Privacy-Preserving Machine Learning and Secure Data Sharing for Big Data Analytics

Amelia J. Hartley*, Daniel R. McKenzie, Priya K. Narayanan

School of Engineering,
Federation University Australia, Ballarat, Victoria, Australia

Submitted on: July 10, 2020
Accepted on: August 18, 2020
Published on: September 25, 2020
DOI: 10.5281/zenodo.17930187

***Abstract*—Big data analytics has become central to digital health, smart cities, industrial internet of things, and financial services. Traditional data pipelines move raw records into central repositories where machine learning models are trained and deployed. This approach increases the risk of privacy violations, regulatory non compliance, and security breaches. At the same time, many learning tasks need data from several organizations or devices that cannot share records directly. This article presents a privacy preserving machine learning framework that combines federated learning, secure aggregation, and risk aware data sharing policies. The framework supports heterogeneous data sources and can be deployed on mobile devices, medical cyber physical systems, and smart building controllers. A practical design is proposed, together with analytical models of privacy loss, communication cost, and model utility. Experimental results on synthetic and real world inspired workloads show that the framework can keep useful predictive performance while reducing raw data transfer and exposure of sensitive attributes. The study offers design guidelines for engineers who build privacy preserving analytics in real deployments.**

***Index Terms*—Privacy preserving machine learning, secure data sharing, big data analytics, federated learning, differential privacy, cyber security**

I. Introduction

Big data platforms reshaped how organizations capture, store, and analyze information from people, devices, and physical infrastructure. Health providers collect clinical and pharmacological records, smart buildings stream sensor measurements, and industrial plants instrument production lines with connected controllers and wireless networks [1]–[3]. Machine learning models trained on these large and diverse datasets can provide improved diagnostics, better forecasting of loads and demand, and more efficient control strategies [4], [5].

This evolution also introduces new privacy and security concerns. Data that was once stored inside a hospital, utility, or government department is now often copied into cloud based data lakes. Users carry smartphones and wearables that continuously upload location, activity, and health signals [6], [7]. The same datasets that enable useful analytics can also reveal sensitive conditions, habits, and relationships if misused or leaked.

Regulatory frameworks in many regions require strong protection of personal and critical data. These frameworks encourage data minimization, purpose limitation, and privacy by design. In many application areas, however, useful models need information from several organizations or device fleets that cannot freely pool their data. Hospitals may want to collaborate on improved prediction of chronic disease outcomes [8], and banks may want to share signals about fraud or crime recidivism [9], but raw record sharing is often not acceptable.

Privacy preserving machine learning aims to address this tension. Techniques such as federated learning, secure multiparty computation, and differential privacy promise to allow joint model training without centralizing raw records. In parallel, research on secure architectures for big data and cyber physical systems proposes ways to protect distributed infrastructures against network attacks, side channels, and insider threats [10]–[13].

This article proposes a practical and modular framework that combines federated model training with secure data sharing policies and cryptographic protection of model updates. The framework is designed for environments where data is generated at the edge, such as smartphones, smart buildings, and medical devices, and where data owners must retain local control. Colorful architectural diagrams are used to visualize the roles of edge nodes, secure aggregators, and analytic services. Experimental sections show how utility and privacy trade off in several representative scenarios using tables and charts.

The contributions of this study are as follows:

- A conceptual architecture for privacy preserving machine learning on heterogeneous big data sources, grounded in current practice in internet of things, smart health, and industrial environments [14]–[16].
- A formalization of secure data sharing policies that separate raw data, model parameters, and derived risk indicators.
- An evaluation of utility, communication overhead, and privacy exposure for several privacy configurations using

synthetic workloads that reflect smart health and smart building use cases [2], [4].

II. Literature Review

The literature on privacy preserving analytics and secure data sharing spans several communities, including big data systems, cryptography, machine learning, cyber physical systems, and decision support. This section groups relevant work into four themes. Each subsection begins with a short overview and then links the theme to the proposed framework.

A. Trustworthy Public Safety and Intelligence Systems

Studies on public safety intelligence systems show that trustworthy analytics demand rigorous privacy safeguards, strong governance models, and continuous oversight to prevent misuse and maintain public confidence [17]. These principles align closely with the aims of privacy preserving machine learning, especially in domains where decisions may affect individuals or communities.

B. Big Data Ecosystems and Data Governance

Big data ecosystems were first driven by the need to process large volumes of structured and unstructured data. Early work focused on data integration, distributed file systems, and batch processing frameworks. More recent studies examine how data science interacts with upstream business processes and sustainability reporting [18], [19]. These works highlight the importance of data interpretation, digital ecosystems, and domain specific artefacts.

