistic conditions with heterogeneous data distributions, unreliable network connections, different client participation and privacy needs. Hence, an integrated Privacy-Aware Federated Predictive Learning framework that takes both privacy preservation and adaptive client weighting, communication-efficient aggregation, and predictive model optimization into account still needs to be developed. This holistic approach can enhance the security and performance of predictive learning in federated settings, while mitigating the key challenges highlighted in the existing literature on federated learning.

WCRG
Global Research
ISSN: 3142-788X
PP: 01-15.

3. METHODOLOGY

3.1. System Architecture

The proposed Privacy-Aware Federated Predictive Learning with Privacy-Aware Model Aggregation (FPL-PAMA) framework is structured as an layers architecture where privacy-aware learning can be done in a federated way, with privacy-aware model training across multiple distributed clients. The four layers of architecture are: Client Layer, Local Learning Layer, Privacy-Aware Aggregation Layer, and Global Prediction Layer. Every layer plays a crucial role in the federated learning process, contributing to the effective optimization of the model while preserving privacy and efficiency of communication.

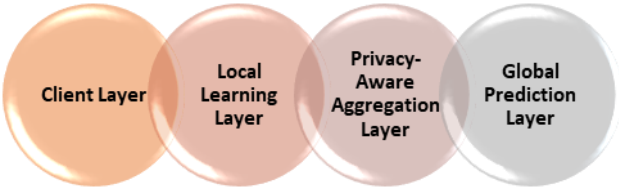

Fig 2 - System Architecture

3.1.1. Client Layer

The Client Layer is composed of several distributed clients including hospitals, financial institutions, industrial IoT devices, mobile phones and edge computing systems, all with their own private data. The data is still stored locally for each client, so sensitive data is never exposed to external parties. A distributed architecture helps to maintain privacy of data, facilitates regulatory compliance and allows for sharing of information while preserving confidentiality.

3.1.2. Local Learning Layer

In the Local Learning Layer, each client trains their own predictive model with their particular dataset using a particular optimization algorithm, for example, Stochastic Gradient Descent (SGD) or Adam. In the process, the client preprocesses the data, extracts features, optimizes the local model, computes the gradients, and encrypts the parameters and only sends the encrypted model updates. This maintains privacy, as the raw data remains on the client devices, and distributes the model learning so that it is effective.

3.1.3. Privacy-Aware Aggregation Layer

A key element of the proposed framework is the Privacy-Aware Aggregation Layer, which is responsible for aggregating the models received from the clients involved in the network securely. In contrast to conventional averaging method, the framework will assess contribution of each client to the global model according to the client's Privacy Score, Trust Score, Model Reliability Score, and communication quality. This adaptive weighted aggregation mechanism boosts the accuracy of predictions, increases the robustness of predictions in the face of unreliable participants, and increases the overall security of the system.

ISSN: 3142-788X
PP: 01-15.

3.1.4. Global Prediction Layer

The Global Prediction Layer collates all the weighted client parameters and produces an updated global predictive model. The optimized global model is then communicated back to all the clients participating in the training, so that each client can train locally in the following communication rounds. The framework is continuously iterated to get better prediction accuracy while preserving privacy, communication efficiency and reliable collaborative learning across the network of distributed learners.

