Pharmacophore

ISSN-2229-5402

Journal home page: http://www.pharmacophorejournal.com

FEDERATED LEARNING FRAMEWORK FOR ADVERSE DRUG REACTION PREDICTION FROM MULTI-INSTITUTIONAL HEALTH RECORDS

Ricardo Alves[1], Bruno Costa[1]*, Helena Martins[2], Nuno Faria[1]

1. ***Department of AI Drug Analytics, Faculty of Pharmacy, University of Porto, Porto, Portugal.***
2. ***Department of Computational Pharmaceutical Systems, Faculty of Medicine, University of Aveiro, Aveiro, Portugal.***

ARTICLE INFO

Received:
16 October 2025
Received in revised form:
12 January 2026
Accepted:
14 January 2026
Available online:
28 February 2026

Keywords: Federated learning, Pharmacovigilance, Adverse drug reactions, Electronic health records, Privacy-preserving artificial intelligence, Differential privacy

ABSTRACT

Adverse drug reactions are often under-detected when analyses are limited to single institutions, and centralized pooling of electronic health records is frequently constrained by privacy regulations, institutional governance, and technical barriers to data transfer. Current multi-institutional pharmacovigilance typically relies on aggregate summaries rather than collaborative machine learning over distributed individual-level records, leaving complex temporal, clinical, and medication-related patterns difficult to detect across health systems. To address these challenges, this article proposes a federated learning framework that enables hospitals to jointly train adverse drug reaction prediction models using their own electronic health records, keeping patient-level data within each institution while exchanging only model updates through a controlled workflow. The framework incorporates a local model trainer at each hospital, a secure aggregation server, a differential privacy module, and a global model distribution layer, supporting structured electronic health record data, clinical text features, and interoperable pharmacovigilance definitions. By facilitating learning from more diverse clinical populations while preserving institutional data sovereignty, this federated approach could enhance adverse drug reaction prediction, promote earlier safety signal detection, and enable more reliable risk–benefit assessments in routine care, contingent on careful attention to privacy safeguards, data harmonization, governance, and workflow integration.

To Cite This Article: Alves R, Costa B, Martins H, Faria N. Federated Learning Framework for Adverse Drug Reaction Prediction from Multi-Institutional Health Records. Pharmacophore. 2026;17(1):1-11. https://doi.org/10.51847/f2RnAaSwj2

Introduction

Adverse drug reactions remain a major concern for patient safety because they may be rare, delayed, multifactorial, or difficult to distinguish from the underlying illness in routine care. Electronic health records offer a rich source for identifying medication exposures, laboratory abnormalities, diagnoses, and clinical narratives, yet single-site models are often limited by local prescribing practices, documentation habits, and the scarcity of confirmed adverse events. Reviews of electronic health record-based adverse drug event prediction emphasize that model development remains highly dependent on data quality, endpoint definition, and clinical context. Recent systematic work on machine learning for adverse drug events similarly indicates that prediction models must account for heterogeneous data sources and clinical workflows before they can support practical safety surveillance.

Multi-institutional drug safety initiatives have historically relied on distributed observational networks, standardized analytics, and site-level summary outputs rather than fully collaborative machine learning over individual-level trajectories. The Observational Health Data Sciences and Informatics ecosystem has shown how common data models can support scalable patient-level prediction while preserving local control of observational health data [1]. Drug safety signal identification from electronic health records has also been described as a distributed and methodologically diverse task, with substantial variation in data representation and outcome ascertainment across participating sites [2]. These approaches provide an important foundation, but they do not by themselves solve the problem of jointly learning complex predictive representations from multi-site clinical records.

Corresponding Author: Bruno Costa; Department of AI Drug Analytics, Faculty of Pharmacy, University of Porto, Porto, Portugal. E-mail: bruno.costa@gmail.com.

Federated learning offers a technical paradigm in which institutions train local models and contribute model updates rather than patient-level records. Early work on predictive modeling from federated electronic health records demonstrated the feasibility of distributed learning for clinical prediction without centralizing sensitive data [3]. Subsequent healthcare applications, including federated prediction of clinical outcomes in hospitalized patients, suggest that collaborative modeling can be organized across institutions while retaining local data control [4, 5]. However, federated learning has not yet become a routine pharmacovigilance tool, particularly for adverse drug reaction prediction from structured codes, laboratory signals, medication histories, and clinical notes.

This article proposes an artificial intelligence systems framework for federated adverse drug reaction prediction across healthcare institutions. The framework is informed by general federated learning principles in healthcare informatics [6], practical multi-institutional medical federation designs [7], and privacy-preserving machine learning methods that can strengthen confidentiality beyond simple data locality [8]. It is conceptual rather than experimental, so it does not report performance results, dataset sizes, or implementation benchmarks. The central thesis is that a purpose-built federated pharmacovigilance framework could balance collaborative learning, privacy protection, regulatory accountability, and deployment feasibility.