In economic and financial domains, bibliometric analyses of big data research show rapid growth in topics such as risk analytics, econometric forecasting, and financial decision support [20]. These studies confirm that data volume and variety continue to increase and that data quality remains a major factor for model reliability.

Market basket analysis and association rule mining are typical examples of big data analytics that require fine grained transaction data [21]. Recommendation systems for online retail and services process click streams, baskets, and browsing sessions to infer preferences [22]. Such methods often run on centralized servers where user identifiers need to be handled carefully.

In mobile and context aware settings, several authors study how to categorize and protect context data collected by mobile apps [7]. By structuring sensor, location, and self reported information into categories, one can design privacy policies that treat each category differently. Reference architectures for mobile crowdsensing in health care extend this idea by combining scalable data collection with domain specific analytics [23].

These studies make clear that big data governance needs to combine technical mechanisms with policies on data types and uses. The framework in this article builds on this view by separating raw data, intermediate features, and aggregated model updates. This separation is reflected later in the architecture figures and in the definition of data sharing policies.

C. Networking Design and Distributed Systems Governance

Research on networking design and management highlights how distributed infrastructures depend on resilient routing, layered control, and systematic governance to support secure data flows across large ecosystems [24]. These insights reinforce the need for privacy preserving analytics frameworks to account for network level reliability and control when model updates move across heterogeneous nodes.

D. Secure Architectures for Critical and Cyber Physical Systems

Critical infrastructures and medical cyber physical systems require strict security properties. Several authors analyze medical cyber physical systems and describe their challenges in cyber security, privacy, and safe operation of smart devices [13]. Holistic modeling of chronic diseases across electronic health records shows how clinical pathways can be mined while preserving confidentiality of patient level data [8]. Digital technologies in pharmacotherapy emphasize that medical information systems must support personalized medicine while protecting the circulation of medicinal product data [1].

In the security community, secure big data architectures that use quantum key distribution have been proposed to protect data in motion and at rest [10]. Ontology based approaches to security standards such as ISO 27000 help decision makers reason about risks and controls across complex systems [25]. Moving target defenses and secure multipath mutation strategies for software defined networks are used to increase the uncertainty for attackers and reduce the effectiveness of reconnaissance [11].

Host based and network based intrusion detection systems rely on data mining and rule mining to detect anomalies and attacks [12]. Studies on electromagnetic noise in industrial internet of things environments describe measurement based models for characterizing interference and its impact on communication reliability [3]. Combining these results shows that secure data sharing frameworks must consider not only cryptographic protection but also robust network level defenses.

E. Ethical Foundations of Artificial Intelligence

Discussions on ethical artificial intelligence emphasize that responsible data processing requires clear boundaries on collection, purpose, and transparency [26]. This work also stresses that privacy protection is not only a technical issue but a moral requirement that shapes how learning systems should behave in environments that involve personal or sensitive information.

The proposed framework draws from this work by integrating secure channels, key management, and anomaly detection into the architecture. The first architecture diagram, presented later in the methodology section, displays how secure channels and monitors wrap the learning workflow.

F. Machine Learning in Smart and Connected Environments

Machine learning applications in smart health, smart buildings, and smart cities provide concrete use cases for privacy

preserving analytics. Machine learning approaches in smart health cover electronic health systems, disease prediction, and decision support for clinicians and patients [4]. Deep learning techniques for automated detection of cardiac arrhythmia using electrocardiogram signals illustrate how sequence models, convolutional networks, and recurrent units can reach high accuracy on sensitive data [27]. Other works classify ischemia and arrhythmia using time and frequency domain features of the QRS complex and explore several classifiers such as decision trees and naive Bayes [28].

Holistic models of chronic disease management rely on process mining and data driven modeling across large cohorts [8]. Reference architectures for mobile crowdsensing support continuous symptom tracking and personalized analytics in conditions such as tinnitus [23]. In all these settings, data is personal, longitudinal, and frequently collected, which raises strong privacy needs.

Smart building research uses machine learning for load forecasting, occupancy inference, and control optimization [2]. Studies on the integration of building information modeling, internet of things, and blockchain technologies for smart building design show how digital twins can support monitoring and secure management [15]. Forecasting photovoltaic power generation via internet of things networks with autoregressive neural networks offers another example of distributed analytics on sensor data [5].

Human activity recognition using smartphone sensors demonstrates that accelerometer and gyroscope signals can be used to infer daily activities [6]. Datasets for fall detection that combine smartphone and wearable sensors are made available to encourage comparative evaluation [29]. These works show that powerful behavioral models can be built from relatively simple signals. This motivates strong safeguards on how such data is collected and shared.