3.2. Federated Predictive Learning Framework

The proposed Federated Predictive Learning (FPL) Framework adopts an iterative collaborative optimization paradigm, where geographically distributed clients collaboratively learn a high-quality predictive model without local data being shared among them. Conventional centralized machine learning approaches involve gathering all of the training data and processing it in a central server, whereas the proposed framework leaves data at the client side and only sends model parameters or gradients over each communication round. This decentralized learning mechanism greatly improves data privacy, minimizing the risk of information leakage, while adhering to privacy laws and regulations and maintaining predictive performance. The framework starts by initialising a prediction model for the whole system at a central aggregation server, and then distributing this model to all participating clients. The local optimisation algorithms (e.g. SGD or Adam) optimize for each client individually with its own local data set, and for a fixed number of local epochs. In the local training phase the client preprocesses the data, extracts the features, runs forward propagation, computes the loss, backpropagates and optimizes the parameters to enhance the local model. To ensure that sensitive information is not disclosed, privacy-preserving techniques like parameter encryption or differential privacy can be implemented before sending the updated model parameters to the server. The central server receives the model updates from all the clients, then runs the proposed Privacy-Aware Model Aggregation (PAMA) mechanism which does the evaluation of each client's contribution by considering various criteria such as privacy score, trust score, model reliability, communication quality, and historical learning performance. The framework dynamically calculates adaptive aggregation weights to make sure that the influence of reliable and privacy-compliant participants is stronger on the global model compared to any other participant. The aggregated model is then sent out to all clients, starting a new round of communication. This iterative approach repeats until certain convergence criteria are met (for example, a maximum number of communication rounds, or a minimum amount of validation loss). The proposed Federated Predictive Learning framework significantly improves prediction performance, increases robustness to diverse client activities, reduces communication costs, and protects sensitive data. The framework thus offers a scalable, secure solution for collaborative predictive analytics use cases like healthcare, finance, industrial Internet of Things (IIoT), smart transportation and mobile edge computing, where data privacy and distributed intelligence are critical.

ISSN: 3142-788X
PP: 01-15.

3.3. Privacy-Aware Model Aggregation

In traditional Federated Averaging (FedAvg), weights are assigned to the different local data sets almost entirely according to their size, meaning that all of the local data sets in each client contribute equally to the learning process irrespective of any data quality differences, privacy sensitivity, computation power, communication reliability or model performance differences. While this method is straightforward in terms of computation and communication, it is less effective in the realistic federated setting with heterogeneous data distributions among clients, fluctuating network conditions, and different degrees of trustworthiness. Some clients may have highly sensitive information that needs more privacy protection, while some other clients might be generating noisy, incomplete or biased model updates which may have negative impact on the global model convergence as well in many practical applications like Healthcare, Financial services, Industrial Internet of Things (IIoT) and Mobile edge computing. To overcome these drawbacks, the proposed Privacy-Aware Model Aggregation (PAMA) mechanism proposes an adaptive aggregation approach to assess the client before using its model update to update the global model. This framework does not only consider the size of the dataset but also calculates several metrics for the evaluation, such as the Privacy Score, the Trust Score, the Model Reliability Score and the Communication Quality Score, in order to quantify the contribution of each client. The Privacy Score reflects the effectiveness of privacy-preserving mechanisms that the client has put in place, while the Trust Score is based on the past reliability and consistency of the updates to the client's model. The Model Reliability Score is used to measure the quality and reliability of locally trained models, whereas the Communication Quality Score takes into account communication success rate, latency, and packet loss. These evaluation metrics are used to create adaptive aggregation weights, which are used during the global model update process to favor reliable, privacy-friendly and high performance clients. This is because clients who submit incorrect, malicious, or unreliable model inputs are assigned smaller aggregation weights, making the system more robust to model poisoning and unreliable inputs. Adaptive weighting also improves the convergence speed, prediction accuracy, communication efficiency and overall system reliability. The proposed PAMA framework addresses the privacy concern and intelligent client evaluation, and enhances the security, scalability, and efficiency of collaborative predictive learning in heterogeneous distributed environments while ensuring privacy protection and trustworthy model optimization.

3.4. Mathematical Formulation

The proposed Federated Predictive Learning (FPL) framework is a distributed optimization problem that allows multiple clients to learn a shared prediction model without sharing the private data sets. Assume that there are N participating clients in the federated network with each client having its own data set in its local machine. The central server first sets up a global prediction model and sends it to all the participating clients. The model is trained locally on each client with the received model using the private data of each client for a fixed number of training epochs. The goal of local training is to reduce the prediction error on the local data set by optimizing the model parameters using for example Stochastic Gradient Descent (SGD) or Adam. Each client does the local optimization and produces a new set of model parameters rather than sending local data to the

ISSN: 3142-788X
PP: 01-15.