Background

Adverse Drug Reaction Detection in Electronic Health Records

Adverse drug reaction detection in electronic health records has used spontaneous reports, rule-based clinical triggers, structured diagnosis and medication codes, laboratory abnormalities, and supervised machine learning. Natural language processing has expanded this evidence base by extracting medication events and adverse outcomes from free-text clinical documentation [9, 10]. Studies of adverse drug event and medication extraction show that deep learning models can identify relevant entities and relations in clinical text, although their behavior depends on annotation quality and local documentation conventions [11, 12]. Because single-site adverse reaction models can reflect local coding, prescribing, and monitoring practices, a federated design would aim to preserve local data control while learning from broader clinical variation.

Federated Learning in Healthcare

Federated learning in healthcare is based on the idea that each institution trains a model locally and shares only model parameters, gradients, or other update summaries for aggregation. Foundational healthcare reviews describe federated learning as a way to support collaborative clinical prediction while reducing the need for direct data sharing [6, 13]. Multi-institutional demonstrations in medicine have further shown that distributed model training can be organized across sites with different infrastructure and clinical populations [5, 7]. For adverse drug reaction prediction, the same paradigm could allow hospitals to contribute evidence about rare medication-related harms without exporting identifiable patient records.

Table 1 shows key aspects of federated learning applications in healthcare, including the workflow, benefits, and illustrative use cases.

Table 1. Federated Learning in Healthcare

Aspect	Description	Example / Use Case
Local Model Training	Each institution trains a model on its own data without sharing raw patient information	Hospital A trains a prediction model for patient readmission using its EHR data
Shared Updates	Only model parameters, gradients, or update summaries are shared for aggregation	Hospitals share gradients to update a global model
Collaborative Clinical Prediction	Enables joint model development while minimizing direct data sharing	Multi-hospital risk prediction for sepsis or COVID-19 outcomes
Multi-Institutional Implementation	Distributed training can operate across sites with varying infrastructure and populations	Hospitals with different EHR systems and patient demographics collaborate on adverse drug reaction prediction
Privacy Preservation	Protects patient identity by avoiding transfer of raw data	Hospitals contribute rare adverse drug reaction data without exporting patient records

Privacy-Preserving Techniques beyond Federation

Federation alone does not guarantee privacy, because model updates may still leak information under some threat models. Secure aggregation protocols were developed to allow a server to learn only the aggregate of client updates rather than each individual client contribution [14]. Hybrid privacy-preserving federated learning approaches combine distributed training with additional safeguards such as differential privacy and cryptographic protection [15]. In clinical settings, these techniques would complement governance controls by reducing the risk that patient-specific information could be inferred from model updates [8].

Heterogeneity across Hospital Data

Hospital data are heterogeneous because institutions differ in patient demographics, formularies, laboratory measurement practices, diagnosis coding, note-writing conventions, and adverse reaction documentation. Federated learning in electronic

health records must therefore address non-identically distributed data rather than assuming that each site is a random sample from the same population [3]. Clustered federated learning and related multitask approaches have been proposed to handle variation among clients while preserving the benefits of shared learning [16]. In pharmacovigilance, such heterogeneity may be especially important because adverse reaction risk can depend on local prescribing patterns, comorbidity profiles, and monitoring intensity.

Prior Multi-Institutional Pharmacovigilance Frameworks

Prior multi-institutional pharmacovigilance frameworks provide important infrastructure for distributed evidence generation, but they are not always designed for iterative machine learning across sites. Standardized patient-level prediction pipelines in observational health data illustrate how common data models can support reproducible model development and validation [1]. Scoping reviews of electronic health record data for drug safety signal identification show growing interest in multi-source surveillance, but also highlight persistent challenges in outcome definition, data completeness, and method comparability [2]. A federated learning framework could build on these distributed data network principles while adding collaborative model training for adverse drug reaction prediction.

Figure 1 illustrates how prior multi-institutional pharmacovigilance networks support distributed evidence generation, while a federated learning layer enables iterative adverse drug reaction prediction across sites without centralizing patient-level data.