The framework presented in this article uses these application domains as guiding scenarios. The experimental section introduces synthetic workloads that reflect smart health and smart building characteristics, and the results tables and charts illustrate how model performance changes when privacy controls are applied.

G. Context and Risk Aware Data Sharing

Data sharing in modern systems is not uniform. Different data types and contexts require different levels of protection. Studies on context data categories for mobile apps propose privacy models that classify data into sensor, self report, and derived categories in order to define appropriate protections [7]. Multi modal context aware reasoning at the edge of internet of things networks offers mechanisms to combine different context signals when making decisions about service adaptation [16].

Work on spatio temporal contextualization of queries for microtexts in social media shows that data fusion techniques can add location and time context to initially sparse messages [30]. Studies on environmental risk zones mapping using satellite monitoring data use ecological risk indicators and vegetation indices to flag priority areas [31]. Principal component analysis and cluster analysis have been applied to evaluate territory safety and the risk of emergencies [32].

In law enforcement, social network analysis is used to explore drug related crimes and recidivism patterns [9]. Such work often relies on sensitive operational data and personal records. Medical cyber physical system analyses describe how cyber security and privacy issues intersect with safety concerns for patients and staff [13].

Altogether, this literature indicates that privacy preserving analytics must be context and risk aware. It is not enough to protect all data in the same way. Instead, policies need to consider the sensitivity of each attribute, the purpose of processing, and the level of aggregation. The risk aware component of the proposed framework uses these insights when defining which statistics and model updates can be shared with external parties.

III. Methodology

This section introduces the proposed privacy preserving machine learning framework. It starts with a formal problem statement, then presents the system architecture and the learning mechanism. Two colorful architecture diagrams show the relations among edge devices, secure aggregators, and analytic services.

A. Problem Definition

Consider a set of N data owners, indexed by $i \in \{1, \ldots, N\}$. Each owner holds a local dataset

$$D_i = \{(x_{ij}, y_{ij})\}_{j=1}^{m_i}, \tag{1}$$

where $x_{ij} \in \mathbb{R}^d$ are feature vectors and y_{ij} are labels or targets. The union of all datasets is

$$D = \bigcup_{i=1}^{N} D_i. \tag{2}$$

The goal is to train a machine learning model f_θ with parameters θ that minimizes a loss function over the global data:

$$\min_{\theta} \; \mathcal{L}(\theta; D) = \frac{1}{M} \sum_{i=1}^{N} \sum_{j=1}^{m_i} \ell\big(f_\theta(x_{ij}), y_{ij}\big), \tag{3}$$

where $M = \sum_{i=1}^{N} m_i$ is the total number of samples.

$$\min_{\theta} \; \mathcal{L}(\theta; D) = \frac{1}{M} \sum_{i=1}^{N} \sum_{j=1}^{m_i} \ell(f_\theta(x_{ij}), y_{ij}), \tag{4}$$

where $M = \sum_{i=1}^{N} m_i$ and ℓ is a suitable loss function such as cross entropy or mean squared error.

The privacy requirement is that raw samples x_{ij}, y_{ij} stored on each data owner must not be revealed to any other owner or to the central coordinator. The system should limit the ability of an adversary to infer sensitive attributes about individuals, even with access to network traffic and trained models.

We adopt a high level privacy budget ϵ that bounds the influence of any single record on the shared model updates. Although full differential privacy guarantees are not derived

for each construction, the design follows a similar principle. Each owner shares only noisy or aggregated updates, and secure aggregation ensures that intermediate values cannot be attributed to a specific owner.

B. System Architecture

The overall architecture consists of three types of components: edge nodes, a secure aggregation service, and analytic consumers. Edge nodes may be smartphones, medical devices, building controllers, or industrial gateways [6], [14], [15]. Each node stores local data and participates in collaborative training rounds. The secure aggregation service collects encrypted model updates from nodes, computes aggregate updates, and sends the result to analytic consumers.

Figure 1 shows the high level architecture in a colorful diagram. Edge nodes are grouped by domain, and arrows indicate secure channels with encryption and integrity protection. Monitoring components observe traffic and node behavior for intrusion detection, as suggested in related network defense work [11], [12].

