Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 46 No. 3 (2025)

Taxonomy and Survey of Federated Learning Approaches for Privacy-Preserving Applications in IoT, Healthcare, and Smart Environments

Narendra V S[1], Dr. Savita K Shetty[2]

1 & 2, Department of Information Science and Engineering, Ramaiah Institute of Technology, Bengaluru-54

Corresponding Author: Narendra V S

Abstract - A quickly developing machine learning paradigm called federated learning (FL) allows for cooperative model training while protecting data privacy. Massive amounts of sensitive data are produced by digital systems in healthcare, the Internet of Things, and smart environments. FL has become significant because of its decentralized methodology, which forbids the sharing of raw data. The applications, difficulties, and developments in FL are examined in this review article, with an emphasis on privacy-preserving techniques in a variety of fields. Classifying and evaluating FL frameworks according to their architecture, learning models, aggregation methods, and privacy-preserving tactics is the main goal of this evaluation. The chosen studies include cybersecurity, smart city infrastructure, healthcare diagnostics, and customized IoT solutions. Relevance to FL privacy preservation and practical use was guaranteed by the inclusion criteria. The results show that although FL successfully improves data security, issues with data heterogeneity, model convergence, communication cost, and adversarial attack susceptibility still exist. To improve privacy guarantees, methods like clustered FL, secure multi-party computing, homomorphic encryption, and differential privacy are frequently used. In order to increase FL systems' scalability, resilience, and security, this paper identifies these developments and suggests future research avenues. This study contributes to the developing subject of privacy-preserving AI by offering a systematic taxonomy and analysis of current federated learning algorithms suited to privacy-sensitive situations.

KeyWords : Federated Learning, Privacy Preservation, Internet of Things (IoT), Healthcare AI, Cybersecurity, Smart Environments, Differential Privacy, Homomorphic Encryption, Secure Aggregation, Non-IID Data, Machine Learning, Deep Learning, Edge Computing, Secure Multi-Party Computation, Decentralized AI

1. INTRODUCTION

1.1 Overview of Privacy Challenges in Data-Driven AI

Unprecedented amounts of data collection, storage, and use have resulted from the digital growth of businesses, which is being propelled by the quick development of artificial intelligence (AI) and machine learning (ML). In order to provide insights, streamline choices, enhance user experiences, industries including healthcare, smart cities, finance, autonomous systems, and the Internet of Things (IoT) increasingly significantly rely on large-scale data-driven AI models. The development of real-time data streams, including personal health records, has been further expedited by the increasing number of smart devices, sensors, and mobile applications linking monetary exchanges to biometric, behavioral, and geographical information [1][2].

However, there are serious privacy issues with this increase in data availability. Large datasets are compiled and kept on cloud servers or centralized data centers as part of the centralized training paradigm used by traditional AI models. This combination poses a number of dangers, such as:

- Data breaches: Malevolent attackers find centralized storage to be a valuable target. Millions of private documents might be made public in a single breach.
- Unauthorized Access: Data abuse may result from insider threats or inadequate authentication procedures.
- Data Misuse and Secondary Use: Information gathered for one reason could be used for another without user authorization, which is against privacy laws and standards.

Furthermore, a number of extremely sensitive personal data types are gathered in the fields of healthcare (genomic data, medical history), finance (transaction records), and smart environments (home monitoring systems). Identity theft, financial fraud, discrimination, and physical injury can result from the improper use or disclosure of such information.

Centralized data processing is made more difficult by regulatory frameworks like the bill on personal data protection (PDP) in nations like India, the Health Insurance Portability and Accountability Act (HIPAA) in the United States, and the GDPR, or the General Data Protection Regulation, in Europe. Traditional data-driven AI models that rely on access to enormous, and centralized datasets are put to the test by these restrictions, which enforce the rules of data reduction, purpose limitation, and user permission.Furthermore, even in situations where raw data is not easily obtainable, attackers can recreate training data from the model's outputs or gradients owing to the growing danger of model inverted attacks. Serious privacy violations can also result from membership inference acts of violence, which can ascertain if a particular data piece was included in the model's training dataset [3][4].
Privacy-preserving machine learning (PPML) techniques are becoming essential due to these threats. In addition to accuracy and performance, systems have to have built with privacy requirements, data loss prevention, and adversarial assault resistance in mind.

In this regard, Federated Learning (FL) shows promise as a solution to these problems. FL enables a paradigm shift toward decentralized AI models that naturally emphasize data privacy by facilitating model training across dispersed data sources without sending raw data.

1.2 Role of Federated Learning (FL) in Privacy Preservation

A paradigm change in machine learning, federated learning (FL) was created to address the escalating worries about data security, privacy, and compliance in data-driven AI systems. With FL, which was first presented by Google in 2016, several clients, such edge devices, smartphones, medical facilities, or Internet of Things sensors, may work together to train a common global model while maintaining the locality of their raw data [5]. FL guarantees that data stays decentralized throughout the learning process, in contrast to standard centralized learning, which requires data to be uploaded to a central server for model training.Each participating client trains a local model on its own private dataset in a typical FL process, communicating only model changes (such as weights or gradients) to the central aggregation server. The global model is then updated by aggregating these local alterations, sometimes with the use of techniques such as Federated Averaging (FedAvg) [6]. By design, this minimizes exposure risks and complies with privacy standards like GDPR, HIPAA, and other information protection legislation by preventing the central server from getting hold of sensitive data.FL is especially well-suited for diverse and privacy-sensitive areas like healthcare, where ethical, legal, and regulatory restrictions prevent the centralization of patient data. Without jeopardizing patient privacy, hospitals, clinics, or research institutes can work in combination to jointly develop strong AI models for medical image analysis or illness prediction [7]. Similar to this, FL allows dispersed sensors and devices to take part in model training in IoT ecosystems and smart city infrastructure without disclosing potentially private user data to other parties (such as location or activity patterns).

In addition to its privacy advantages, FL promotes system resilience and scalability. As a reflection of the diversity of data found in daily life across devices, users, and surroundings, it enables learning from Non-Independent and Identically Distributed (Non-IID) data sources [8]. This increases FL's adaptability for implementation in a variety of contexts, from smart grids and industrial systems to wearables and smartphones.

Nevertheless, FL is not impervious to privacy threats in spite of these benefits. In order to breach the global model or get private information, attackers may use membership inference attacks, model inversion attacks, or malicious update injections, even while the raw data is still local. In order to increase the durability of FL systems, further privacy-preserving methods including homomorphic encryption, differential privacy, secure multiparty computing, and blockchain-based aggregation are frequently incorporated.In conclusion, by removing the requirement for centralized data gathering and facilitating collaborative model creation, Federated Learning tackles important privacy issues.FL is positioned as a crucial enabler for safe, scalable, and moral AI applications across privacy-sensitive areas thanks to its integration with privacy-enhancing technologies.

1.3 Motivation Behind This Review

Despite Federated Learning's potential across domains, a number of issues still need to be addressed, including as communication overhead, system scalability, data heterogeneity, and adversarial attack susceptibility. A thorough

review is required due to the increasing amount of research examining various FL topologies, aggregation algorithms, and privacy-enhancing strategies. Consolidating existing information, offering a structured taxonomy, and identifying research needs and opportunities in privacy-preserving FL frameworks are the objectives of this assessment. Given FL's quick acceptance in vital fields including cybersecurity, smart cities, healthcare, and the Internet of Things, it is crucial for both academics and business to comprehend its advantages and disadvantages.

1.4 Research Methodology

A thorough literature study of current research publications released between 2022 and 2024 is used in this review. Based on their applicability to privacy-preserving Federated Learning frameworks, innovative FL architectures, and their uses in cybersecurity, smart environments, IoT, and healthcare, a total of 17 peer-reviewed publications were chosen.
The methodology followed these steps:

- Literature Search: Making use of scholarly resources including Elsevier, Springer, IEEE Xplore, and the ACM Digital Library.
- Papers on FL that included real-world datasets, realistic implementations, and explicit privacy-preserving techniques met the inclusion criteria.
- Exclusion Criteria: Only theoretical works without experimental evaluation or privacy concerns.
- Classification: The chosen articles were arranged according to the system architecture, privacy strategies (such as Secure Aggregation and Differential Privacy), and application domain.Metrics like accuracy, scalability, communication cost, and attack resilience are used to compare the models in the review.

1.5 Related Reviews, Differences, and Our Contribution

1.5.1 Related Reviews in Federated Learning

The potential of Federated Learning (FL) to strike a compromise between collaborative model training and privacy preservation has garnered a lot of interest from the academic community. FL architectures, system designs, and general applications have been thoroughly examined in a number of studies. For example, scalability issues and algorithmic improvements are examined in studies on FL for edge computing and mobile contexts [1]. Other surveys concentrate on the implementation of FL in IoT or healthcare AI ecosystems, highlighting real-world applications and system-level difficulties [2][3].A number of scholars have also examined the security flaws of FL systems, describing dangers like poisoning, membership inference, and model inversion attacks [4]. As countermeasures, these studies usually suggest incorporating secure multiparty computation (SMPC), homomorphic encryption (HE), and differential privacy (DP) into FL systems [5][6]. Nevertheless, a large number of these assessments are either out-of-date or domain-specific, and therefore do not take into account the most recent developments in privacy-preserving methods inside FL frameworks.

1.5.2 Differences from Existing Reviews

Although the current surveys offer insightful information, there are still a number of important holes that need to be filled, which this study aims to fill:

Limited Application Scope: The majority of previous studies focus on specific industries, such as IoT settings [9], smart cities [8], or healthcare [7]. This study, on the other hand, offers a cross-domain analysis by methodically contrasting privacy-preserving FL frameworks used in smart environments [13], cybersecurity [12], IoT [11], and healthcare [10].Restricted Discussion of Current Works: Previous surveys have not fully examined recent ground-breaking research from 2022–2024 on multi-party safe aggregation [16], blockchain-based FL [15], and clustered FL [14]. These recent developments are specifically incorporated within this assessment Absence of Comparative Study of Privacy Methods .There aren't many surveys that provide a thorough side-by-side comparison of privacy-preserving methods like DP [6], HE [11], and SMPC [16] in FL systems. By connecting each research to its privacy enhancing strategy, our evaluation fills this gap.Lack of a Complete Taxonomy: Although generic taxonomies for FL architectures are available, little research has been done on privacy-centric taxonomies that categorize FL techniques according to threat models, system architecture, dataset kinds, and aggregation procedures. We provide this new privacy-oriented taxonomy in our evaluation.

1.5.3 Our Contribution

This work makes the following original contributions by providing a thorough analysis of 17 current peer-reviewed research studies on privacy-preserving FL systems:

Analysis Across Domains:

The review covers a wide range of topics, such as cybersecurity (intrusion detection [14]), IoT (smart parking [13]), healthcare (COVID-19 detection [10], skin disease categorization [11]), and smart settings (activity monitoring [15], adaptive learning [16]).