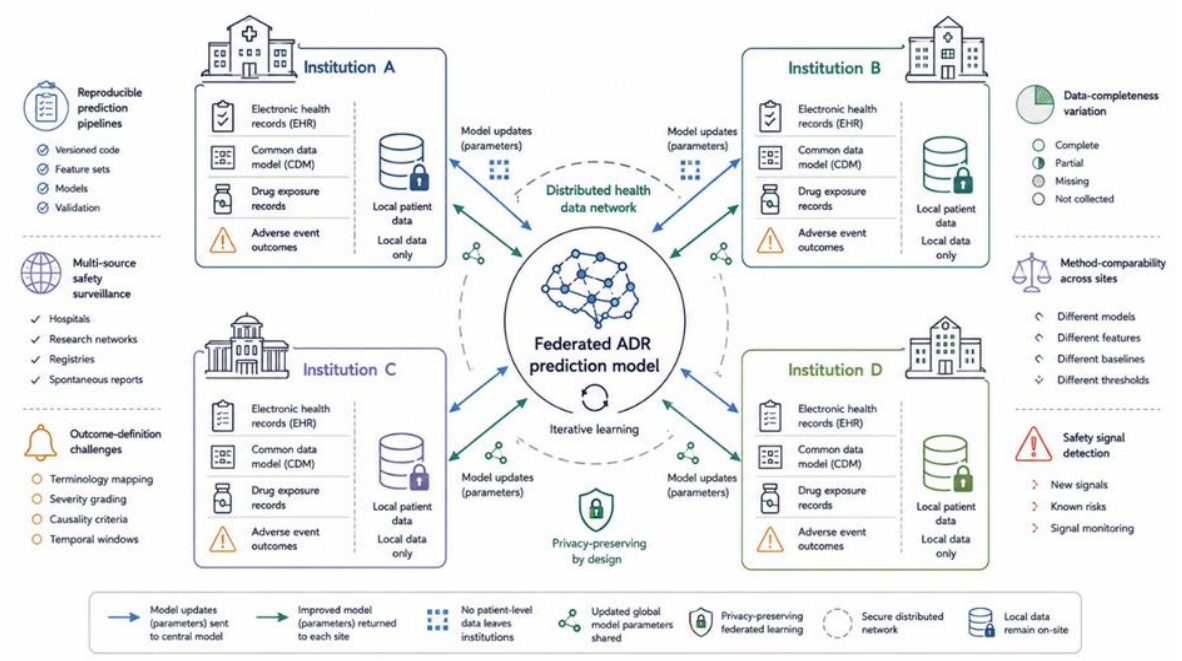

Figure 1. From Distributed Pharmacovigilance Infrastructure to Iterative Federated ADR Prediction

Framework Architecture Overview

High-Level Design

The proposed framework consists of hospital-local compute nodes, a coordinating aggregation server, privacy modules, and a model distribution layer. Each hospital keeps its electronic health record data behind its firewall, trains a local copy of the adverse drug reaction prediction model, and transmits only protected model updates to the aggregator. The aggregator combines these updates into a global model and redistributes the updated model for the next local training cycle, following the broad design logic used in federated clinical prediction systems [3, 4]. This architecture would allow local inference within each hospital while supporting shared learning across the network.

Figure 2 illustrates the proposed federated learning architecture in which hospitals retain patient-level electronic health records locally while exchanging privacy-protected model updates for shared adverse drug reaction prediction and locally governed pharmacovigilance deployment.

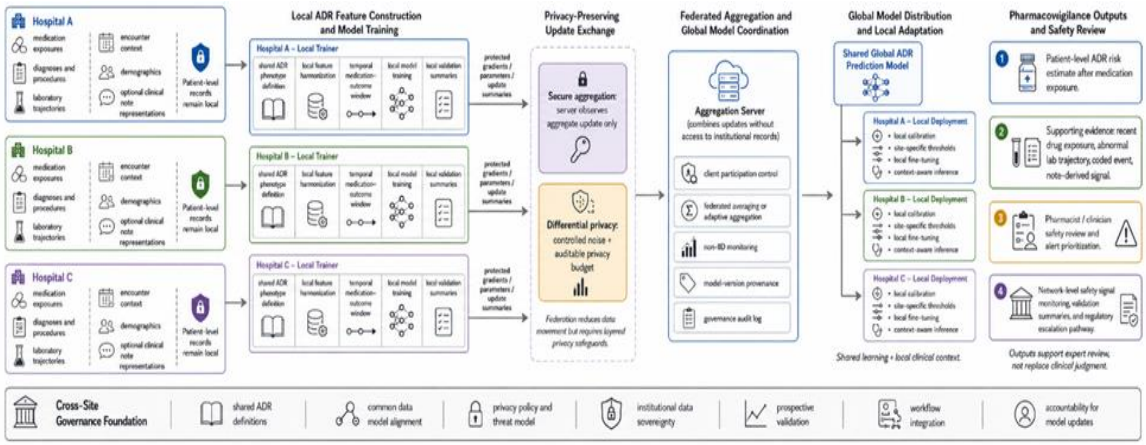

Figure 2. Federated Learning Framework for Adverse Drug Reaction Prediction from Multi-Institutional Health Records

Core Input Data and Prediction Task

The core input data would include demographics, medication exposures, diagnoses, procedures, laboratory results, encounter context, and, where appropriate, clinical note representations. Natural language processing methods for medication and adverse event extraction could support optional text-derived features when clinical notes are available and locally permitted for modeling [9, 17]. The prediction task would be defined as estimating the likelihood of an adverse drug reaction after drug exposure within a clinically meaningful time window, with endpoints such as kidney injury, hepatotoxicity, severe hypersensitivity, or other coded and clinically adjudicated reactions. Pretrained patient trajectory representations based on common data model electronic health records suggest one possible direction for representing longitudinal medication and outcome sequences in such a framework.

Design Principles

The framework is guided by data-never-leaves-the-site operation, privacy-by-design, modular deployment, interoperability with common clinical data models, and alignment with pharmacovigilance governance expectations. Secure and privacy-preserving medical machine learning literature emphasizes that technical architecture must be matched with institutional trust, auditability, and clearly specified data access boundaries [8]. The framework should therefore separate local feature construction, model training, aggregation, privacy accounting, and deployment governance into auditable components. These principles would make the system more compatible with distributed health networks and with hospitals that differ in technical maturity.

Table 2 maps the proposed federated pharmacovigilance framework from local electronic health record custody to privacy-protected model updating, global aggregation, local adaptation, and clinical safety deployment.