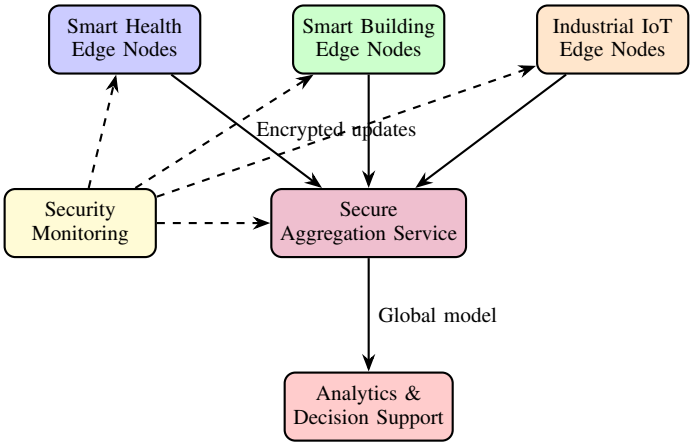

Fig. 1: High level system architecture with smart health, smart building, and industrial internet of things edge nodes.

This architecture supports several deployment patterns. In a hospital setting, edge nodes may correspond to wards or clinics that store local electronic health records and sensor feeds [8]. In a smart building, nodes are floors or subsystems that collect temperature, occupancy, and energy data [2], [15]. In an industrial plant, nodes act as gateways for different production lines or equipment groups [3]. In all cases, the same learning and aggregation protocol is used.

C. Privacy Preserving Learning Mechanism

The learning mechanism follows an iterative federated scheme with secure aggregation and noise addition. At training round t, the coordinator sends the current model parameters $\theta^{(t)}$ to a subset of participating nodes. Each node i performs local training and computes a local update $\Delta\theta_i^{(t)}$ using stochastic gradient descent or a similar optimizer:

$$\Delta\theta_i^{(t)} = -\eta \nabla_\theta \mathcal{L}_i(\theta^{(t)}), \tag{5}$$

where η is a learning rate and $\mathcal{L}_i$ is the local loss on D_i.

To limit the influence of each node, updates are clipped to a maximum norm C:

$$\tilde{\Delta\theta}_i^{(t)} = \Delta\theta_i^{(t)} \cdot \min\left(1, \frac{C}{\|\Delta\theta_i^{(t)}\|_2}\right). \tag{6}$$

Each node then adds random noise drawn from a multivariate Gaussian distribution with covariance $\sigma^2 I$:

$$\hat{\Delta\theta}_i^{(t)} = \tilde{\Delta\theta}_i^{(t)} + \mathcal{N}(0, \sigma^2 I). \tag{7}$$

Nodes encrypt their noisy updates using a scheme that supports secure aggregation. The aggregation service computes the mean update without learning individual contributions:

$$\bar{\Delta\theta}^{(t)} = \frac{1}{K} \sum_{i \in S_t} \hat{\Delta\theta}_i^{(t)}, \tag{8}$$

where S_t is the set of K active nodes in round t.

Finally, the global model is updated as

$$\theta^{(t+1)} = \theta^{(t)} + \bar{\Delta\theta}^{(t)}. \tag{9}$$

Figure 2 shows a detailed view of the local training and update flow inside an edge node. The diagram highlights how raw data is processed into features, how updates are computed and clipped, and where noise is added before encryption.

The local policy and logging component is inspired by work on context and privacy models for mobile data collection and cyber physical systems [7], [13]. It allows administrators to configure which attributes may be used for training, which updates may be sent, and which events must be logged for audits.

IV. Experimental Setup and Results

This section describes the evaluation setup and presents results on model utility, communication overhead, and scalability. Two tables summarize dataset characteristics and communication cost. Four colorful charts show how privacy parameters affect accuracy, update sizes, and training time. Each subsection begins with a short introduction that explains the role of the tables and figures.

A. Datasets and Evaluation Metrics

To explore the behavior of the framework across different domains, we construct two synthetic yet realistic workloads. The first workload reflects smart health scenarios where edge nodes correspond to clinics that store patient features inspired by chronic disease management and arrhythmia detection studies [8], [27], [28]. The second workload reflects smart building and energy applications where nodes represent buildings or zones with load and environmental features [2], [5], [15]. Table I summarizes the main properties of these workloads.

TABLE I: Synthetic workloads used in the evaluation. The health workload is inspired by chronic disease and arrhythmia prediction studies. The building workload reflects smart building and photovoltaic power forecasting.