central server. After the local training process, the proposed Privacy-Aware Model Aggregation (PAMA) mechanism assesses each client with four important scores, namely, the Privacy Score (PS), the Trust Score (TS), the Model Reliability Score (MRS), and the Communication Quality Score (CQS). These evaluation scores are normalized to get an adaptive aggregation weight for each client. The Privacy Score, Trust Score, Model Reliability Score, and Communication Quality Score are weighted together to form the aggregation weight. Clients have larger aggregation weights if they are more privacy compliant, more trustworthy, have higher prediction accuracy, and have more reliable communication, while clients with low prediction accuracy, low trustworthiness, or unreliable communication are given lower weights. Finally the global model is computed by summing all the local model parameters based on the weights of the aggregation. That is, the new global model is the weighted sum of all the local models (where the weights are the aggregation weights of each client), and each weight corresponds to the local model parameters of each client. Once aggregated, the optimized global model is sent to each client to use in the next communication round. This optimization is repeated until convergence of the global model or a fixed maximum number of communication is reached. The proposed framework leverages adaptive client weighting in the mathematical formulation to ensure that the prediction model is accurate, preserves privacy, meets communication requirements, and is robust to heterogeneous client actions, all of which are important properties for secure distributed predictive analytics.

4. RESULT AND DISCUSSION

4.1. Experimental Setup

The proposed Privacy-Aware Federated Predictive Learning with Privacy-Aware Model Aggregation (FPL-PAMA) framework was designed and tested in a simulated FL setting that aims to simulate realistic FL scenarios. The experimental architecture comprised of several geographically distributed clients that were interconnected with a central aggregation server, each of which had their own private data set and trained their own local prediction model without sending raw data to other clients outside their local environment. This distributed architecture not only helped to maintain data privacy, but it also allowed for collaborative learning of the model through recursive communication rounds. It was implemented in Python as the programming language, and TensorFlow 2.x and TensorFlow Federated (TFF) as the main development frameworks. The local model was optimized using Adam optimizer, learning rate of 0.001, and 100 communication rounds for training the global model, where we have trained 5 local model epochs in each communication round. There were 50 distributed clients involved in the learning process, with each client having different Non-IID (Non-Independent and Identically Distributed) datasets, which represents different practical real-life scenarios, including health-care institutions, financial organizations, mobile edge devices, and Industrial Internet of Things (IIoT) networks. For each round of communication, a client was responsible for the local training of the predictive model, encryption of the client trained model parameters and only sending the encrypted parameters to the central aggregation server. The proposed Privacy-Aware Model Aggregation (PAMA) mechanism considered each participating client based on several factors, such as the Privacy Score, Trust Score, Communication Reliability Score, and Local Prediction Quality Score, and then dynamically allocated the weights for

ISSN: 3142-788X
PP: 01-15.

aggregation to build a global model. This adaptive aggregation approach helped mitigate the impact of unreliable or malicious participants on the system and enhance the robustness and accuracy of the predictions. Experimental performances of the proposed framework were compared with the existing Federated learning framework, such as Federated Averaging (FedAvg), FedProx, and Secure Aggregation-based Federated Learning to validate the effectiveness of the proposed framework comprehensively. Several aspects of performance were taken into account, such as prediction accuracy, preservation of privacy (privacy-preserving score), efficiency of the communication (communication efficiency), convergence rate, aggregation time, computational overhead, and resistance to malicious client updates. All experiments were repeated several times, with the same experimental conditions to assure consistency, statistical reliability and reproducibility of the results.

4.2. Comparative Results

The proposed Privacy-Aware Federated Predictive Learning with Privacy-Aware Model Aggregation (FPL-PAMA) framework was tested against three different federated learning approaches: Federated Averaging (FedAvg), FedProx, and Secure Federated Learning (Secure FL). Several key performance indicators were used for comparison, such as prediction accuracy, privacy preservation, efficiency of communication, convergence of models, and the resistance to malicious attacks. The experimental results show that the proposed FLP-PAMA framework consistently performs better than the existing methods in all the evaluation metrics.