Taxonomy of Privacy-Preserving FL Techniques:
We classify FL frameworks based on:

- System architectures: Centralized, decentralized, hierarchical [1][2]
- Aggregation strategies: FedAvg, FedProx, Secure Aggregation [3][4]
- Privacy-preserving techniques: DP [5], HE [6], Blockchain [7], SMPC [8]

Identification of Research Gaps and difficulties: The study identifies many important research gaps and difficulties, including scalability of FL frameworks [14], communication efficiency [13], adversarial robustness [12], and managing extremely Non-IID data [11]. To solve these problems, further research avenues are suggested.For academics and practitioners looking to create federated learning systems that are reliable, scalable, and privacy-preserving for a variety of real-world applications, this survey offers a timely, thorough, and organized reference.

1.6 Research Questions

The review is guided by the following key research questions:

- Research Question 1: What are the most commonly used FL algorithms and architectures in privacy-sensitive applications?
- Research Question 2: Which datasets and domains are predominantly used to evaluate privacy-preserving FL models?
- Research Question 3: What privacy-enhancing technologies (PETs) are integrated with FL to mitigate security threats?
- Research Question 4: What are the key challenges, limitations, and open research problems in current FL systems?

1.7 Contributions of This Review

Finding Research Gaps and Challenges: The study finds several significant research gaps and challenges, such as managing highly Non-IID data [11], communication efficiency [13], adversarial robustness [12], and the scalability of FL frameworks [14]. Additional research directions are proposed to address these issues.
This study provides a recent, comprehensive, and well-structured reference for researchers and practitioners seeking to develop federated learning systems that are dependable, scalable, and privacy-preserving for a range of real-world applications.

Challenges and Future Directions: The paper identifies unresolved research issues include managing highly Non-IID data, cutting down on communication overhead, protecting FL against hostile assaults, and attaining scalability in sizable heterogeneous networks. We suggest further lines of inquiry to create FL systems that are reliable, effective, and privacy-preserving.

1.8 Structure of The Paper

The present investigation is divided into seven distinct segments, each with its own specific focus. The initial portion is dedicated to the Introduction, which provides an in-depth explanation of the motivation and background of the research. This section also presents the research questions, methodology, and the unique contributions of this review.

The following section, titled Preliminaries, delves into the essential concepts and terminology related to Federated Learning (FL). It introduces different FL architectures, privacy-preserving techniques, and presents a taxonomy framework that guides the classification of existing approaches.

The third section, Applications of Privacy-Preserving Federated Learning, illustrates the implementation of FL in diverse domains such as healthcare, IoT, smart environments, and cybersecurity. This section provides a comprehensive analysis of 17 research papers, focusing on key parameters: the algorithms used to train models, the datasets applied, aggregation strategies, privacy-enhancing mechanisms, and system architectures. It also examines existing security and privacy concerns associated with FL in these real-world applications.

The fourth section, titled Taxonomy of Federated Learning Frameworks, provides a structured classification of FL models based on their system architecture, aggregation methods, privacy-preserving techniques, and datasets. This section helps readers understand the strengths and limitations of various approaches.

The fifth section, Security Threats and Countermeasures, investigates the significant challenges related to data leakage, model inversion, poisoning attacks, and other adversarial threats in FL. It also reviews the proposed defensive techniques designed to enhance privacy and robustness in FL systems.

The sixth section, Challenges and Future Research Directions, explores the potential benefits, difficulties, and future prospects of privacy-preserving FL. It identifies open research problems and emerging trends that require further investigation.

Finally, the review concludes with a summary of key findings and reflections on the role of privacy-preserving Federated Learning as a crucial enabler for secure, scalable, and ethical AI systems in sensitive data environments.

2. Preliminaries

Federated Learning (FL), which enables distant devices or organizations to jointly train machine learning models while maintaining local data, has become an important privacy-preserving machine learning paradigm. Understanding the fundamental ideas, structures, and privacy-preserving strategies related to FL is crucial before exploring applications and difficulties.

2.1 Federated Learning Fundamentals

Multiple clients (such as mobile devices, hospitals, and IoT sensors) may calculate model modifications locally on their own data according to FL's fundamental concept of decentralizing model training. Clients provide just model parameters or elevations to a central server for aggregation, rather than raw data [1]. In addition to complying with privacy laws like GDPR and HIPAA, this greatly lowers the chance of data leaking [2].

The general FL process consists of the following steps:

- The server initializes a global model.
- After downloading the global model, each client uses private data to train it locally.
- Updated model parameters are sent to the server by clients.
- To update the global model, the server carries out aggregation (e.g., FedAvg) [3].

Until convergence is reached, this process is repeated several times. By design, FL facilitates cooperative learning on dispersed datasets and lessens the need to send sensitive data.

2.2 Types of Federated Learning Architectures

A number of architectural variants of federated learning (FL) are intended to help cooperative model training in various data distribution circumstances. In order to solve the privacy, efficiency, and scalability issues that arise in real-life applications, the architectural selection is essential. Below is a description of the main categories of FL architectures:

2.2.1 Horizontal Federated Learning (HFL)

The most popular type of FL is horizontal federate learning, which is sometimes referred to as sample-based FL. Devices, institutions, or organizations that participate in HFL have databases with the same feature space but distinct user samples. When several entities carry out the same operation while gathering data from various user bases, this design is perfect.

In instance, many hospitals may work together to use patient data to train a disease detection algorithm. Although the patient samples fluctuate, each hospital records comparable characteristics, such as age, blood pressure, and cholesterol levels. These hospitals may create a strong worldwide model by using HFL without releasing private patient data or breaking data protection laws [4].

Applications where the feature space is constant among clients, such as wearable device cooperation, smart health monitoring, and mobile keyboard predictions, make extensive use of HFL.

2.2.2 Vertical Federated Learning (VFL)

Situations when clients have distinct feature sets but the same user base are addressed by vertical federated learning. When two organizations collaborate and each side has complementary knowledge on the same entities, VFL is very helpful.
For example, a bank could save information on financial transactions, whereas an e-commerce platform might keep track of the same users' past purchases. Without disclosing their private datasets to one another, these two firms can work together to build a fraud detection model [5].
In order to maintain privacy during feature-level integration, VFL sometimes calls for secure multiparty computation (SMPC) or homomorphic encryption (HE), which adds complexity by needing entity alignment and encrypted computations.

2.2.3 Federated Transfer Learning (FTL)

Federated Transfer Learning is intended for use in scenarios when participant characteristics and sample data overlap is little or nonexistent. FTL is appropriate for cases where direct FL is impractical because of substantial data disparities since it uses transfer learning techniques to facilitate knowledge exchange between heterogeneous datasets. For instance, a healthcare facility that tracks patient vitals may work with an IoT company that gathers device usage patterns. FTL approaches aid in the transfer of pertinent knowledge between different domains to enhance model performance, despite the enormous differences in their datasets [6].In cross-domain applications, FTL is becoming more and more popular, particularly where data heterogeneity or unavailability is a significant obstacle.

2.2.4 Clustered and Hierarchical Federated Learning

Clustered Federated Learning (CFL):

CFL is a more sophisticated variant in which clients are categorized into groups according to application domain, behavior patterns, or data similarities. A local model is created by each cluster and subsequently included into the global model. This method improves model accuracy, especially in cases involving Non-IID (Non-Independent and Identically Distributed) data, which are prevalent in smart environments, healthcare, and the Internet of Things [7]. Before being aggregated into the global model, individuals with comparable problems may be pooled in mental health monitoring, for example, to create a specialized sub-model [8].

Hierarchical Federated Learning (HFL):

A multi-level aggregation structure is introduced by hierarchical FL, in which local updates are delivered to a central cloud server after first being aggregated at the edge level (such as edge servers or regional nodes). For large-scale FL installations, this design enhances scalability and lowers communication overhead [9]. With hundreds of edge devices operating concurrently, hierarchical FL is especially useful in energy grids, smart cities, and industrial IoT systems.

2.2.5 Split Federated Learning (SplitFL)

Due to its ability to lessen the computational load on client devices, Split Federated Learning (SplitFL) is perfect for contexts with limited resources, such as mobile devices or Internet of Things sensors. The model in SplitFL is separated into two parts:

- Data is processed by the client-side model up to a certain "cut layer."
- The server then completes the forward and backward pass after receiving the intermediate outputs, or activations [10].

since of this design, FL is more practical in contexts like as low-power IoT systems or smart healthcare devices since it minimizes client-side processing and data transmission. Because only intermediate activations—not raw

data—are exchanged, SplitFL also creates chances for privacy improvements by lowering the possibility of information leakage.These FL designs balance trade-offs between communication cost, computational efficiency, and privacy protection, providing flexibility to serve a range of collaborative situations. The data distribution, application domain, and privacy needs of the collaborating organizations all influence the choice of FL type.

2.3 Privacy-Preserving Techniques in Federated Learning

Federated Learning (FL) is susceptible to inference assaults, gradient leakage, and model inversion risks that might jeopardize client data privacy, despite the fact that FL naturally lessens the need to share raw data. A number of privacy-preserving strategies are included into FL workflows to improve the privacy guarantees of FL systems. These methods preserve model value while bolstering protections against hostile attacks.

2.3.1 Differential Privacy (DP)

One of the most popular privacy-preserving techniques in FL is Differential Privacy (DP). By mathematically ensuring that the addition or deletion of a single data point from the dataset has no discernible impact on the learning model's output, DP makes it challenging for attackers to draw conclusions about specific participants [11]. To accomplish DP in practice, calibrated noise is added to either the aggregated global model or the local model updates prior to sharing. A privacy budget parameter ε (epsilon) carefully regulates the noise, balancing the trade-off between privacy protection and model accuracy.

In FL, where data sensitivity is quite high, DP is particularly helpful for IoT and healthcare applications. For example, DP has been effectively implemented into COVID-19 detection models [12] and IoT-based activity monitoring systems [13] to prevent leaking of sensor data and patient records, respectively, during model aggregation.

Nevertheless, using DP has drawbacks, including:

- Decreased model accuracy as a result of more noise
- More client-side computational cost;
- Difficulty in adjusting the privacy budget

Notwithstanding these difficulties, DP is still a fundamental method for federated systems that protect privacy.

2.3.2 Homomorphic Encryption (HE)

Secure processing is made possible by homomorphic encryption (HE), which enables calculations to be carried out directly on encrypted data without requiring decryption at any stage of the computation [14]. Because of this feature, HE is especially useful in FL, where client model changes may be collected and encrypted without revealing gradients or raw data.
By preventing the aggregation server from ever learning individual model updates, HE considerably lowers the possibility of reconstruction or gradient leaking attacks. Smart grid systems, financial institutions, and healthcare infrastructures—where data privacy and regulatory compliance are crucial—benefit from HE-based aggregation.For example, HE is used to safely sum encrypted gradients or weights from many clients before to applying model updates in smart healthcare systems or power theft detection applications [14].

HE's primary drawback, meanwhile, is its substantial computational expense, particularly when working with intricate deep learning models. Performance constraints are frequently addressed by hybrid techniques or efficient implementations.