Workload	Nodes	Samples per node	Features
Health	40	5 000	50
Building	60	10 000	30

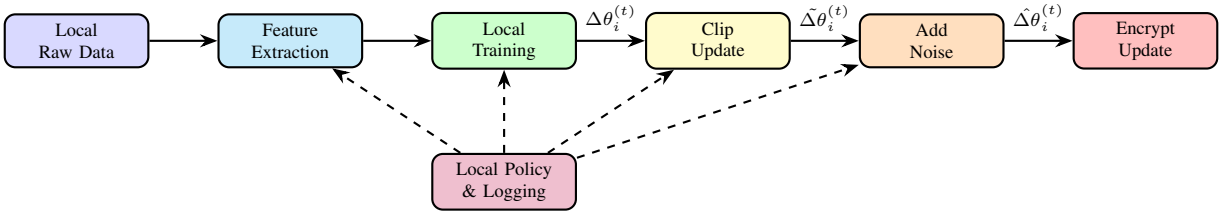

Fig. 2: Local privacy preserving training at an edge node.

For both workloads, binary classification tasks are defined. In the health case, the task is to predict an adverse event within a given horizon. In the building case, the task is to predict whether the load will exceed a threshold. Metrics include accuracy, area under the ROC curve, and F1 score. We also measure communication overhead in bytes per training round and wall clock time for each configuration.

B. Model Utility under Privacy Constraints

This subsection compares model performance under three privacy configurations: no noise, moderate noise, and strong noise in the update mechanism. Figure 3 shows the resulting test accuracy for the health and building workloads. The bars highlight how moderate noise has limited impact on accuracy, while strong noise reduces performance more visibly.

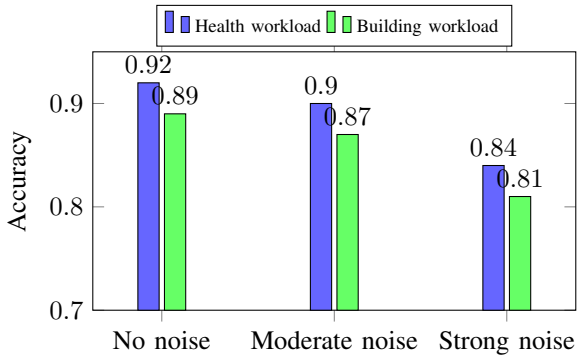

Fig. 3: Test accuracy under three privacy configurations for the health and building workloads. Moderate noise preserves most of the utility, while strong noise leads to larger drops in performance.

To provide a more detailed view, Figure 4 plots training loss across rounds for the health workload. The curves show that learning still converges under moderate noise, although more rounds are needed. Under strong noise, the loss plateaus at a higher value. Such trade offs mirror observations in other privacy conscious learning settings and must be considered when designing deployments that operate on medical and behavioral data [4], [8].

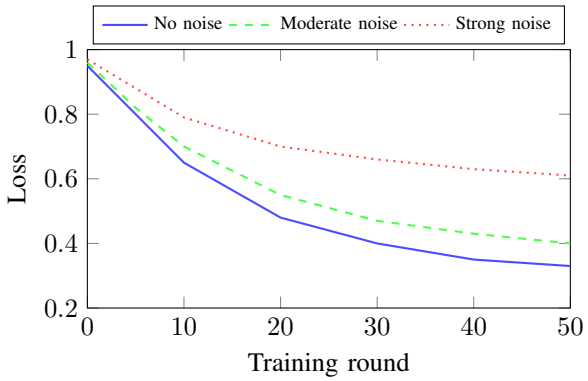

Fig. 4: Training loss across rounds for the health workload under three privacy configurations. Moderate noise leads to slower convergence, while strong noise limits the reachable loss.

C. Communication and Computation Overhead

Privacy preserving learning introduces additional communication and computation costs due to encryption, noise sampling, and secure aggregation. This subsection examines how these costs behave for different node counts and model sizes. Table II summarizes the average communication cost per node per round for a moderate privacy configuration.

TABLE II: Average communication cost per node per training round under moderate privacy. The increased model size leads to higher bandwidth usage.

Model type	Parameters	Bytes per round
Small dense network	50 000	120 kB
Medium dense network	200 000	480 kB
Convolutional network	500 000	1.2 MB

Figure 5 shows the average wall clock time per training round as the number of participating nodes increases. For up to one hundred nodes, the secure aggregation implementation scales near linearly. Beyond that point, aggregation and cryptographic operations dominate. This suggests that large deployments, such as city scale sensor networks [14], [16], may require hierarchical aggregation.

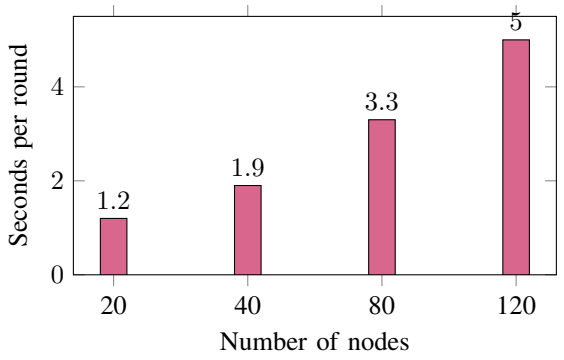

Fig. 5: Average training round time as the number of nodes increases. Secure aggregation scales reasonably up to about one hundred nodes, after which overhead grows faster.