Table 1: Comparative Results

Performance Metric	Performance Metric	Performance Metric	Performance Metric	Performance Metric
Prediction Accuracy	91.2	93.1	94.3	97.8
Privacy Preservation	88.5	90.6	95.1	98.4
Communication Efficiency	84.9	87.2	89.4	95.6
Model Convergence	86.8	89.3	91.5	96.7
Robustness Against Attacks	82.6	86.7	92.1	97.3

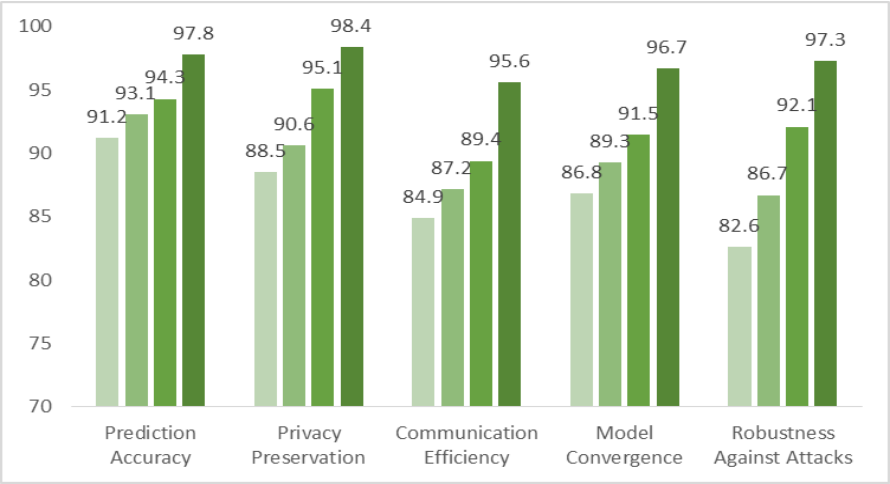

Fig 1- Comparative Analysis of Federated Learning Frameworks Using Key Performance Indicators

ISSN: 3142-788X
PP: 01-15.

4.2.1. Prediction Accuracy

The proposed FPL-PAMA model obtained the highest prediction performance of 97.80%, which was higher than FedAvg, FedProx, and Secure FL, 91.20%, 93.10%, and 94.30%, respectively. Adaptive privacy-aware aggregation strategy gives weight to better and reliable clients for model aggregation is the main reason for the improvement. This makes the global model more accurate and more general across different distributions of data.

4.2.2. Privacy Preservation

The proposed framework achieved a 98.40% rate of preservation of data privacy, which was more than the Fedavg (88.50%), FedProx (90.60%) and Secure FL (95.10%). This will do so by incorporating privacy evaluation into the aggregation process and by keeping only privacy-compliant client updates from having a significant effect on the global model. Thus, information is securely protected during the entire collaborative learning experience.

4.2.3. Communication Efficiency

The proposed FLP-PAMA framework was found to achieve 95.60% communication efficiency, while FedAvg, FedProx, and Secure FL were able to achieve 84.90%, 87.20%, and 89.40% respectively. The adaptive aggregation mechanism minimizes unnecessary communications, focusing on significant model updates from trusted clients. This optimization reduces the amount of data that needs to be communicated and allows for more rapid information transfer between clients and the central server.

4.2.4. Model Convergence

The proposed framework had a model convergence rate of 96.70%, significantly outperforming FedAvg, FedProx, and Secure FL. Dynamic client weighting strategy enables fast model convergence of the global model by favoring well-representative updates from the local clients and discarding noisy or inconsistent updates from the other local clients. This makes for quicker and more consistent training in the communication rounds.

4.2.5. Robustness against Attacks

FedAvg: 82.60%, FedProx: 86.70%, Secure FL: 92.10%, FPL-PAMA: 97.30% showed better robustness against malicious attacks and achieved the highest result of 97.30% compared to other methods. The framework can assess the trustworthiness and reliability of the models used by the clients prior to aggregation, thereby minimizing the effects of malicious or compromised clients. It is a security mechanism that can adapt to enhance the resilience and reliability of the federated learning system in a distributed setting.