Table 2. Federated Pharmacovigilance Architecture: Component Logic, Data Boundaries, and ADR Prediction Function

Framework layer	Primary function in the proposed system	Information handled at this layer	Privacy or sovereignty boundary	Analytical contribution to ADR prediction	Implementation dependency
Local institutional EHR environment	Maintains patient-level medication, diagnosis, laboratory, encounter, demographic, and clinical-note data within each hospital	Raw or locally processed patient-level records	Data remain inside the institutional firewall and are not exported to the consortium	Captures local medication exposure, comorbidity context, laboratory change, and adverse reaction evidence	EHR access governance, data quality review, local compute capability
Shared ADR phenotype specification	Defines adverse reaction endpoints, exposure windows, candidate drug classes, temporal risk windows, and adjudication logic	Common endpoint rules and local phenotype mappings	Phenotype definitions are shared, but local records used to instantiate them remain private	Reduces cross-site endpoint drift and improves comparability of safety labels	Consortium agreement on ADR definitions and clinical validity
Local feature harmonization	Converts heterogeneous site data into a compatible modeling representation	Structured codes, laboratory trajectories, medication histories, visit context, optional text-derived features	Feature construction is performed locally under site-specific permissions	Enables collaborative training despite differences in coding, formulary, documentation, and monitoring practices	Common data model alignment and local terminology mapping

Local model trainer	Trains a site-specific copy of the ADR prediction model using local records	Model parameters, gradients, loss values, validation summaries	Only compatible update objects are prepared for transmission; patient-level examples remain local	Learns local medication-outcome associations and contributes them to the shared model	Sufficient local sample size, compute infrastructure, model compatibility
Privacy module	Applies differential privacy, update clipping, privacy accounting, or related protections before transmission	Protected model updates and privacy-budget metadata	Reduces risk of patient-level inference from model updates	Supports participation by institutions with strict confidentiality requirements	Locally approved privacy parameters and auditable privacy-budget tracking
Secure aggregation server	Combines protected local updates into an aggregate model update	Aggregated update summaries rather than individual patient records	Server should not inspect raw hospital data or unprotected individual-site updates	Produces a global model that reflects multi-institutional learning	Secure aggregation protocol, client authentication, participation management
Non-IID and heterogeneity monitor	Detects variation in site-level learning behavior, label prevalence, calibration, and feature distributions	Locally computed validation summaries and aggregate diagnostics	Site-level summaries may be shared under governance rules, not raw records	Identifies whether one global model is sufficient or whether clustered/personalized adaptation is needed	Standardized reporting metrics and governance-approved heterogeneity thresholds
Global model distribution layer	Redistributes updated global model versions to participating hospitals	Versioned model objects and documentation	Shared model travels to sites; local data do not travel to the central server	Provides a common predictive foundation trained from broader clinical variation	Model-version control, provenance documentation, deployment approval
Local adaptation and calibration layer	Tunes the global model to each hospital's population, monitoring practices, and workflow thresholds	Local calibration statistics, thresholds, and optional fine-tuning data	Adaptation remains institution-specific and does not expose local patient distributions	Improves local usability, calibration, and clinical trust	Local validation cohort, pharmacy safety review input, threshold governance
Pharmacovigilance workflow interface	Presents risk estimates and supporting evidence to pharmacists, clinicians, safety teams, or surveillance dashboards	Local risk scores, supporting features, alert rationale, review status	Patient-level outputs remain within the hospital's operational systems	Translates prediction into actionable safety review and signal detection	EHR/CDS integration, alert governance, prospective monitoring

Federated Training and Secure Aggregation

Local Model Architecture and Training

Each hospital would train a local model on its own adverse drug reaction dataset using a selected architecture such as gradient-boosted trees, a neural network, a temporal sequence model, or a text-enhanced model. Deep neural approaches to adverse drug reaction discovery from electronic health records show how longitudinal clinical features can be used for safety-related prediction tasks [18]. Multimodal adverse drug reaction classification models also indicate that combining text and drug representations may help capture complementary evidence, provided that all feature extraction remains local under federation [19]. The framework would not prescribe a single model family, but it would require each local trainer to produce compatible update objects for aggregation.

Federated Aggregation Algorithm

The aggregation server would combine protected local updates into a shared global model using federated averaging, secure multi-party computation, or a related privacy-preserving aggregation strategy. Practical secure aggregation protocols allow client updates to be summed without exposing individual client contributions to the server [16]. General federated learning theory identifies aggregation design as a central issue for convergence, robustness, and privacy in distributed optimization [20]. In this pharmacovigilance framework, aggregation would be configured so that no hospital can inspect another hospital's patient-level data or unprotected model update.

Communication Efficiency and Participation

Communication efficiency is important because hospitals may differ in network capacity, information technology resources, and tolerance for computational burden. Federated healthcare systems have been discussed as practical collaborations that

must minimize data movement while still enabling meaningful shared model updates [7, 13]. Model compression, selective update transmission, gradient sparsification, and adaptive client participation could be considered as implementation strategies, although they should be evaluated carefully for their effects on safety-relevant model behavior. Participation rules would also need to account for hospital downtime, local governance review, and the possibility that some institutions contribute intermittently rather than continuously.