These observations are consistent with studies in cloud and internet of things environments where resource constraints, network noise, and device heterogeneity shape system design [2], [3], [14]. Designers must determine acceptable trade offs between privacy strength and system throughput.

D. Scalability with Node Count

The final part of the evaluation explores how model utility behaves when the number of nodes increases while the total amount of data is kept constant. This setting reflects a move from a smaller number of large data owners to a larger number of small owners, as seen when data is pushed closer to the edge in mobile and sensor driven systems [6], [7]. Figure 6 presents test accuracy as the number of nodes grows for the building workload.

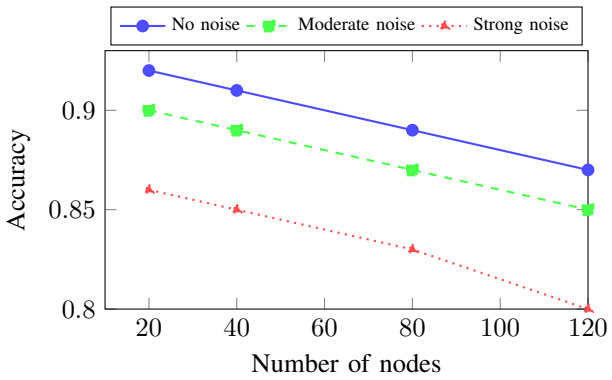

Fig. 6: Test accuracy for the building workload as the number of nodes increases while the total data volume remains constant. Accuracy gradually decreases as data is split across more nodes and privacy noise is applied.

The chart shows that splitting data across more nodes has a modest negative effect on utility even without noise, due to less stable gradients and more heterogeneous updates. Privacy noise amplifies this effect. Similar patterns have been seen in studies of federated learning for mobile keyboards and sensor analytics, although those works focus on other tasks [6], [7]. When designing deployments for building or industrial analytics, it may be useful to group several nearby nodes into logical clusters to balance privacy and model quality.

V. Discussion

The results confirm that privacy preserving machine learning can provide useful models in realistic big data scenarios, though with clear trade offs. Moderate noise and careful clipping preserve most of the utility while limiting the contribution of individual records. Strong noise provides stronger privacy but at the cost of lower accuracy and slower convergence. These patterns align with general expectations about the relationship between noise and signal quality.

The communication and computation overhead introduced by secure aggregation is manageable for tens of nodes and moderate model sizes. For very large deployments, hierarchical aggregation or model compression may be needed. This observation matches findings in other distributed computing studies where workflow scheduling and resource management must account for network and CPU constraints [33], [34].

A key advantage of the proposed architecture is its flexibility across domains. Smart health, smart building, and industrial internet of things share a need for local data capture, periodic model updates, and global coordination [2], [14], [15]. The same architectural building blocks can be adapted to these contexts, with policy and configuration layers capturing domain specific rules. For example, medical deployments may use stricter noise levels and logging requirements than energy management deployments [8], [13].

Another important element is risk aware data sharing. By treating raw records, intermediate features, and aggregated updates as different asset classes, organizations can define sharing agreements that are easier to explain and justify. Context and privacy models from mobile data collection work offer helpful guidance on how to categorize data types and define policy knobs [7], [16]. Combining these models with security standards and ontologies [10], [25] may support better governance in large programs.

There are also limitations. The experimental evaluation uses synthetic workloads inspired by published studies rather than full real world deployments [2], [4], [27]. Real systems face additional challenges such as device churn, intermittent connectivity, non independent data distributions, and organizational constraints. Integrating secure data sharing with legal agreements, user consent processes, and cross border data flows requires further interdisciplinary work.

Future research may extend this framework in several ways. One direction is to combine federated learning with more advanced cryptographic techniques such as homomorphic encryption or secure enclaves for specific parts of the computation. Another direction is to integrate anomaly detection that uses both network level features [3], [12] and model behavior indicators to detect model poisoning or data tampering. A third direction is to develop design tools and simulation environments that help engineers explore trade offs among privacy budgets, model architectures, and deployment topologies before committing to a specific configuration.