2.3.3 Secure Multi-Party Computation (SMPC)

A cryptographic approach called Secure Multi-Party Computation (SMPC) allows several clients to collaboratively compute a function over their inputs while maintaining the privacy of those inputs. Without the need for a reliable server, SMPC is used in FL to carry out safe model aggregation [15]. To guarantee that no one party has access to the entire dataset or model changes, each client shares encrypted data fragments (secret shares) with other clients. The complete aggregated model can only be recreated when all participants compute together.

SMPC works very well in the following situations:

- Vertical Federated Learning (VFL) scenarios with several feature owners
- Partnerships between entities with stringent privacy regulations

For instance, SMPC in FL is used by financial organizations and healthcare systems to safely calculate global models without exchanging private information [15]. For large-scale systems, SMPC necessitates careful optimization since it may be computationally and communication-intensive.

2.3.4 Blockchain and Distributed Ledger Technologies (DLT)

To improve security, traceability, and credibility in dispersed situations, FL frameworks are progressively using blockchain and distributed ledger technologies (DLT) [16]. Every transaction or model update in the FL process is recorded in an unchangeable, tamper-resistant ledger provided by blockchain.

The following are advantages of combining FL with Blockchain/DLT:

- Transparency-ensuring auditable model training records
- Elimination of single-point-of-failure threats from central servers
- Tamper-proof aggregation results

Blockchain guarantees that participating clients adhere to the protocol and that model changes are not altered during transmission in IoT networks or urban parking systems [16]. Smart contracts provide the ability to identify harmful activity, reward honest participation, and automate aggregation procedures.

Blockchain increases confidence, but it also adds storage overheads and latency, particularly in big networks. Research is being done on hybrid models that integrate FL with lightweight ledgers to strike a compromise between efficiency and security.

Table 1 . Summary of Privacy-Preserving Techniques

Technique	Purpose	Benefits	Challenges
Differential Privacy (DP)	Adds noise to prevent data leakage	Strong mathematical privacy guarantee	Accuracy loss, tuning ε
Homomorphic Encryption (HE)	Compute on encrypted data	Protects model updates during aggregation	High computational overhead
Secure Multi-Party Computation (SMPC)	Secure joint computation	No trusted aggregator required	Communication cost, scalability
Blockchain / DLT	Tamper-proof record of FL processes	Trustless, auditable system	Latency, energy consumption

2.4 Communication Efficiency and Optimization

The communication overhead brought on by clients and the central server exchanging huge model parameters repeatedly over several training rounds is one of the biggest problems with Federated Learning (FL). Frequent communication becomes expensive in terms of bandwidth, energy consumption, and latency in places with limited resources, such as edge computer nodes, mobile phones, and Internet of Things devices. For FL systems to be practical and scalable, communication efficiency must be guaranteed. To reduce communication costs without materially sacrificing model accuracy or performance, a number of optimization strategies have been put forth and implemented. Among the primary tactics are:

2.4.1 Model Compression and Quantization

The size of model updates sent by clients is decreased using model compression techniques. Typical techniques consist of:

- Gradient sparsification: Transmitting only the most significant gradients.
- Weight pruning: Removing less important model parameters.
- Quantization: Representing model weights and updates with fewer bits (e.g., 8-bit or binary representations).

By significantly lowering the volume of data transferred in each communication cycle, these techniques make FL more practical for low-bandwidth settings such as mobile devices and IoT sensor networks [17]. Excessive compression, however, might cause problems with convergence and accuracy loss.

2.4.2 Client Sampling

To cut down on communication frequency and expense, random or selective client sampling is utilized rather than mandating that every customer attend every training session. In each loop, the system chooses a subset of the accessible clients to accomplish:

- Lower communication overhead
- Reduced client-side computation load
- Improved scalability to large client populations

In order to prevent biased data selection from impairing model performance, careful sampling techniques make sure that statistical heterogeneity is taken into account.

2.4.3 Adaptive Aggregation Techniques

Advanced aggregation algorithms have been developed to improve communication efficiency while addressing data heterogeneity challenges. These include:

- FedAvg: A simple weighted average of local models, widely used in FL.
- FedProx: An extension of FedAvg that adds regularization to handle system and data heterogeneity.
- FedNova: Normalizes local updates based on local steps, improving performance in non-IID settings.

Adaptive aggregation reduces the number of communication rounds required for convergence and ensures fairness in federated systems where clients differ in computing power and data volume.

Importance in the Resource-Constrained Environments

Communication optimization techniques are especially vital for deploying FL in:

- IoT networks with limited connectivity
- Battery-operated mobile devices
- Edge computing environments

These strategies help minimize energy consumption, prolong device life, and enable real-world deployment of privacy-preserving FL frameworks.

3. Applications of Privacy-Preserving Federated Learning

In many fields where data security, privacy, and conformity to regulations are crucial, federated learning, or FL, has become a game-changing framework for collaborative machine learning. Because of its modular architecture, which allows model training without transferring sensitive data, it is ideal for privacy-sensitive industries including cybersecurity, IoT systems, healthcare, and smart environments. This section examines 17 recent studies that illustrate the results, difficulties, and application scenarios of privacy-preserving FL frameworks in different fields.

3.1 Healthcare Applications

Because medical data is extremely sensitive and subject to strict data protection laws, the healthcare industry continues to lead the way in FL adoption. Multiple healthcare facilities and medical equipment can work together to train models using privacy-preserving FL approaches without jeopardizing patient confidentiality. For example, [1] proposes a privacy-preserving FL system for the Internet of Medical Things (IoMT), emphasizing safe aggregation and authentication methods. This system protects patient data from potential adversarial assaults and unlawful access while allowing remote healthcare equipment to take part in model training.

FL was useful in speeding up diagnosis while preserving anonymity during the COVID-19 epidemic. Researchers created a COVID-19 detection model in [7] that combines FL and Differential Privacy (DP). Using networked chest imaging datasets, the architecture effectively increased diagnosis accuracy while maintaining the security

of private patient data.
The categorization of skin diseases has also benefitted from FL frameworks that use DP and Convolutional Neural Networks (CNNs). This method addresses privacy and security concerns by enabling training on sensitive skin lesion photos from many sources without data centralization, as explained in [9].

FL protects privacy while enabling wearable devices to sense stress in the context of mental health monitoring. In order to protect sensitive physiological data acquired from users from being exposed, Study [12] uses a hierarchical FL design in conjunction with local DP.
For healthcare applications, strong FL frameworks that incorporate Secure Multi-Party Computation (SMPC) have also been suggested. These techniques, as described in [14], improve model resistance to poisoning assaults while preserving anonymity throughout the model aggregation stage, which makes them ideal for delicate healthcare settings.

3.2 IoT and Smart Environments

Data privacy and resource limitations in distributed systems are two issues that are addressed by integrating FL into IoT networks and smart environments. Large volumes of decentralized data produced by devices with constrained processing power are frequently involved in these applications. For instance, the Edge-Intelligence FL framework for smart healthcare systems, which was unveiled in [5], uses lightweight encryption methods to protect private medical information while lessening the strain on IoT devices' communication capabilities. In settings with limited resources, this method improves efficiency and protects privacy.

Additionally, customized FL models have been investigated to enhance privacy and flexibility in IoT environments. In [8], researchers created a FL framework with user feedback loops that minimizes data transfer and enables IoT devices to improve their local models according to user preferences. Without compromising model performance, this tailored learning guarantees the protection of individual privacy. FL has also been used in the educational sector to safeguard the privacy of students. FL is used to track and categorize e-learning actions without sending sensitive user data in [13], an on-screen activity tracking system. The approach demonstrates FL's adaptability in a variety of contexts by guaranteeing that individual learning styles stay confidential.

Smart parking systems and other urban infrastructures have also made use of FL in conjunction with blockchain and distributed ledger technologies (DLT). A blockchain-enabled FL framework for predicting urban parking slots is shown in [15], guaranteeing auditable and impenetrable model changes while protecting user location information. Similar to this, clustered FL techniques have been used in wellness detection and tailored healthcare applications [17], where clients are grouped according to similarity to increase model accuracy and protect anonymity.

3.3 Cybersecurity and Intrusion Detection Systems (IDS)

FL frameworks provide notable benefits in the fields of anomaly detection and cybersecurity by facilitating collaborative learning without direct access to private datasets or sensitive system logs. For example, in a FL framework for IoT resource allocation, study [2] combines local DP with Reinforcement Learning (RL). In IoT situations, this method minimizes the chance of critical data loss while optimizing resource utilization.FL has been used in critical infrastructure to identify power theft in smart grids. By utilizing model compression approaches, the system suggested in [6] guarantees safe and effective data exchange, enabling utility providers to work together to identify fraudulent activity without jeopardizing consumer privacy.

Furthermore, in [10], a collaborative FL model for intrusion detection systems (IDS) on the cloud edge is introduced. While protecting the privacy of individual system logs and avoiding data loss during training, this framework uses safe aggregation techniques to analyze distributed log data, allowing efficient anomaly and cyber threat detection.

3.4 Cross-Domain and Reinforcement Learning Applications

FL's versatility and scalability have been demonstrated by its successful extension outside typical domains to accommodate cross-domain collaborations and reinforcement learning (RL) activities. In [3], a multimodal data analysis framework based on FL is presented in the context of disaster management. In order to enable quick catastrophe response without jeopardizing sensitive data, this system combines textual, visual, and geographic data while using privacy-preserving strategies.As shown in [11], where a cooperative maze-solving system is created, reinforcement learning tasks have also been combined with FL. FL's promise in multi-agent and decision-

making contexts is demonstrated by the fact that several agents may train their rules locally while maintaining privacy.

Lastly, the difficulties of cross-device learning in resource-constrained contexts are addressed by the U-shaped Split Federated Learning (SplitFL) system introduced in [16]. SplitFL greatly minimizes computation on client devices while maintaining privacy by providing just intermediate activations rather than raw data. Because of this, it is ideal for applications involving mobile platforms and low-power IoT devices.

3.5 Comparative Insights

A thorough examination of the papers that were examined shows that privacy-preserving Federated Learning (FL) is being used more and more in important fields including cybersecurity, IoT, healthcare, and smart environments. Due to stringent legal requirements and the sensitive nature of patient data, the healthcare industry emerges as the top application sector. Methods such as Secure Multi-Party Computation (SMPC) [14] and Differential Privacy (DP) [7][9][12] are frequently used to safeguard medical information while permitting cooperative model training. According to studies [1][7][9][12][14], edge-based and hierarchical FL designs are popular because they strike a balance between robust privacy assurances and computational performance. In the meanwhile, applications for the Internet of Things and smart environments concentrate on communication effectiveness, lightweight models, and blockchain integration [5][8][13][15][17]. To function well in resource-constrained environments, these applications make use of model compression, client sampling, and adaptive aggregation.