Privacy-Preserving Mechanisms and Differential Privacy

Integrated Differential Privacy

The framework would incorporate differential privacy by allowing each hospital to perturb its model updates before transmission according to a locally approved privacy policy. Hybrid privacy-preserving federated learning research shows how differential privacy can be combined with collaborative training to reduce the risk of revealing sensitive information through updates [15]. In clinical prediction, this mechanism would be used to limit the possibility that a trained model memorizes or exposes individual patient records, particularly for rare adverse reactions or uncommon medication combinations. The privacy module would maintain an auditable privacy budget while allowing institutions to decide how much privacy protection they require for participation.

Additional Privacy Layers

Additional privacy layers could include secure aggregation, trusted execution environments for aggregation logic, cryptographic protocols, and, in highly sensitive settings, homomorphic encryption. Secure and federated medical machine learning reviews emphasize that no single technique is sufficient for every threat model, especially when clinical data are high-dimensional and institutionally sensitive [8]. Secure aggregation can protect individual hospital updates during each training cycle [14], while broader federated learning research highlights the need to consider adversarial clients, inference attacks, and governance assumptions [20]. A layered privacy architecture would therefore treat federation as one component of a wider confidentiality and accountability system.

Privacy-Utility Trade-Off Management

Privacy-utility trade-off management would be handled as a governance and evaluation function rather than as a fixed technical setting. Hospitals could select privacy parameters consistent with local policy, and the consortium could monitor whether stronger privacy protection changes calibration, discrimination, or clinical usability in ways that require model revision. The future of digital health with federated learning depends on balancing practical deployment, privacy protection, and clinical value across institutions [13]. For adverse drug reaction prediction, this balance should be evaluated conceptually and prospectively before the framework is used to guide high-stakes clinical or regulatory decisions.

Table 3 shows key considerations for managing the privacy-utility trade-off in federated learning for healthcare applications.

Table 3. Privacy-Utility Trade-Off in Federated Learning for Healthcare

Consideration	Description	Example / Implication
Governance Approach	Privacy-utility trade-offs are managed through oversight rather than fixed technical settings	Consortium sets review policies and monitors outcomes
Local Privacy Control	Hospitals can choose privacy parameters aligned with their internal policies	Hospital A selects differential privacy level according to local regulations
Evaluation Metrics	Monitor effects of privacy settings on model performance and usability	Assess changes in calibration, discrimination, and clinical decision support effectiveness
Clinical Impact	Balance between privacy protection and model utility is crucial for patient care	Strong privacy settings may reduce predictive accuracy for adverse drug reactions
Prospective Assessment	Evaluate trade-offs before high-stakes deployment	Conceptual and simulation studies for ADR prediction to guide clinical or regulatory decisions

Handling Data Heterogeneity and Non-Iid Challenges

Feature Harmonization across Sites

Feature harmonization would begin with a shared data specification that maps medication exposures, diagnoses, procedures, laboratory values, and adverse reaction outcomes into a common representation before local training. Standardized patient-level prediction pipelines in observational health data show how common data models can support reproducible feature construction across institutions [1]. For adverse drug reaction detection, structured vocabularies such as diagnosis codes, drug concepts, and adverse event terminologies would need to be aligned with locally documented clinical evidence, including laboratory signals and narrative descriptions. Text mining studies of adverse drug reactions in hospital records show that free-text notes may capture information that structured fields miss, but they also require consistent annotation and extraction policies before they can be used across sites [21, 22].

Dealing with Differing ADR Prevalence and Covariate Distributions

Differing adverse reaction prevalence, prescribing patterns, comorbidity burden, and monitoring intensity can cause federated clients to learn from non-identically distributed data. Clustered federated learning provides one conceptual strategy by grouping sites with similar learning behavior rather than forcing all institutions into a single undifferentiated model [16]. Federated adverse drug reaction prediction on distributed health data has already been framed as a setting where local clinical differences must be respected while still enabling shared learning [23]. The proposed framework would therefore support stabilization strategies such as proximal regularization, adaptive aggregation, and local adaptation when site-level label distributions or covariate patterns diverge substantially.

Federated Hyper-Parameter Tuning and Model Selection

Federated hyper-parameter tuning would need to evaluate candidate model configurations without exposing validation records from participating hospitals. General federated learning research identifies model selection, client sampling, and optimization under heterogeneous data as unresolved system-level issues [20]. In a pharmacovigilance setting, candidate models should be compared using locally computed validation summaries that are aggregated in a privacy-preserving manner, rather than by transferring patient-level validation data. Swarm learning and related decentralized clinical machine learning approaches suggest that coordination can occur without a conventional central data repository, which may be relevant when institutions prefer distributed governance structures [24].

Model Personalization and Local Adaptation

Local Fine-Tuning of the Global Model

After the global model is trained through federation, each hospital could adapt it locally to reflect site-specific prescribing, monitoring, and documentation patterns. Personalized or clustered federated approaches are relevant because they acknowledge that a single global model may not represent every hospital equally well [16]. Federated healthcare informatics literature also emphasizes that clinical deployment often requires balancing shared model performance with local usability and trust [6]. In this framework, local fine-tuning would be treated as a controlled adaptation step, with governance rules ensuring that local changes do not undermine safety monitoring or interpretability.