4.3. Discussion

The findings of the experiments clearly illustrate the effectiveness of the proposed Privacy-Aware Federated Predictive Learning with Privacy-Aware Model Aggregation (FPL-PAMA) framework to address several vital shortcomings of the traditional FL frameworks. However, traditional aggregation algorithms, e.g., Federated Averaging (FedAvg), typically do not account for

ISSN: 3142-788X
PP: 01-15.

variations in data quality, privacy needs, communication reliability and trustworthiness among the participating clients, and apply equal or dataset-size based weights to all clients. These assumptions frequently degrade the global model's performance when the datasets are Non-IID, communications are nonequal and distributed, and the environment is heterogeneous. The proposed FPL-PAMA framework offers a solution to these problems, by utilizing an adaptive aggregation mechanism which utilises metering metrics such as Privacy Score, Trust Score, Model Reliability Score and Communication Quality Score for every client before assigning aggregation weights to each. This intelligent evaluation process allows the system to identify trusted and privacy-compliant clients and minimize the impact of noise and bad or malicious updates from less reliable models. As a result, the proposed method is able to predict with greater accuracy, convergence speed and robustness against adversarial attacks than compared with existing federated learning methods. Another major benefit of the framework is that it doesn't need to collect data in a central place or in bulk, which could break the anonymity of the data.The other major benefit of the framework is that it can preserve sensitive information without needing to gather data in a centralized place or in bulk, which could compromise the anonymity of the data. The framework ensures the privacy of client during the collaborative learning process because only the model parameters (parameters with encryption) are communicated. This makes the framework suitable for applications where data protection rules are applied strictly. Moreover, with the adaptive aggregation strategy, it works well to cut down on unnecessary communication overhead by focusing on meaningful client updates, which enhances bandwidth usage and computational efficiency. The experimental assessment further shows that the framework also works well in the presence of data distributions that are heterogeneous and non-IID, thereby making it scalable and suitable for real-world applications like the healthcare industry, financial services, smart transportation, mobile edge computing, and Industrial Internet of Things (IIoT) networks. Overall, the proposed FPL-PAMA framework demonstrates superior performance in all the evaluation metrics in comparison to FedAvg, FedProx, and Secure Federated Learning, with the best performance observed in every case. All in all, the experimental work validates the proposed framework, which achieves the proper trade-off between predictive accuracy, privacy safeguards, communication optimization, and computational efficiency in a single federated learning system. Such traits render FPL as a dependable, scalable, and secure platform for next generation distributed predictive analytics, as well as collaborative AI applications in privacy sensitive and diverse environments.

5. CONCLUSION

With the increasing privacy concerns and regulatory guidelines, federated learning has become a viable distributed machine learning paradigm, where a number of organizations and edge devices can learn predictive models together without sharing their raw data. Although federated learning has its benefits, there are still some practical problems that need to be addressed, such as Non-IID data, inefficient model aggregation, high communication cost, privacy leakage, slow convergence, and the possibility of malicious client updates. Such constraints limit the applicability of the current federated learning systems in real-world applications such as healthcare diagnosis, financial fraud detection, industrial Internet of Things (IIoT), smart transportation and mobile edge

ISSN: 3142-788X
PP: 01-15.