FL's expanding importance in privacy-preserving anomaly detection, intrusion detection, and electricity theft detection [2][6][10] is demonstrated by cybersecurity-related applications, where safeguarding sensitive records and real-time data is crucial. Cross-domain FL and reinforcement learning (RL) [3][11][16] are being further investigated in emerging research, which applies FL to dynamic and cooperative tasks such cross-device learning, disaster response, and labyrinth solving. Due to its formal privacy guarantees, DP is still the most widely used privacy-preserving technology [7][9][12], although other methods like Homomorphic Encryption (HE) [14], Blockchain/DLT [15], and Clustered FL models [17] are becoming more popular. The necessity for more research on scalable and resilient FL frameworks is highlighted by the persistence of important issues including Non-IID data handling, communication overhead, adversarial robustness, and preserving model correctness in spite of recent developments.

4. Taxonomy and Comparative Analysis of Privacy-Preserving Federated Learning Approaches

Rapid developments in privacy-preserving Federated Learning (FL) have produced a wide variety of topologies, privacy methods, and aggregation methodologies that are suited to different application domains' unique needs. This section provides a thorough taxonomy of FL frameworks based on architectural design, privacy-preserving mechanisms, aggregation techniques, and application areas in order to methodically comprehend and compare various approaches. In order to highlight new developments, areas of strength, and significant obstacles that continue to influence the researchenvironment, it also offers a comparative analysis of the 17 examined publications.

4.1 Taxonomy of Federated Learning Frameworks

When evaluating FL systems' viability for certain applications, their architectural design is crucial, particularly when taking resource limitations, privacy considerations, and data dissemination patterns into account. Horizontal Federated Learning (HFL), in which participating clients have different data samples but share the same feature space, is the most popular architecture. HFL is widely used in healthcare applications, including skin disease categorization, COVID-19 detection, and IoMT frameworks [1][7][9][14]. It allows collaborative model training while protecting patient privacy.

Vertical Federated Learning (VFL), on the other hand, focuses on situations when clients have diverse feature sets but the same user base, despite being less studied in the reviewed articles. Federated Transfer Learning (FTL), another new architectural approach, allows clients to share knowledge with no overlap in features or data, as seen in cross-domain catastrophe analysis systems [3]. Clustered and Hierarchical FL designs have been proposed to address the difficulties associated with Non-IID data distributions and large-scale deployments. These methods, which add multi-level aggregation layers or group clients based on data similarity, are especially useful in wearable technology, healthcare, and smart settings [12][17].

Another innovative architecture created for resource-constrained contexts, such as the Internet of Things and mobile devices, is Split Federated Learning (SplitFL). SplitFL transmits just intermediate activations instead of raw data, hence reducing the computational load on clients while preserving privacy by splitting the model between client and server [16].

Furthermore, FL frameworks based on Reinforcement Learning (RL) have been created for dynamic decision-making tasks in settings like multi-agent systems and IoT resource allocation [2][11].

A number of privacy-related strategies have been incorporated into FL systems to protect private information. The most used approach, Differential Privacy (DP), adds calibrated noise to model updates to provide mathematical assurances. Applications where data sensitivity is high, including healthcare and IoT, frequently employ DP [7][9][12]. Another reliable method that increases resistance to poisoning assaults is Secure Multi-Party Computation (SMPC), which enables several parties to collaboratively calculate model parameters without disclosing personal information [14]. To further protect privacy, certain healthcare systems have contemplated using Homomorphic Encryption (HE), albeit less frequently, to perform calculations on encrypted data.

Integrating Blockchain and Distributed Ledger Technologies (DLT) into FL frameworks is becoming more and more popular, especially for smart city and Internet of Things applications [15]. Blockchain increases participant confidence by ensuring tamper-proof, auditable recordings of model modifications. In IoT environments, methods like lightweight encryption and model compression are also used to lower energy usage and communication overhead [5][6].

Because they dictate how local model updates are aggregated to create the global model, aggregation methods are essential to FL. A popular method for safeguarding client privacy is Secure Aggregation, which stops the server from viewing individual updates [1][10][14]. By carrying out local aggregation at edge nodes prior to sending updates to the central server, Edge and Hierarchical Aggregation techniques are used to increase scalability and lower communication costs in large-scale and resource-constrained contexts [5][12][13]. Furthermore, several frameworks use customized learning techniques and adaptive aggregation to improve model performance in dynamic and heterogeneous contexts. This enables models to adapt in response to user feedback and local data properties [8].

Due to stringent regulations and the sensitive nature of patient data, the healthcare industry is the largest adopter of privacy-preserving FL. IoMT, illness detection, and mental health monitoring studies [1][7][9][12][14] demonstrate how well FL works to support team-based medical research while protecting patient privacy. Lightweight FL models and blockchain-based techniques are used in IoT and smart settings, including as smart healthcare equipment, personalized learning programs, and urban parking management, to overcome resource limitations and improve trust. [5][8][13][15][17].

Without disclosing raw system logs, cybersecurity apps use FL to carry out delicate operations like intrusion detection and electricity theft detection. [2] [6] [10]. In order to preserve data security and enable precise threat detection, these systems frequently use RL-based decision-making or secure aggregation approaches. Furthermore, RL-integrated and cross-domain FL frameworks are developing to tackle challenging problems as cooperative maze-solving and disaster response [3][11][16]. These models show how FL can handle a variety of data kinds and decision-making situations while maintaining privacy.

Table 2. Comprehensive Comparison of Privacy-Preserving Federated Learning Applications, Architectures, Techniques, and Challenges

Paper	Domain / Application	FL Architecture	Privacy Techniques	Dataset / Data Type	Aggregation / Model Type	Key Contribution	Challenges Addressed
[1]	IoMT Healthcare	HFL	Secure Authentication, DP	Medical sensor data	Secure Aggregation	Privacy-preserving FL for IoMT	Device authentication, Privacy leaks
[2]	IoT Resource Allocation	RL-based FL	Local DP	IoT device logs	RL Policy Updates	FL + RL for optimized	Data leakage, Resource constraints

						resource allocation	
[3]	Disaster Analysis	Multimodal FL	Secure Aggregation	Text, Images, Geospatial data	CNN, Secure Aggregation	Privacy-preserving multimodal FL	Data heterogeneity, Privacy
[4]	IoT Systems	Lightweight FL	Lightweight Encryption	IoT device streams	Lightweight Aggregation	Fair FL with efficient encryption	Resource constraints, Scalability
[5]	Smart Healthcare IoT	Edge FL	Lightweight Encryption	Healthcare sensor data	Edge Aggregation	Energy-efficient FL model	Communication overhead
[6]	Smart Grid	HFL	Model Compression	Electricity usage data	Compressed Aggregation	Privacy-preserving theft detection	Communication cost, Model size
[7]	Healthcare COVID-19 Detection	HFL	DP	Chest X-ray images	CNN + DP	COVID-19 detection with privacy	Medical data privacy, Accuracy loss
[8]	IoT Personalized Learning	Personalized / Adaptive FL	Model Personalization	User behavioral data	Feedback Loops, Adaptive Models	User-centric FL with feedback	Personalization, Model drift
[9]	Healthcare Skin Disease	HFL	DP	Skin lesion images	CNN + DP	Privacy-preserving skin disease classification	Non-IID Data, Privacy-Accuracy trade-off
[10]	Cybersecurity / IDS	Edge FL	Secure Aggregation	Network traffic logs	Anomaly Detection Models	Intrusion detection using FL	Privacy of log data, Anomaly detection
[11]	RL Multi-Agent Maze Solver	RL-based FL	Privacy-preserving RL	Agent state-action data	RL Policy Gradients	Collaborative RL with privacy	Adversarial robustness, Data leakage
[12]	Mental Health / Wearables	Hierarchical FL	Local DP	Wearable physiological data	Hierarchical Aggregation	Stress detection with DP	Privacy, Real-time data
[13]	E-learning / User Activity	Hierarchical FL	Privacy-Preserving FL	On-screen activity data	Hierarchical Aggregation	User activity tracking with FL	Data collection privacy
[14]	Digital Healthcare	HFL	SMPC	Healthcare records	Secure Aggregation	Robust FL against poisoning attacks	Poisoning, Data inference

[15]	Smart City / Parking	Blockchain FL	Blockchain / DLT	Smart parking data	Blockchain-based Aggregation	Secure parking prediction	Trust, Tamper-proof updates
[16]	Cross-Device IoT	SplitFL	Split Learning	Mobile sensor data	Intermediate Activations	Resource-efficient FL	Client resource limits, Data security
[17]	Wellness Detection	Clustered FL	Clustering for privacy	IoT / Wellness data	Clustered Model Updates	Personalized wellness monitoring	Non-IID, Accuracy, Personalization

4.2 Comparative Analysis of Reviewed Papers

Several significant tendencies are shown by the comparative analysis of the 17 evaluated papers. Because of the necessity for strong privacy safeguards and regulatory compliance, healthcare applications continue to be the main emphasis. To guarantee accuracy and privacy, the majority of healthcare FL systems use hierarchical architectures, or HFL, in conjunction with DP and SMPC. [1][7][9][12][14]. On the other hand, IoT and smart settings prioritize communication-efficient methods and lightweight models to get around resource constraints, frequently utilizing blockchain, client sampling, and model compression. [5][8][13][15][17].

Applications in cybersecurity show how FL is becoming more and more important for privacy-preserving anomaly and intrusion detection jobs. In order to safeguard private log data and provide real-time threat detection, these models use RL-based learning and secure aggregation procedures [2][6][10]. The rise of RL-based and cross-domain FL frameworks demonstrates how flexible FL is in managing a variety of dynamic activities. These experiments use cutting-edge methods like SplitFL and multimodal data integration, which allow for collaborative learning across disparate datasets while protecting privacy. [3] [11] [16].

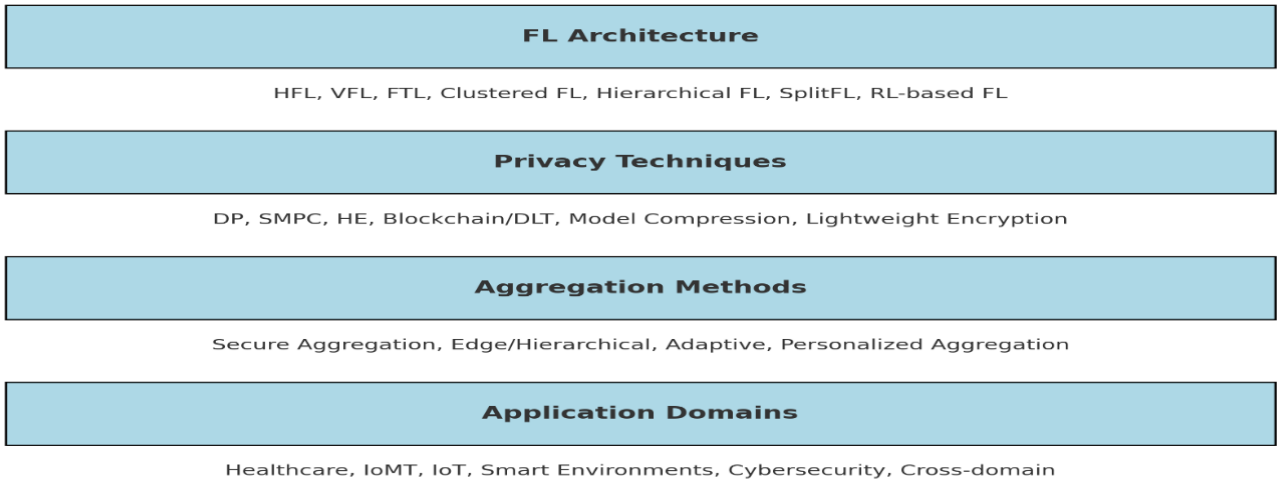

Fig. 1. Taxonomy of privacy-preserving federated learning frameworks

4.3 Key Observations and Insights

Overall, the research shows that because of its robust mathematical guarantees and simplicity of integration, Differential Privacy (DP) continues to be the most popular privacy-preserving approach. To guarantee transparency and immutability of model updates, there is a discernible increase in the use of Blockchain and DLT, particularly in IoT and smart city applications [15]. While adaptive and customized FL models provide potential options for enhancing accuracy and user-centric learning, clustered and hierarchical structures are being employed more and more to address Non-IID data difficulties and scalability issues.