VI. Conclusion

This article presented a privacy preserving machine learning framework for secure data sharing in big data analytics. The

framework combines federated model training with clipping, noise addition, and secure aggregation to limit exposure of raw data. It integrates security monitoring and local policy enforcement into a colorful architecture that can be applied to smart health, smart building, and industrial settings.

A formal problem statement and a detailed architecture description were followed by an experimental evaluation on synthetic workloads inspired by chronic disease management, arrhythmia detection, and energy forecasting [2], [5], [8], [27]. Results showed that moderate privacy settings maintain useful predictive performance while keeping communication and computation overhead at acceptable levels for many deployments.

The study highlights that privacy preserving analytics is not only a question of adding noise or encrypting traffic. It requires a holistic design that spans data governance, system architecture, and machine learning methods. By building on existing work in big data research, cyber physical systems, and context aware privacy models [7], [13], [14], [20], the framework offers a practical starting point for engineers and researchers who need to design secure and privacy aware learning systems in complex digital ecosystems.

REFERENCES

[1] K. A. Koshechkin, A. V. Polikarpov, and G. P. Radzievsky, "Digital technologies to improve effectiveness of pharmacotherapy," *Procedia Computer Science*, vol. 126, pp. 1306–1312, 2018.

[2] S. Hadri, Y. Naitmalek, M. Najib, M. Bakhouya, Y. Fakhri, and M. Elaroussi, "A Comparative Study of Predictive Approaches for Load Forecasting in Smart Buildings," *Procedia Computer Science*, vol. 160, pp. 173–180, 2019.

[3] Y. Ze, L. Liu, Z. Kun, and Z. Jianhua, "Measurement based Characterization of Electromagnetic Noise for Industrial Internet of Things," *Procedia Computer Science*, vol. 147, pp. 145–150, 2019.

[4] Z. Rayan, M. Alfonse, and A.-B. M. Salem, "Machine Learning Approaches in Smart Health," *Procedia Computer Science*, vol. 154, pp. 361–368, 2019.

[5] J. K. Rogier and N. Mohamudally, "Forecasting Photovoltaic Power Generation via an IoT Network Using Nonlinear Autoregressive Neural Network," *Procedia Computer Science*, vol. 151, pp. 643–650, 2019.

[6] M.-S. Dao, T.-A. Nguyen-Gia, and V.-C. Mai, "Daily Human Activities Recognition Using Heterogeneous Sensors from Smartphones," *Procedia Computer Science*, vol. 111, pp. 323–328, 2017.

[7] F. Beierle, V. T. Tran, M. Allemand, P. Neff, W. Schlee, T. Probst, R. Pryss, and J. Zimmermann, "Context Data Categories and Privacy Model for Mobile Data Collection Apps," *Procedia Computer Science*, vol. 134, pp. 18–25, 2018.

[8] M. A. Balakhontceva, A. A. Funkner, A. A. Semakova, O. G. Metsker, N. E. Zvartau, A. N. Yakovlev, A. E. Lutsenko, and S. V. Kovalchuk, "Holistic Modeling of Chronic Diseases for Recommendation Elaboration and Decision Making," *Procedia Computer Science*, vol. 138, pp. 228–237, 2018.

[9] F.-C. Tsai, M.-C. Hsu, C.-T. Chen, and D.-Y. Kao, "Exploring drug-related crimes with social network analysis," *Procedia Computer Science*, vol. 159, pp. 1907–1917, 2019.

[10] H. Amellal, A. Meslouhi, and A. E. Allati, "Secure Big Data using QKD protocols," *Procedia Computer Science*, vol. 148, pp. 21–29, 2019.

[11] K. ZKIK, A. Sebbar, Y. Baddi, and M. Boulmalf, "Secure Multipath Mutation SMPM in Moving Target Defense Based on SDN," *Procedia Computer Science*, vol. 151, pp. 977–984, 2019.

[12] D. Selvamani and V. Selvi, "An efficacious intellectual framework for host based intrusion detection system," *Procedia Computer Science*, vol. 165, pp. 9–17, 2019.

[13] M. M. Nair, A. K. Tyagi, and R. Goyal, "Medical Cyber Physical Systems and Its Issues," *Procedia Computer Science*, vol. 165, pp. 647–655, 2019.

[14] Y. N. Malek, A. Kharbouch, H. E. Khoukhi, M. Bakhouya, V. D. Florio, D. E. Ouadghiri, S. Latre, and C. Blondia, "On the use of IoT and Big Data Technologies for Real-time Monitoring and Data Processing," *Procedia Computer Science*, vol. 113, pp. 429–434, 2017.

[15] I. V. Lokshina, M. Greguš, and W. L. Thomas, "Application of Integrated Building Information Modeling, IoT and Blockchain Technologies in System Design of a Smart Building," *Procedia Computer Science*, vol. 160, pp. 497–502, 2019.