Context-Aware Inference

Context-aware inference would allow the locally deployed model to account for hospital-specific population characteristics without transmitting those characteristics to other sites. For example, local calibration layers or institution-specific decision thresholds could reflect differences in age distribution, renal function monitoring, comorbidity patterns, or drug utilization while keeping local statistics private. Smart healthcare federated learning systems illustrate how distributed learning can be combined with local deployment requirements in privacy-sensitive clinical environments [25]. This approach would preserve the benefits of a shared adverse drug reaction model while allowing hospitals to adapt predictions to their own care context.

Integration Into Pharmacovigilance Workflows

Deployment within Hospital Safety Systems

The federated model would run locally within hospital safety infrastructure, such as electronic health record-integrated clinical decision support, pharmacy review queues, or medication surveillance dashboards. Natural language processing reviews show that adverse drug event detection can draw on clinical notes as well as structured fields, making local workflow integration important for presenting interpretable evidence to pharmacists and clinicians [26]. The framework should therefore return not only risk estimates but also locally derived supporting signals such as recent medication exposure, abnormal laboratory trajectories, or relevant note-derived evidence. These outputs would be intended to support expert review rather than replace clinical judgment.

Network-Wide Signal Detection and Reporting

At the network level, the framework could support pharmacovigilance by sharing aggregate model behavior, site-level validation summaries, and de-identified safety signal patterns rather than patient-level records. Drug safety signal identification from electronic health records has been described as a promising but operationally complex field, especially when institutions differ in data capture and event definitions [2]. Distributed adverse drug reaction prediction using federated learning provides a direct conceptual bridge between privacy-preserving model training and multi-site pharmacovigilance [23]. Regulatory reporting would still require clear rules for signal adjudication, provenance tracking, and escalation from model-generated alerts to formal safety review.

Evaluation Strategy

Model Performance under Federation

Evaluation should compare the federated model conceptually against locally trained single-site models and, where governance permits, a reference centralized model trained under approved conditions. Federated clinical outcome prediction studies provide examples of how distributed models can be assessed across participating institutions while preserving local control of data [4, 5]. For adverse drug reaction endpoints, evaluation should consider discrimination, calibration, clinical interpretability, and consistency across medication classes and patient subgroups, without assuming that a single metric is sufficient. Any

comparison should be framed as prospective validation of feasibility and clinical value rather than as a claim of guaranteed superiority.

Table 4 consolidates the evaluation and governance requirements that determine whether a federated adverse drug reaction prediction model is ready for prospective pharmacovigilance use.

Table 4. Evaluation and Governance Matrix for Federated ADR Prediction Readiness

Readiness domain	Core question for implementation	Recommended assessment approach	Minimum evidence needed before deployment	Main risk if neglected	Governance owner or reviewer
ADR phenotype validity	Are adverse drug reaction labels clinically meaningful, temporally plausible, and comparable across hospitals?	Cross-site phenotype review, clinician adjudication samples, exposure-window sensitivity checks	Documented endpoint definitions, medication exposure rules, and site-level phenotype mapping	The model learns inconsistent or clinically invalid safety labels	Pharmacovigilance committee, clinical pharmacy, safety reviewers
Data harmonization	Can medication, diagnosis, laboratory, encounter, and text-derived variables be represented consistently across sites?	Common data model mapping, terminology audits, missingness profiling, feature availability comparison	Site-level data quality reports and harmonized feature dictionaries	Apparent multi-site learning reflects coding artifacts rather than true ADR risk	Data governance team, informatics leads
Federated model performance	Does the federated model improve or complement local single-site prediction without masking site-level weaknesses?	Compare local-only models, federated global model, and locally adapted versions using site-held validation data	Discrimination, calibration, subgroup performance, and medication-class performance reported by site	A global model performs acceptably on average but poorly in specific hospitals or patient groups	Model evaluation group, clinical AI oversight board
Calibration and threshold suitability	Are risk scores clinically interpretable and aligned with local alert burden and review capacity?	Calibration plots, decision-curve analysis, threshold simulation, pharmacist workload review	Locally approved thresholds and documented alert-volume projections	Excessive alerts, missed signals, or poor clinical trust	Pharmacy leadership, CDS governance committee
Privacy protection	Are model updates protected against plausible inference or reconstruction risks?	Differential privacy accounting, secure aggregation validation, threat-model review, privacy audit	Documented privacy budget, aggregation protocol, and institutional approval	Sensitive patient information may be inferred from model updates	Privacy office, cybersecurity, legal counsel
Communication and operational feasibility	Can institutions participate reliably without excessive technical burden?	Update-size monitoring, synchronization testing, downtime simulation, client participation tracking	Evidence that training cycles complete under realistic site infrastructure constraints	Sites drop out, updates become unstable, or training cycles fail	IT operations, federation coordinator
Non-IID robustness	Does the framework handle differences in prescribing patterns, ADR prevalence, laboratory monitoring, and documentation?	Site-stratified validation, heterogeneity diagnostics, clustered or personalized model comparison	Evidence that performance does not collapse under site-level distribution differences	The model overfits dominant institutions and under-serves smaller or different hospitals	Federated modeling team, consortium analytics board
Interpretability and evidence presentation	Can clinicians understand why a case is flagged for review?	Local explanation review, case-based pharmacist assessment, supporting-signal audit	Clear display of medication exposure, temporal laboratory change, coded evidence, and note-derived signals	Alerts are ignored because they lack actionable rationale	Clinical pharmacy, safety surveillance team
Workflow integration	Does the prediction output enter the right safety process at the right time?	Prospective pilot in pharmacy queues, CDS systems, or surveillance dashboards	Documented review pathway, escalation criteria, and user feedback results	Technically valid predictions fail to influence safety action	Medication safety leadership, CDS governance
Accountability and model provenance	Can the consortium trace model versions, update sources, validation summaries, and deployment decisions?	Version-control audit, update provenance logs, governance documentation review	Complete model cards, update logs, validation reports, and approval records	Responsibility for model behavior becomes unclear across institutions	Consortium steering committee, regulatory affairs, institutional compliance