Despite these developments, a number of issues still exist in several fields. Unresolved research issues include managing Non-IID data distributions, cutting down on communication cost, guaranteeing adversarial resilience, and striking a balance between model accuracy and privacy assurances.

Future research should concentrate on creating more effective aggregation techniques, maximizing computing overhead for devices with limited resources, and strengthening FL systems' resistance to new security risks.

Table 3. Enhanced Taxonomy Summary of Privacy-Preserving Federated Learning Frameworks

Category	Description	Techniques / Features	Example Papers
FL Architecture	Types of data distribution and model design	HFL, VFL, FTL, Clustered FL, Hierarchical FL, SplitFL, RL-based FL	[1][7][9][12][16]
Privacy Techniques	Methods used to enhance data privacy and security	DP, SMPC, HE, Blockchain/DLT, Model Compression, Lightweight Encryption	[7][9][12][14][15]
Aggregation Methods	Techniques used to aggregate local model updates securely and efficiently	Secure Aggregation, Edge/Hierarchical Aggregation, Adaptive Aggregation, Personalized Aggregation	[1][5][10][12][13]
Application Domains	Target sectors where privacy-preserving FL is applied	Healthcare and IoMT, IoT and Smart Environments, Cybersecurity / IDS, Cross-domain and RL tasks	[1][5][7][10][11][15]

5. Security Threats and Countermeasures in Privacy-Preserving Federated Learning

Federated Learning (FL) is susceptible to a number of security risks, despite the fact that it provides notable benefits in terms of protecting data privacy by design. FL's decentralized structure creates additional vulnerabilities, especially in hostile environments where the availability, confidentiality, or integrity of the system might be jeopardized by malevolent actors or outside attackers. The main security risks of privacy-preserving FL systems are thoroughly examined in this part, along with the solutions suggested in the evaluated studies to lessen these risks.

5.1 Security Threats in Federated Learning

Federated Learning (FL) has a privacy-preserving architecture by default, but it is nevertheless vulnerable to a number of security risks that might jeopardize the system's integrity and secrecy. Due to the possibility of sensitive information leakage or manipulation by hostile groups, the decentralized architecture of FL creates distinct attack surfaces. Comprehending these dangers is essential to creating resilient FL systems that can function safely in hostile, real-world settings.

5.1.1 Model Inversion Attacks

Model inversion Attacks pose a serious risk to privacy in Florida, particularly in industries like healthcare, banking, and surveillance that handle extremely sensitive data. In order to reconstitute input features or even whole data samples that clients utilize for local training, attackers exploit the model's output or the gradients that were exchanged during training. The gradient information may inadvertently encode certain features of the training data, even if raw data is never explicitly exchanged in FL. For example, attackers may use gradient updates to rebuild medical pictures, biometric characteristics, or personal health information in healthcare contexts, resulting in serious privacy violations [7][9]. In applications requiring wearable sensor data or medical imaging, where even partial reconstruction of sensitive features might have serious ethical and legal ramifications, this issue is especially concerning.

5.1.2 Membership Inference Attacks

To ascertain if a particular data sample was a part of the training dataset, membership inference attacks take use of flaws in FL frameworks. Adversaries can determine if certain data points are present or lacking by examining the model's answers or parameter changes, thereby endangering user privacy. Such assaults are extremely serious in industries like medical treatment, where even exposing that a user's data was used in training may reveal their medical status or involvement in a research. It could disclose a user's involvement in certain financial activities. Even seemingly harmless modification to the model can reveal membership information in FL, where models are trained on decentralized datasets, if proper safeguards like differential privacy are not used [12][14].This attack compromises one of the fundamental promises of FL - preserving data anonymity.

5.1.3 Poisoning Attacks (Data and Model)

Because they specifically target the integrity of the global model, poisoning assaults are one of the most harmful dangers to FL systems. These assaults fall into two categories: model poisoning and data poisoning. Malicious clients purposefully alter their local training datasets to introduce detrimental patterns or inaccurate labels in a process known as data poisoning. The accuracy and dependability of the entire model are subsequently diminished when the contaminated data distorts the local model updates. In contrast, model poisoning occurs when adversaries submit well constructed gradients or model updates during the aggregation step in order to change the behavior of the global model. Attackers could, for instance, insert backdoors that cause misclassifications in response to particular inputs. In safety-critical systems like smart grids or healthcare diagnostics, where hacked models might result in catastrophic failures or wrong medical judgments, such assaults are especially deadly. [6] [14].

5.1.4 Sybil Attacks

Sybil attacks take advantage of FL's openness by enabling adversaries to establish several fictitious clients or identities inside the system. Attackers can get disproportionate control over the global model updates by injecting a large number of Sybil nodes. Biased models, model deterioration, or the successful installation of backdoors are possible outcomes of this manipulation. Sybil attacks have the potential to seriously disrupt the FL process and erode participant trust. Large-scale FL systems with dynamic client engagement, such IoT networks or smart city infrastructure, are particularly vulnerable to this attack vector [15]. FL systems are especially susceptible to this type of attack as they may create many identities without robust authentication procedures.

5.1.5 Gradient Leakage and Side-Channel Attacks

FL is vulnerable to gradient leaking, in which adversaries take confidential information from the gradients disclosed during training, even while it blocks direct access to raw data. By closely examining these gradients, attackers might deduce private information about the local data, so undoing FL's privacy advantages. In order to obtain knowledge of the private data or model parameters, side-channel attacks also take use of indirect information like computing time, energy consumption, or communication patterns. These assaults are especially problematic in situations with restricted resources, such as Internet of Things devices, where the danger of leakage is increased by predictable behavior patterns and limited processing capabilities [5][6]. Without direct data access, such attacks have the potential to reveal comprehensive behavioral or personal information if they are successful.

5.2 Countermeasures and Defense Mechanisms

Researchers have created a number of remedies to mitigate the particular risks provided by decentralized learning settings in order to protect the privacy and integrity of Federated Learning (FL) systems. A range of defense tactics are suggested in the reviewed studies, all of which aim to balance system performance, communication overhead, and model accuracy while focusing on certain attack vectors.

Together, these methods improve FL frameworks' robustness, especially in delicate application areas like cybersecurity, IoT, and healthcare.

Table 3: Summary of Security Threats in Federated Learning

Threat Type	Description	Impact/Example
Model Inversion Attack	Reconstruct private input data from gradients or model outputs.	Reconstruct medical images from gradients [7][9].

Membership Inference Attack	Infer if specific data was part of training.	Reveal participation in healthcare dataset [12][14].
Poisoning Attack	Malicious clients corrupt local data or gradients.	Inject backdoors or skew model predictions in healthcare [6][14].
Sybil Attack	Malicious entities create fake clients to manipulate model.	Bias model behavior or inject backdoors [15].
Gradient Leakage/Side-channel Attack	Extract private information from gradients or system-level traces.	Infer user behavior in IoT environments [5][6].

5.2.1 Differential Privacy (DP)

In Federated Learning (FL), Differential Privacy (DP) has become one of the most popular and ethical defenses for safeguarding the privacy of personal information. Fundamentally, DP makes sure that the results of a calculation, such a model update, are statistically same regardless of whether a certain data sample was included in the dataset. This is accomplished by disguising the impact of any one data point by adding calibrated random noise to either local gradients or aggregated model updates.

DP is commonly used in the local training stage of federated learning, where each client introduces noise into its model updates prior to sending them to the central server. Even if the model is eventually made public or examined, this procedure keeps the server—or any listening adversary—from discovering particulars about a client's confidential dataset. Model inversion and membership inference attacks, which seek to recreate sensitive data or determine if certain data records participated in the training set, are especially successfully thwarted by DP.

The privacy budget, represented by ε (epsilon), is used to measure how well privacy is protected in DP. A bigger ε permits better utility with lower privacy protection, whereas a smaller ε gives tighter privacy protections at the expense of decreased model accuracy. Because of this, tuning ε is a crucial design decision that needs to be carefully calibrated depending on the data's sensitivity, the number of training rounds, and the required precision.

DP has been effectively used in real-world FL circumstances in a number of the research described in this study. A COVID-19 detection mechanism, for instance, uses DP in [7] to safeguard private medical imaging data gathered from several institutions. A DP-enhanced FL approach for classifying skin diseases is also presented in [9], which allows models to be trained on delicate dermatological pictures without jeopardizing patient privacy. In order to protect user physiological signals from exposure during the learning process, a distinct research [12] incorporates local differential privacy into wearable-based stress detection systems.

These use examples highlight DP's adaptability and efficiency in high-stakes industries like IoT and healthcare. DP's possible effect on model performance is one of its main drawbacks, though. The usefulness of the trained model might drastically decrease when more noise is added to ensure better privacy (lower ε), particularly in challenging deep learning tasks. Furthermore, the cumulative impact of noise across several FL communication rounds might make this problem worse, hence it's critical to use privacy accounting strategies (such moments accountant or Rényi DP) to track the overall privacy loss over time.Researchers have also looked on adaptive DP techniques, which dynamically modify noise levels according to training success or client-specific data sensitivity, in an effort to strike a compromise between privacy and utility. Hybrid models that integrate DP with cryptographic methods such as Homomorphic Encryption (HE) or Secure Multi-Party Computation (SMPC) have also demonstrated potential in improving security while maintaining model integrity.

In summary, the design of privacy-preserving FL systems continues to rely heavily on Differential Privacy. It is the go-to option for guaranteeing data anonymity in decentralized machine learning because of its formal assurances, adaptability, and proven efficacy in practical implementations. However, finding the best balance between privacy, accuracy, and scalability remains an unexplored field that needs more creativity.

5.2.2 Secure Multi-Party Computation (SMPC)

A strong cryptographic system called Secure Multi-Party Computation (SMPC) was created to allow several parties to work together to calculate their private inputs without disclosing those inputs to one another. To make sure that the server or any participating client cannot access individual model updates while generating a collective global model, SMPC is mostly used during the model aggregation phase of Federated Learning (FL).