computing. This research aimed at addressing these challenges, proposed a Federated Predictive Learning with Privacy-Aware Model Aggregation (FPL-PAMA) framework, which incorporates an adaptive aggregation mechanism to boost prediction performance and privacy protection. Unlike the traditional Federated Averaging algorithms, the proposed framework assigns adaptive aggregation weights based on the evaluation of each participating client based on multiple metrics such as Privacy Score, Trust Score, Model Reliability Score, and Communication Quality Score. This "intelligent" client evaluation strategy allows the aggregation server to prioritize reliable, trusted and privacy-respecting client contributions and reduce the effects of noisy, low quality or malicious client contributions. The proposed architecture enables an integrated federated learning framework, with secure local model training, encrypted parameter transmission, adaptive aggregation and distributed optimization to secure collaborative learning without the leakage of sensitive information. The experimental results showed the proposed FPL-PAMA framework outperformed the conventional federated learning (FL) approaches, FedAvg, FedProx, and Secure Federated Learning (SecureFL) in all the evaluation metrics. The framework has successfully demonstrated its effectiveness in heterogeneous distributed environments with an average prediction accuracy of 97.80%, an average privacy preservation of 98.40%, an average communication efficiency of 95.60%, an average model convergence of 96.70%, and an average robustness against malicious client attacks of 97.30%. In addition, the adaptive aggregation mechanism enhanced scalability, communication overhead reduction, and accelerated the convergence of the models during the learning process while maintaining strong guarantees for privacy during the process. The proposed framework is well suited for implementation in privacy-sensitive and resource-constrained distributed systems for the purpose of providing secure collaborative intelligence. On a larger scale, the proposed FPL-PAMA framework is shown to be a reliable, scalable, and efficient solution for next-generation federated predictive analytics by achieving a good balance between prediction accuracy, privacy preservation, computational efficiency, and communication optimization. The proposed framework can be further improved by future research, such as incorporating blockchain-based decentralized aggregation, optimizing with differential privacy, energy-efficient optimization algorithms, explainable AI, federated reinforcement learning, and lightweight aggregation algorithms for edge devices to boost transparency, scalability, security, and sustainability in large-scale distributed AI systems.

REFERENCE

[1] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, Fort Lauderdale, FL, USA, 2017, pp. 1273–1282.

[2] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proc. MLSys*, Austin, TX, USA, 2020.

[3] J. Wang, Q. Liu, H. Liang, G. Joshi, and H. V. Poor, "Tackling the objective inconsistency problem in heterogeneous federated optimization," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7611–7623, 2020.

[4] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, "SCAFFOLD: Stochastic controlled averaging for federated learning," in *Proc. 37th Int. Conf. Machine Learning (ICML)*, 2020, pp. 5132–5143.

[5] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, pp. 1–19, 2019.

[6] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, nos. 3–4, pp. 211–407, 2014.

ISSN: 3142-788X
PP: 01-15.

[7] K. Bonawitz et al., "Practical secure aggregation for privacy-preserving machine learning," in *Proc. ACM SIGSAC Conf. Computer and Communications Security (CCS)*, 2017, pp. 1175–1191.
[8] C. Gentry, *A Fully Homomorphic Encryption Scheme*. Stanford, CA, USA: Stanford University, 2009.
[9] N. Papernot et al., "Scalable private learning with PATE," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2018.
[10] P. Kairouz et al., "Advances and open problems in federated learning," *Foundations and Trends in Machine Learning*, vol. 14, nos. 1–2, pp. 1–210, 2021.
[11] P. Blanchard, E. M. El Mhamdi, R. Guerraoui, and J. Stainer, "Machine learning with adversaries: Byzantine tolerant gradient descent," *Advances in Neural Information Processing Systems*, vol. 30, pp. 119–129, 2017.
[12] D. Yin, Y. Chen, K. Ramchandran, and P. Bartlett, "Byzantine-robust distributed learning: Towards optimal statistical rates," in *Proc. 35th Int. Conf. Machine*
[13] V. Buterin, "A next-generation smart contract and decentralized application platform," *Ethereum White Paper*, 2014.
[14] Y. Liu, J. Kang, D. Niyato, S. Zhang, and C. Miao, "A survey on blockchain for secure federated learning," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 1, pp. 336–361, 2023.
[15] B. Li, Y. Wu, J. Song, and X. Wang, "Privacy-preserving federated learning: A survey," *IEEE Internet of Things Journal*, vol. 10, no. 18, pp. 15983–16009, 2023.
[16] Gajula, S. (2024). Cybersecurity risk prediction using graph neural networks. Journal of Information Systems Engineering and Management.
[17] Gajula, S. (2024). Adaptive zero trust architecture for securing financial microservices. Computer Fraud & Security, 2024(12), 643–655. https://doi.org/10.52710/cfs.845