Privacy and Communication Metrics

Privacy evaluation should document the protection mechanisms used during model update exchange, including secure aggregation, differential privacy accounting, and any cryptographic safeguards. Practical secure aggregation research provides the basis for evaluating whether the server can aggregate updates without observing individual client contributions [14]. Hybrid privacy-preserving federated learning methods further show that privacy must be evaluated together with model utility and system practicality [15]. Communication evaluation should consider update size, synchronization burden, client availability, and resilience to interrupted participation, while avoiding claims about efficiency until the system is tested in real deployments.

Prospective Real-World Pilot

A prospective pilot would evaluate whether the federated training cycle can be governed, operated, monitored, and integrated into live pharmacovigilance processes across participating hospitals. Multi-institutional medical federated learning studies demonstrate that collaboration without patient data exchange is feasible in principle, but pharmacovigilance adds additional requirements for adverse reaction definitions, safety review, and regulatory accountability [7]. Pretrained patient trajectory modeling for adverse drug event prediction suggests that longitudinal representations may be useful in future implementations, provided that local validation confirms their clinical relevance. The pilot should therefore focus on feasibility, workflow fit, privacy review, and the reliability of known signal detection pathways before broader deployment.

Limitations

Data Quality and Coding Consistency

The framework would remain vulnerable to inconsistent coding, incomplete medication reconciliation, missing laboratory data, variable note quality, and differences in adverse reaction documentation across institutions. Electronic health record-based adverse drug event prediction reviews emphasize that endpoint definition and data completeness are central limitations for model development. Studies using text mining to identify adverse drug reactions from hospital notes also show that relevant safety information may be present in unstructured documentation but inconsistently expressed across clinical settings [21, 22]. Strong data governance, local quality checks, and shared phenotype definitions would therefore be prerequisites for meaningful federated pharmacovigilance.

Regulatory and Legal Barriers

Technical privacy safeguards do not eliminate the need for institutional agreements, ethics review, cybersecurity assessment, and compliance with national and regional data protection requirements. Secure medical machine learning literature emphasizes that privacy-preserving architecture must be accompanied by governance, auditability, and a clear understanding of threat models [8]. Federated learning for digital health also raises practical questions about responsibility for model updates, monitoring, failure modes, and cross-site accountability [13]. These barriers may slow adoption even when the framework does not require patient-level data sharing.

Conclusion

A federated learning framework for adverse drug reaction prediction would allow healthcare institutions to collaborate on model development while keeping patient-level records within local environments. Each hospital would train on its own electronic health record data, transmit protected model updates, and receive an updated global model for local deployment. This design could help overcome the fragmentation that limits single-site pharmacovigilance models. It would also preserve institutional control over sensitive clinical data.

The main strength of the proposed framework is its ability to combine broader clinical learning with privacy-preserving design. Multi-institutional participation could expose the model to more diverse prescribing patterns, comorbidities, laboratory trajectories, and adverse reaction presentations. Layered safeguards such as secure aggregation, differential privacy, and local governance would support confidentiality while enabling collaboration. Local adaptation would further allow hospitals to align the shared model with their own clinical context.

Important challenges remain before such a system could be considered ready for routine pharmacovigilance. Data harmonization, adverse reaction phenotype definition, privacy governance, and integration into clinical workflows would all require careful design. Prospective validation would be necessary to determine whether the framework supports reliable safety surveillance in real care environments. Regulatory acceptance would also depend on transparent documentation of model behavior, update provenance, and institutional accountability.