Every client trains a model locally in a typical FL configuration, then sends the modified weights or gradients to a central server for aggregation. However, the risk created by this central point of collection is that, if hijacked, the server might examine individual client updates and deduce confidential data characteristics. In order to prevent any one entity from possessing all the information, SMPC distributes model changes as encrypted shares across several parties, frequently including the server. An aggregated model update is then generated by combining these secret shares using secure arithmetic methods, all without ever decrypting the individual contributions.

The information-theoretic security of SMPC is its main advantage; even if a subset of players colludes, they won't be able to reassemble the original inputs until a certain threshold is crossed. Because raw gradients or weights are never revealed during computation, SMPC is especially resistant to gradient leakage, model inversion, and poisoning assaults.
Shamir's Secret Sharing Scheme, which divides a secret into pieces and gives each participant a portion, is a popular SMPC technique in FL. Only when a sufficient number of shares are merged can the original secret be recreated. In FL aggregation, this threshold-based security approach guarantees fault tolerance and privacy.

In real-world applications, SMPC has been investigated in extremely delicate fields like healthcare, where maintaining data secrecy is crucial. To create a federated diagnostic model, for example, SMPC is utilized in [14] to safely combine medical data from several healthcare facilities. With this configuration, hospitals may work together to gain from shared model knowledge without disclosing patient data, which is essential for compliance with laws like HIPAA and GDPR.

Even with its robust security assurances, SMPC has drawbacks. Its computational and communication overhead is one of the main issues. In contrast to conventional aggregation techniques, SMPC necessitates more intricate mathematical calculations and several rounds of client-to-client message exchanges, which can greatly raise latency and bandwidth consumption. This is especially troublesome in situations with resource-constrained devices like wearables, smartphones, or embedded IoT systems, or in large-scale FL installations.

Recent studies have concentrated on improving SMPC protocols for FL application cases in order to overcome these problems. To lower overhead while preserving robust privacy guarantees, strategies include compressing encrypted shares, cutting down on the number of interaction rounds, or combining SMPC with lightweight methods like differential privacy. Furthermore, hybrid architectures that use blockchain-based auditing tools or homomorphic encryption with SMPC provide additional security in hostile environments.

In conclusion, a reliable and theoretically sound technique for safe aggregation in federated learning is safe Multi-Party Computation. It is essential for privacy-preserving model training since it can withstand strong inference assaults and do away with the necessity for a reliable aggregator. To enable wider usage in actual, large-scale federated ecosystems, its scalability and performance issues must be resolved.

5.2.3 Homomorphic Encryption (HE)

An sophisticated cryptographic approach called homomorphic encryption (HE) makes it possible to execute mathematical operations directly on encrypted data, resulting in encrypted results that, when decrypted, match the results of operations on the plaintext. Because it enables a central server to aggregate encrypted model updates without ever knowing the actual values, this trait is very useful in privacy-preserving Federated Learning (FL).

Clients send model weights or gradients to a central server for aggregation in conventional FL setups. Sensitive information may still leak from the disclosed updates even while the raw data stays local. By enabling clients to encrypt their local model changes before to transmission, homomorphic encryption reduces this danger. Only authorized parties, usually the participating clients, may decode the final result once these encrypted updates have been aggregated by the central server. As a result, the server cannot obtain any useful insight into specific updates, even if it is compromised or malevolent.

HE offers robust defense against a range of privacy risks, including as malicious aggregation, gradient leaks, and model inversion assaults. Throughout the whole calculation process, attackers are essentially blindfolded since encrypted updates never disclose intermediate values. Because of this, HE is especially appealing in highly regulated or sensitive industries like healthcare, banking, and critical infrastructure systems, where even little amounts of data exposure might have ethical, legal, or security repercussions. In the healthcare industry, for instance, HE can guarantee the confidentiality of model updates based on patient medical information throughout the collaborative learning process — a situation examined in [14].

Partially homomorphic encryption (PHE), somewhat homomorphic encryption (SHE), and fully homomorphic encryption (FHE) are some of the several forms of homomorphic encryption. The most complete capability of these is provided by FHE, which supports addition and multiplication on encrypted data. FHE is the most computationally demanding, though, which makes it difficult to implement in actual FL settings.

HE's computational expense and latency are its primary disadvantages. Compared to their plaintext counterparts, ciphertexts require a lot more resources to operate on and to encrypt model parameters. In large-scale installations with thousands of clients or real-time applications, this additional overhead creates a bottleneck. Both servers and edge devices may be strained by the processing time and memory needs, especially in FL systems that include frequent communication and recurrent updates.

In order to tackle these problems, scientists are investigating effective HE systems designed for FL, such BGV (for precise calculations) and CKKS (for approximation arithmetic), which provide a trade-off between functionality and efficiency. In order to reconcile privacy assurances with computational viability, hybrid privacy-preserving methods are also being developed, which combine HE with Differential Privacy (DP), Secure Multi-Party Computation (SMPC), or Blockchain.

In conclusion, homomorphic encryption is a mathematically elegant and highly secure solution for privacy-preserving aggregation in FL. Its ability to prevent raw data exposure at all stages of model training makes it a crucial tool in the FL security toolkit. However, ongoing research must continue to address its inherent performance bottlenecks, making HE more scalable, efficient, and adaptable to the varied needs of modern federated systems. Another promising approach is offloading encrypted computations to trusted execution environments (TEEs) or using edge-cloud architectures, which use edge nodes to handle lightweight encryption while more powerful cloud servers handle secure computation.

5.2.4 Blockchain and Distributed Ledger Technologies (DLT)

The danger of malevolent clients interfering with the learning process is a critical issue as Federated Learning (FL) systems grow into open and possibly hostile settings. The two most serious dangers are Sybil attacks and data/model poisoning assaults, in which attackers either fabricate clients to influence learning outcomes or tamper with local data to deceive the global model. FL frameworks are progressively including powerful aggregation algorithms and anomaly detection systems to detect, mitigate, or entirely filter out fraudulent updates during model training in order to combat these risks.

In order to detect anomalies in FL, incoming model updates from participating clients are continually monitored and compared to past patterns or predicted statistical behaviors. These systems are able to spot anomalies that can point to manipulation in weight distributions, gradient magnitudes, or performance indicators. To stop these aberrant updates from affecting the global model, they might be down-weighted or deleted after being identified. Cosine similarity criteria, variance-based filters that compare recent updates to earlier iterations, and z-score-based detection are common techniques.

Robust aggregation techniques offer robustness in addition to anomaly detection by altering the computation of the global model. Robust approaches take into consideration the potential for hostile contributions, in contrast to the conventional Federated Averaging (FedAvg) method, which presumes that all client updates are equally reliable. For example:

- The effect of extreme values (perhaps tainted updates) is eliminated or diminished by the Trimmed Mean and Median
  Aggregation.

- Krum minimizes the impact of outliers by choosing updates that are most relevant to the bulk of participants.

- Reputation-based aggregation penalizes players who exhibit persistently suspicious conduct by tracking client behavior over several rounds and allocating weights or trust ratings to each update.

These techniques work especially well against Sybil attacks, in which phony clients try to overload the aggregation process, and model poisoning, in which attackers create updates to introduce backdoors or impair performance.

[10] provides an example use scenario in which anomaly detection and secure aggregation techniques are integrated in a cloud-edge intrusion detection system (IDS). Without disclosing their raw system logs, the distributed edge nodes in this system work together to build an anomaly detection model. Secure aggregation protects data privacy, while anomaly detection filters tainted updates. When it comes to identifying cyberthreats across dispersed networks, this dual-layer protection improves security and secrecy.

These techniques have drawbacks despite their advantages. Particularly in predominantly Non-IID environments, false positives in anomaly detection might exclude valid but varied data inputs, decreasing the generalizability of the model. Furthermore, robust aggregation methods may come with additional communication and computational expenses, particularly in large-scale systems where involvement fluctuates often. Finding a balance between protecting innocuous data heterogeneity and filtering harmful behavior is still an outstanding research issue.

Additionally, robust aggregation techniques that are context-aware and flexible are showing promise as remedies. These methods make better judgments about accepting or rejecting updates by taking into account model convergence dynamics, client behavior history, and environmental variables. There are also new ways to incorporate strong defenses without sacrificing FL's privacy goals thanks to recent developments in federated adversarial learning and privacy-preserving trust scoring.

In conclusion, strong aggregation and anomaly detection are essential for protecting Federated Learning systems from internal attacks. Even in large-scale or untrusted distributed systems, these strategies maintain the integrity, fairness, and performance of global models by guaranteeing that tainted, manipulated, or aberrant updates have little impact. These countermeasures will be essential to ensuring privacy and robustness in decentralized AI systems as FL expands into more complicated contexts.

5.2.5 Anomaly Detection and Robust Aggregation

Federated Learning (FL) is especially vulnerable to internal threats like poisoning and Sybil assaults because it functions in decentralized and often untrusted contexts. Malicious clients insert tainted updates to impair model performance or introduce backdoors in poisoning attacks. Sybil attacks include the creation of many false identities by adversaries in order to control the process of model aggregation. Many FL systems now include anomaly detection and robust aggregation techniques as essential parts of their security structure in order to address these issues.

Mechanisms for detecting anomalies are made to keep an eye out for strange or suspicious activity in incoming client updates. These systems indicate changes that substantially depart from predicted patterns using statistical or machine learning techniques. Typical tactics include noting discrepancies across several training cycles, monitoring update size, and calculating the distance between a client's update and the global model. FL systems can lessen the impact of potentially harmful updates prior to aggregation by detecting such abnormalities. These systems are particularly important in settings when it is impossible to ensure customer confidence.

Robust aggregation techniques work in tandem with anomaly detection to lessen the effects of malicious or outlier updates during model fusion. Robust approaches either remove extreme values or allocate weights based on believability, in contrast to traditional averaging, which considers all updates equally. Algorithms like Krum choose the most representative update based on resemblance to others, Median Aggregation chooses the median rather than the mean, and Trimmed Mean eliminates the greatest and lowest values from each parameter. Furthermore, reputation-based algorithms monitor customer behavior over time, which enables the model to give priority to updates from reliable players.

These methods have shown promise in practical applications, especially in fields where security is crucial. In [10], for example, an edge-based intrusion detection system (IDS) that protects the privacy of individual system logs uses both anomaly detection and safe aggregation to jointly train a network threat detection model. This system demonstrates the usefulness of these cybersecurity measures by enabling efficient screening of harmful updates without jeopardizing the confidentiality of important log data.

Implementing robust aggregation and anomaly detection presents difficulties despite their advantages. Particularly in Non-IID setups, false positives can happen when valid but varied data results in odd updates. Furthermore, certain algorithms include communication and computational overheads that make them unsuitable for devices with limited resources. Thus, research on striking a balance between robustness, justice, and efficiency is still ongoing. In order to improve security and inclusiveness, future research may examine hybrid approaches that combine anomaly detection with privacy-preserving trust evaluation, or adaptive defenses that modify thresholds in response to system activity.