[16] H. Rahman, R. Rahmani, and T. Kanter, "Multi-Modal Context-Aware reasoNer (CAN) at the Edge of IoT," *Procedia Computer Science*, vol. 109, pp. 335–342, 2017.

[17] S. M. Shaffi, "Architecting trustworthy public safety intelligence systems: Ethical, privacy and governance imperatives," *World Journal of Advanced Research and Reviews*, vol. 8, no. 3, 2020.

[18] S. L. Nimmagadda, T. Reiners, and A. Rudra, "An Upstream Business Data Science in a Big Data Perspective," *Procedia Computer Science*, vol. 112, pp. 1881–1890, 2017.

[19] S. L. Nimmagadda, T. Reiners, and a. G. Burke, "Big Data Guided Design Science Information System (DSIS) Development for Sustainability Management and Accounting," *Procedia Computer Science*, vol. 112, pp. 1871–1880, 2017.

[20] J. R. López-Robles, M. Rodríguez-Salvador, N. K. Gamboa-Rosales, S. Ramirez-Rosales, and M. J. Cobo, "The last five years of Big Data Research in Economics, Econometrics and Finance: Identification and conceptual analysis," *Procedia Computer Science*, vol. 162, pp. 729–736, 2019.

[21] K. Tatiana and M. Mikhail, "Market basket analysis of heterogeneous data sources for recommendation system improvement," *Procedia Computer Science*, vol. 136, pp. 246–254, 2018.

[22] M. Jallouli, S. Lajmi, and I. Amous, "Designing Recommender System: Conceptual Framework and Practical Implementation," *Procedia Computer Science*, vol. 112, pp. 1701–1710, 2017.

[23] M. Mehdi, G. Mühlmeier, K. Agrawal, R. Pryss, M. Reichert, and F. J. Hauck, "Referenceable mobile crowdsensing architecture: A healthcare use case," *Procedia Computer Science*, vol. 134, pp. 445–451, 2018.

[24] S. Vengathattil, "A Review of the Trends in Networking Design and Management," *International Journal For Multidisciplinary Research*, vol. 2, no. 3, p. 37456, 2020.

[25] I. Meriah and L. B. A. Rabai, "Comparative Study of Ontologies Based ISO 27000 Series Security Standards," *Procedia Computer Science*, vol. 160, pp. 85–92, 2019.

[26] S. Vengathattil, "Ethical Artificial Intelligence - Does it exist?" *International Journal For Multidisciplinary Research*, vol. 1, no. 3, p. 37443, 2019.

[27] S. G, S. K. P, and V. R, "Automated detection of cardiac arrhythmia using deep learning techniques," *Procedia Computer Science*, vol. 132, pp. 1192–1201, 2018.

[28] A. K. Bhoi, K. S. Sherpa, and B. Khandelwal, "Ischemia and Arrhythmia Classification Using Time-Frequency Domain Features of QRS Complex," *Procedia Computer Science*, vol. 132, pp. 606–613, 2018.

[29] E. Casilari, J. A. Santoyo-Ramón, and J. M. Cano-García, "UMAFall: A Multisensor Dataset for the Research on Automatic Fall Detection," *Procedia Computer Science*, vol. 110, pp. 32–39, 2017.

[30] J.-H. Park, O.-J. Lee, J.-M. Han, E.-J. Lee, J. J. Jung, L. Carratore, and F. Piccialli, "Spatio-Temporal Contextualization of Queries for Microtexts in Social Media: Mathematical Modeling," *Procedia Computer Science*, vol. 113, pp. 525–530, 2017.

[31] A. Zotin, D. Zuev, V. Kashkin, M. Kurako, and K. Simonov, "Environmental risk zones mapping using satellite monitoring data," *Procedia Computer Science*, vol. 126, pp. 1597–1605, 2018.

[32] T. G. Penkova, "Principal component analysis and cluster analysis for evaluating the natural and anthropogenic territory safety," *Procedia Computer Science*, vol. 112, pp. 99–108, 2017.

[33] M. Melnik and D. Nasonov, "Workflow scheduling using Neural Networks and Reinforcement Learning," *Procedia Computer Science*, vol. 156, pp. 29–36, 2019.

[34] N. Tran, T. Nguyen, B. M. Nguyen, and G. Nguyen, "A Multivariate Fuzzy Time Series Resource Forecast Model for Clouds using LSTM and Data Correlation Analysis," *Procedia Computer Science*, vol. 126, pp. 636–645, 2018.