A multi-site pharmacovigilance consortium would be a practical next step for piloting this framework. Such a consortium could define shared adverse reaction phenotypes, governance standards, privacy requirements, and evaluation protocols. The goal would not be to replace existing pharmacovigilance systems, but to augment them with collaborative machine learning that respects patient privacy and institutional boundaries. Over time, this approach could support a shared global model that benefits all participating health systems while maintaining local control.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Reps JM, Schuemie MJ, Suchard MA, Ryan PB, Rijnbeek PR. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. J Am Med Inform Assoc. 2018;25(8):969-75.
2. Davis SE, Zabotka L, Desai RJ, Wang SV, Maro JC, Coughlin K, et al. Use of electronic health record data for drug safety signal identification: a scoping review. Drug Saf. 2023;46(8):725-42.
3. Brisimi TS, Chen R, Mela T, Olshevsky A, Paschalidis IC, Shi W. Federated learning of predictive models from federated electronic health records. Int J Med Inform. 2018;112:59-67.
4. Vaid A, Jaladanki SK, Xu J, Teng S, Kumar A, Lee S, et al. Federated learning of electronic health records to improve mortality prediction in hospitalized patients with COVID-19: machine learning approach. JMIR Med Inform. 2021;9(1):e24207.
5. Dayan I, Roth HR, Zhong A, Harouni A, Gentili A, Abidin AZ, et al. Federated learning for predicting clinical outcomes in patients with COVID-19. Nat Med. 2021;27(10):1735-43.
6. Xu J, Glicksberg BS, Su C, Walker P, Bian J, Wang F. Federated learning for healthcare informatics. J Healthc Inform Res. 2021;5(1):1-9.
7. Guo P, Wang P, Zhou J, Jiang S, Patel VM. Multi-institutional collaborations for improving deep learning-based magnetic resonance image reconstruction using federated learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2021. p. 2423-32.
8. Kaissis GA, Makowski MR, Rückert D, Braren RF. Secure, privacy-preserving and federated machine learning in medical imaging. Nat Mach Intell. 2020;2(6):305-11.
9. Christopoulou F, Tran TT, Sahu SK, Miwa M, Ananiadou S. Adverse drug events and medication relation extraction in electronic health records with ensemble deep learning methods. J Am Med Inform Assoc. 2020;27(1):39-46.
10. Henry S, Buchan K, Filannino M, Stubbs A, Uzuner O. 2018 n2c2 shared task on adverse drug events and medication extraction in electronic health records. J Am Med Inform Assoc. 2020;27(1):3-12.
11. Wei Q, Ji Z, Li Z, Du J, Wang J, Xu J, et al. A study of deep learning approaches for medication and adverse drug event extraction from clinical text. J Am Med Inform Assoc. 2020;27(1):13-21.
12. Ju M, Nguyen NT, Miwa M, Ananiadou S. An ensemble of neural models for nested adverse drug events and medication extraction with subwords. J Am Med Inform Assoc. 2020;27(1):22-30.
13. Rieke N, Hancox J, Li W, Milletari F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. NPJ Digit Med. 2020;3(1):119.
14. Bonawitz K, Ivanov V, Kreuter B, Marcedone A, McMahan HB, Patel S, et al. Practical secure aggregation for privacy-preserving machine learning. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security; 2017. p. 1175-91.
15. Truex S, Baracaldo N, Anwar A, Steinke T, Ludwig H, Zhang R, et al. A hybrid approach to privacy-preserving federated learning. In: Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security; 2019. p. 1-11.
16. Sattler F, Müller KR, Samek W. Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints. IEEE Trans Neural Netw Learn Syst. 2020;32(8):3710-22.
17. Dai HJ, Su CH, Wu CS. Adverse drug event and medication extraction in electronic health records via a cascading architecture with different sequence labeling models and word embeddings. J Am Med Inform Assoc. 2020;27(1):47-55.
18. Zhang W, Kuang Z, Peissig P, Page D. Adverse drug reaction discovery from electronic health records with deep neural networks. In: Proceedings of the ACM Conference on Health, Inference, and Learning; 2020. p. 30-9.
19. Sakhovskiy A, Tutubalina E. Multimodal model with text and drug embeddings for adverse drug reaction classification. J Biomed Inform. 2022;135:104182.
20. Kairouz P, McMahan HB. Advances and open problems in federated learning. Found Trends Mach Learn. 2021;14(1-2):1-210.
21. Wasylewicz A, van de Burgt B, Weterings A, Jessurun N, Korsten E, Egberts T, et al. Identifying adverse drug reactions from free-text electronic hospital health record notes. Br J Clin Pharmacol. 2022;88(3):1235-45.
22. Van De Burgt BW, Wasylewicz AT, Dullemond B, Jessurun NT, Grouls RJ, Bouwman RA, et al. Development of a text mining algorithm for identifying adverse drug reactions in electronic health records. JAMIA Open. 2024;7(3):ooae070.
23. Choudhury O, Park Y, Salonidis T, Gkoulalas-Divanis A, Sylla I, k Das A. Predicting adverse drug reactions on distributed health data using federated learning. In: AMIA Annual Symposium Proceedings. 2020;2019:313.
24. Warnat-Herresthal S, Schultze H, Shastry KL, Manamohan S, Mukherjee S, Garg V, et al. Swarm learning for decentralized and confidential clinical machine learning. Nature. 2021;594(7862):265-70.

25. Li J, Meng Y, Ma L, Du S, Zhu H, Pei Q, et al. A federated learning based privacy-preserving smart healthcare system. IEEE Trans Ind Inform. 2021;18(3).
26. Golder S, Xu D, O'Connor K, Wang Y, Batra M, Hernandez GG. Leveraging natural language processing and machine learning methods for adverse drug event detection in electronic health/medical records: a scoping review. Drug Saf. 2025;48(4):321-37.