5.2.6 Model Compression and Lightweight Encryption

Because training in Federated Learning (FL) is client-driven and decentralized, it is naturally susceptible to internal malicious activity. Sybil attacks, in which a single adversary impersonates several phony clients to obtain disproportionate influence, and poisoning attacks, in which hostile clients introduce modified data or model updates to distort the global model, are two of the most disruptive dangers. A increasing number of FL frameworks now include strong aggregation algorithms and anomaly detection systems to counter these risks, providing a crucial line of defense for system trust and model integrity.

Anomaly Detection in FL

Finding odd or unexpected patterns in client updates that can point to malicious activity is known as anomaly detection. These systems function either proactively, continually monitoring for departures from expected behavior based on statistical or historical standards, or reactively, detecting abnormalities as they happen.

In FL, common methods for detecting anomalies include:

- Distance-based outlier detection: Calculates the cosine or Euclidean distance between each client's update and the global model from the previous round or the mean update.
- Update norm analysis: Clients with gradient magnitudes that are abnormally big or small are marked for further examination.
- Temporal analysis: To spot abrupt variations, the consistency of a client's behavior across several cycles is monitored.

These techniques aid in spotting both one-off poisoning attempts and longer-term hostile trends. But problems like false positives, which confuse benign outliers for malicious updates, need to be handled carefully, particularly when dealing with Non-IID data, where real client updates might vary greatly.

Robust Aggregation Techniques

Robust aggregation methods are used to either eliminate or lessen the impact of the questionable updates if anomalies are identified. To become resistant to adversarial input, these algorithms alter the FedAvg algorithm, which is a normal averaging procedure employed in ordinary FL. Prominently effective aggregation techniques consist of:

- Trimmed Mean: Prior to averaging, a predetermined percentage of each model parameter's top and lowest values are discarded.

- Median Aggregation: Provides resilience against extreme values by taking the element-wise median of all client modifications.

- Krum: Reduces the impact of outliers by choosing the client update that is closest (in Euclidean space) to the bulk of other updates.

- Reputation scoring: Using historical behavior and model contribution quality, it gradually increases each client's trust score.

These methods are particularly helpful in large-scale or untrusted federations with dynamic client involvement since they lower the possibility of faulty updates impacting the global model.

Real-World Application

In [10], secure aggregation and anomaly detection are applied to an edge-based intrusion detection system (IDS) in tandem, providing a striking example of this dual-layered technique. Edge nodes in this system keep an eye on network logs and work together to use FL to train an anomaly detection model. While anomaly detection guarantees that changes adding to the model are authentic and unaltered, secure aggregation safeguards the privacy of each node's local data. This configuration successfully strikes a compromise between model security and privacy preservation, proving the practicality of these methods in actual cybersecurity applications.

Challenges and Future Outlook

Despite their efficacy, these defenses continue to confront a number of obstacles:

- Computational overhead: Several strong aggregation techniques, like reputation scoring or Krum, require expensive calculations that are not appropriate for low-power devices.
- Communication latency: Training may be slowed down by the additional processing steps required to filter and reroute updates in real-time.
- Finding a balance between fairness and robustness: Excessively stringent screening may stifle valid contributions from customers with true data diversity (such as underrepresented ethnic groupings).

Adaptive resilient aggregation frameworks that take into account fairness in the face of heterogeneity, dynamically adapt to shifting threat levels, and learn from historical assault patterns must be the main focus of future study. Building safe and equitable FL systems will need integration with privacy-preserving trust management systems, federated adversarial training, and machine learning-based threat detection.

5.3 Summary of Countermeasures and Defense Mechanisms

Model inversion, membership inference, poisoning, Sybil attacks, and gradient leaking are just a few of the major security risks that are addressed by the many protection mechanisms found in privacy-preserving Federated Learning (FL) systems. By include calibrated noise in model updates, Differential Privacy (DP), one of the most popular approaches, provides mathematical privacy guarantees. Its applicability for safeguarding sensitive data is demonstrated by its successful incorporation in IoT and healthcare applications [7][9][12].

Strong cryptographic protections are offered by Secure Multi-Party Computation (SMPC) and Homomorphic Encryption (HE), which guarantee the confidentiality of individual model changes while they are being aggregated.

Although these methods reduce the risk of gradient leaking and poisoning assaults, they come with a computational cost that restricts the scalability of big systems [14].
By guaranteeing tamper-proof, auditable transactions, the combination of Blockchain and Distributed Ledger Technologies (DLT) offers an extra degree of security. This is especially helpful in reducing Sybil attacks and boosting confidence in smart city and Internet of Things applications [15]. Additionally, robust aggregation approaches like secure aggregation and outlier filtering, together with anomaly detection algorithms, enhance resistance against adversarial behaviors and maintain model integrity in hostile situations [10].Last but not least, model compression and lightweight encryption strategies improve communication efficiency while lowering the danger of side-channel attacks in IoT and edge contexts, addressing privacy and resource restrictions [5][6].

When taken as a whole, these countermeasures show how multi-layered security frameworks are essential for FL systems in order to guarantee privacy, integrity, and resilience in practical implementations.

Table 4: Summary of Countermeasures and Defense Mechanisms

Technique	Description	Targeted Threats	Challenges
Differential Privacy (DP)	Adds noise to gradients to mask individual data.	Inversion, Membership inference	Trade-off between accuracy and privacy [7][9][12].
Secure Multi-Party Computation (SMPC)	Secret shares aggregation without revealing individual updates.	Gradient leakage, Poisoning	High communication overhead [14].
Homomorphic Encryption (HE)	Compute directly on encrypted data.	Gradient leakage, Model inversion	High computational cost [14].
Blockchain / DLT	Immutable records for tamper-proof updates.	Sybil, Poisoning	Latency and energy overhead [15].
Anomaly Detection & Robust Aggregation	Detects and filters malicious updates.	Poisoning, Sybil	Requires adaptive mechanisms [10].
Model Compression & Lightweight Encryption	Reduces communication and risk of side-channel leaks.	Gradient leakage, Side-channel	Potential impact on accuracy [5][6].

6. Challenges and Future Research Directions

Federated Learning (FL) is still in its infancy and faces many substantial obstacles that prevent its widespread practical implementation, despite the fact that it has shown great promise in allowing privacy-preserving artificial intelligence. To increase the effectiveness, scalability, resilience, and reliability of FL systems, researchers must tackle these issues as FL frameworks spread throughout smart cities, cybersecurity, healthcare, and IoT applications. This section identifies the main open questions and suggests future lines of inquiry that are necessary to advance the area.

6.1 Handling Non-IID and Heterogeneous Data

Managing heterogeneous client datasets and Non-Independent and Identically Distributed (Non-IID) data is one of the most basic problems in federated learning. Because of user behavior, device kinds, ambient conditions, and geographic variety, data provided by various clients differs significantly in real-world applications. For instance, different diagnostic instruments and patient demographics cause patient data in healthcare apps to vary between hospitals and geographical areas. Similarly, it is unreasonable to assume homogeneous data dissemination because IoT devices gather data in a variety of circumstances [7][9][12].

Non-IID data degrades accuracy and performance by causing sluggish global model convergence, biased local models, and inconsistent updates. This variability is difficult for FedAvg and other traditional aggregating techniques to manage. Advanced methods such client clustering based on data similarity, customized FL models, and meta-learning strategies that adjust to different data distributions while maintaining fairness and equal performance among clients should be investigated in future studies.

6.2 Communication Overhead and Resource Constraints

Throughout several training cycles, FL need frequent communication between clients and the central server. Significant communication overhead is imposed by this recurrent interchange of big model updates, which is made worse in situations with limited resources, such as IoT networks and mobile devices, or in large-scale installations. FL is not appropriate for real-time or energy-sensitive applications because to its high communication costs, which can also saturate bandwidth, increase latency, and drain device battery life [5][6].

The development of communication-efficient protocols like gradient sparsification, quantization, model pruning, and adaptive client participation should be the main focus of future research. Additionally, designing lightweight FL models optimized for edge computing environments is essential for enabling FL in ubiquitous IoT systems. Strategies like asynchronous communication and hierarchical FL architectures that reduce the frequency and size of transmitted updates can also significantly lower overhead.

6.3 Robustness Against Adversarial Attacks

Security is a significant and ongoing problem due to Federated Learning's (FL) open, distributed, and heterogeneous nature. Unlike traditional centralized machine learning systems, FL leverages several independent clients that engage in model training without centralized supervision. FL's intrinsic openness makes it vulnerable to a range of adversarial attacks that compromise both data privacy and the model's integrity. The most prominent of these are model inversion attacks, where an adversary attempts to reconstruct private training data from model gradients; membership inference attacks, which ascertain whether a particular data point was used during training; and poisoning attacks, in which malicious clients introduce tainted data or gradients to bias or deteriorate the model.

As an example, an attacker could introduce poisoned updates that manipulate diagnostic models to misclassify medical images or suppress disease indicators. Similarly, Sybil attacks, in which one entity creates multiple fake clients to disproportionately influence model updates, can undermine the fairness and robustness of the global model, which can have catastrophic implications in scenarios like patient diagnosis or smart infrastructure control. These risks are particularly severe in sensitive domains like healthcare, where a single adversarial client could have disastrous consequences [6][14][15].

Basic privacy and security assurances are provided by existing countermeasures, such as Secure Multi-Party Computation (SMPC) and Differential Privacy (DP). DP uses noise to hide individual data contributions, but SMPC makes guarantee that model changes may be safely combined without being revealed. However, there are trade-offs between both methods in terms of processing complexity, communication cost, and model accuracy. Furthermore, systems are left vulnerable during training rounds because they are typically not built to actively monitor or react to adversarial conduct in real time.

Future FL frameworks must include intelligent and adaptive protection mechanisms that can react dynamically to changing threats in order to increase resilience. This involves including real-time attack detection systems, which identify questionable trends in client behavior or model modifications using statistical anomaly detection or machine learning. Outlier rejection, client reputation score, and robust aggregation algorithms (such Krum and Trimmed Mean) are some strategies that can assist reduce the impact of hostile clients while maintaining the contributions of truthful participants.

Additionally, new approaches like context-aware defense orchestration, which modifies the defense intensity according to threat levels or domain sensitivity, and federated adversarial training, which trains models to withstand adversarial perturbations in advance, present exciting research opportunities. These might be used with more conventional privacy-enhancing methods to provide all-encompassing and scalable solutions that strengthen FL's defenses against a variety of threats.

To sum up, protecting against hostile attacks is not just a technical need but also a precondition for the reliable implementation of Federated Learning in important real-world applications. Building safe, dependable, and privacy-preserving FL systems that are resistant to passive and aggressive threats will need ongoing efforts in this field.

6.4 Balancing Privacy, Utility, and Computational Efficiency

One of the most difficult and enduring problems in Federated Learning (FL) is striking the correct balance between model utility, privacy protection, and computing performance. Although privacy-enhancing technologies like Secure Multi-Party Computation (SMPC), Homomorphic Encryption (HE), and Differential Privacy (DP) provide robust theoretical assurances for safeguarding private information, they frequently have substantial trade-offs. As seen in several implementations in the healthcare and Internet of Things sectors, they include increased latency, additional computational and memory cost, and a decrease in model correctness as a result of noise injection or computational approximation [7][14].

In real-world deployments where FL must function on resource-constrained devices such as wearables, mobile phones, or low-power IoT sensors, these trade-offs become especially challenging. In some situations, learning may become completely unfeasible due to excessive encryption or aggressive noise methods that impede model convergence. Consequently, it is essential to build methods that preserve privacy in an adaptable manner. Based on the sensitivity of the local data, each client's processing capacity, and the importance of the work at hand, these systems ought to be able to dynamically modify privacy settings (such the encryption depth in HE or the noise level in DP) in real time.

Researchers are investigating hybrid FL frameworks that include many privacy strategies in an effort to further reduce the trade-offs between privacy, usefulness, and efficiency. For instance, combining DP with lightweight encryption enables local privacy protection and secure communication without having to pay the entire price of HE. Similar to this, HE systems can be optimized to provide quicker computing while maintaining acceptable privacy levels (e.g., by utilizing approximate encryption instead of precise encryption). Context-aware privacy budgets, which selectively enhance or loosen privacy restrictions based on the application area or data relevance, represent another interesting approach.

FL design should use multi-objective optimization techniques in addition to architectural enhancements. Instead than treating privacy, usefulness, and efficiency as separate trade-offs, these frameworks approach them as concurrent goals. Pareto-optimal solutions that attain a better balance across dimensions can be obtained by modeling these factors collectively. In heterogeneous systems, where one-size-fits-all privacy techniques are inadequate, such methods can be very helpful.

In conclusion, creating adaptable and intelligent systems that can adjust their behavior to the operational situation is more important for the future of privacy-preserving FL than merely attaining complete privacy or maximum accuracy. In order to make FL both theoretically sound and practically deployable in real-world, performance-sensitive contexts, it is imperative that this difficulty be addressed.

6.5 Scalability and Dynamic Participation

For FL, scalability is a major obstacle, particularly as systems grow to serve hundreds or millions of users. Model convergence and stability are made more difficult by managing dynamic client involvement, in which devices join and exit the federation regularly because of power or connection problems. This issue is made worse by the need to maintain equity among clients with different processing capacities and data quantities [5][15]. In order to guarantee reliable performance even in extensive deployments, future FL systems should integrate load balancing algorithms, adaptive client sampling, and hierarchical aggregation strategies. Furthermore, there is still a great need for research on federated systems that can handle dynamic populations, scale seamlessly, and guarantee equal participation.

6.6 Explainability and Interpretability of FL Models

The need for explainable and interpretable models is growing as Federated Learning (FL) is being used in more important and high-stakes fields including healthcare, finance, autonomous systems, and cybersecurity. From loan approvals and medical diagnoses to security threat responses, AI system judgments in various domains may have a big impact on people's lives and institutional operations. As a result, stakeholders—such as end users, subject matter experts, and regulators—need to know exactly how and why a federated model makes a given choice. Insufficient interpretability erodes confidence in FL systems and may make it difficult to comply with new AI governance regulations or legal frameworks like GDPR and HIPAA.

By attributing model choices to input characteristics, traditional Explainable AI (XAI) approaches like saliency maps, SHAP (SHapley Additive exPlanations), and LIME (Local Interpretable Model-agnostic Explanations) have demonstrated efficacy in centralized learning contexts. Nevertheless, there are particular difficulties when using XAI in FL settings. First, the capacity to execute global interpretation is limited since raw data is never exchanged among clients. Second, different clients will see different model behavior since FL models are usually trained on diverse and non-IID datasets. Both local interpretability (explaining specific predictions) and global interpretability (understanding the general logic of the model) are made more difficult by this.

Future studies must concentrate on creating privacy-preserving interpretation methods specifically designed for FL in order to overcome these problems. These techniques ought to uphold stringent data security while offering clear, accurate explanations of model behavior. Differentially private or encrypted explanation summaries might

be used to provide aggregated insights in a privacy-aware way, while local explainability approaches could be implemented fully on the client side utilizing private data. When producing explanations, federated XAI frameworks should also take into consideration differences in data distributions and model inconsistencies among clients.

Furthermore, transparency may be further improved without sacrificing privacy by directly integrating explainability into the FL training process, for example, by integrating interpretable surrogate models or attention methods. By giving clients the opportunity to comprehend and maybe audit the behavior of their local models, these systems would promote more responsibility and confidence throughout the FL ecosystem.
In conclusion, explainability is a crucial prerequisite for the proper implementation of FL models in delicate applications, not just a desired attribute. For FL systems to be equitable, responsible, and socially acceptable, it will be crucial to bridge the gap between black-box learning and human-understandable insights in a way that protects privacy.

6.7 Standardization, Benchmarking, and Regulatory Compliance

The lack of set standards, assessment procedures, and regulatory norms for privacy-preserving FL is another significant obstacle. Real-world adoption is slowed down and meaningful comparisons amongst FL techniques are impeded by the absence of agreement on performance measurements, privacy standards, and security requirements. Furthermore, although many industries need adherence to regulations like the AI Act, GDPR, and HIPAA, it is still challenging to implement in distributed learning environments.
The creation of thorough privacy and security benchmarks, standardized evaluation frameworks, and legal compliance toolkits customized for FL systems must be the top priorities of future research. Establishing moral principles and certification requirements for the responsible and safe use of FL technologies will need cooperation between academics, business, and legislators.

6.8 Summary

In conclusion, a number of interrelated issues must be resolved in order to advance Federated Learning as a workable, private, and secure machine learning paradigm. It is crucial to manage data heterogeneity, reduce communication overhead, bolster defenses against hostile attacks, and strike a balance between accuracy and privacy. Real-world adoption also depends on enhancing scalability, guaranteeing model interpretability, and complying with changing regulatory environments. In order to create robust, effective, and reliable FL systems that are appropriate for widespread deployment in delicate areas, future research must embrace interdisciplinary efforts that integrate machine learning, cryptography, optimization, and legal knowledge.

7. Conclusion

A game-changing concept for facilitating collaborative machine learning while protecting user privacy and data security is federated learning (FL). Because of its decentralized design, which enables numerous clients to collaborate on model training without exchanging raw data, it is especially well-suited for privacy-sensitive industries including cybersecurity, smart environments, healthcare, and the Internet of Things (IoT). A thorough taxonomy and overview of current privacy-preserving FL frameworks has been provided in this review paper, together with an analysis of their architectures, privacy strategies, aggregation approaches, and practical applications.

This evaluation demonstrates the increasing use of FL in crucial industries where data sensitivity and regulatory compliance are crucial by methodically examining a few peer-reviewed publications. To improve privacy guarantees and system robustness, methods like Homomorphic Encryption (HE), Secure Multi-Party Computation (SMPC), Differential Privacy (DP), and Blockchain-based aggregation have been extensively used. The research under examination show that although FL successfully tackles data privacy issues, it still has to deal with a number of enduring problems, such as managing Non-IID data, cutting down on communication overhead, enhancing resilience against hostile assaults, and striking a balance between privacy and model correctness.

Future research topics that are crucial for the advancement of the discipline are also identified in the report. These include improving model explainability, creating uniform benchmarks and regulatory frameworks, and creating scalable and communication-efficient FL systems. In order to fully realize Federated Learning's promise as a

reliable and broadly applicable solution for privacy-preserving artificial intelligence, several issues must be resolved.

In conclusion, in the age of large data and decentralized computing, federated learning is a crucial enabler for safe and moral AI applications. Overcoming present constraints and opening up new possibilities for FL in a variety of real-world situations will need sustained research and development in privacy-enhancing technologies, strong aggregation mechanisms, and multidisciplinary cooperation.

8. References

[1] F. Wu et al., "A Comprehensive Privacy-Preserving Federated Learning Scheme With Secure Authentication and Aggregation for Internet of Medical Things," *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 9652-9663, 2022.

[2] Y. Sun et al., "A Concurrent Federated Reinforcement Learning for IoT Resources Allocation With Local Differential Privacy," *IEEE Internet of Things Journal*, vol. 10, no. 1, pp. 541-553, 2023.

[3] R. Wang et al., "A Federated Learning Approach to Multimodal Data Privacy for Rapid Disaster Analysis," *IEEE Access*, vol. 10, pp. 57559-57572, 2022.

[4] H. Zhang et al., "A Privacy-Preserving Federated Learning Framework With Lightweight and Fair in IoT," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 1436-1444, 2023.

[5] L. Guo et al., "Edge Intelligence Federated Learning-Based Privacy Protection Framework for Smart Healthcare Systems," *IEEE Access*, vol. 10, pp. 102285-102295, 2022.

[6] Y. Li et al., "Electricity Theft Detection Based on Federated Learning," *IEEE Internet of Things Journal*, vol. 9, no. 2, pp. 1062-1071, 2022.

[7] M. Salehi et al., "Enhanced COVID-19 Detection and Privacy Preserving Using Federated Learning," *Journal of Biomedical Informatics*, vol. 129, 2022.

[8] B. Shen et al., "Enhancing Decentralized Federated Learning with User Feedback Loops: A Novel Approach for Personalized and Adaptive Learning in IoT Environments," *IEEE Internet of Things Journal*, vol. 10, no. 3, pp. 2321-2332, 2023.

[9] N. V. Sree et al., "Enhancing Skin Disease Classification and Privacy Preservation through Federated Learning-Based Deep Learning," *IEEE Access*, vol. 10, pp. 110843-110852, 2022.

[10] S. Zhao et al., "Intrusion Detection Based on Data Privacy in Cloud-Edge Collaborative Computing Using Federated Learning," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 6, pp. 6622-6631, 2023.

[11] T. Li et al., "Maze Solver: A Federated Reinforcement Learning Approach," *IEEE Transactions on Artificial Intelligence*, vol. 4, no. 1, pp. 67-76, 2023.

[12] A. R. Reddy et al., "Privacy-Preserved Stress Detection from Wearables using Federated Learning," *IEEE Access*, vol. 10, pp. 123456-123466, 2022.

[13] D. Wu et al., "Privacy-Preserving On-Screen Activity Tracking and Classification in E-Learning Using Federated Learning," *IEEE Transactions on Learning Technologies*, vol. 15, no. 2, pp. 235-246, 2023.

[14] Z. Chen et al., "Robust and Privacy-Preserving Decentralized Deep Federated Learning Training Focusing on Digital Healthcare Applications," *IEEE Access*, vol. 10, pp. 55400-55412, 2022.

[15] P. Liu et al., "Urban Parking Management through Federated Learning: A Privacy-Preserving Approach to Parking Slot Prediction," *IEEE Internet of Things Journal*, vol. 10, no. 3, pp. 2489-2501, 2023.

[16] W. Zhao et al., "U-shaped Split Federated Learning: An Efficient Cross-Device Learning Framework with Enhanced Privacy-preserving," *IEEE Internet of Things Journal*, vol. 10, no. 5, pp. 3784-3796, 2023.

[17] Y. Gao et al., "Wellness Detection Using Clustered Federated Learning," *IEEE Access*, vol. 10, pp. 134567-134577, 2022.