ORIGINAL RESEARCH **Open access**

Federated Transfer Learning Framework for Adapting COVID-19 Prognostic Models from High-Resource to Low-Resource Hospitals Using Only Aggregate Statistics

Diego Morales[1*], Andres Gutierrez[1], Lucia Navarro[2], Pablo Rios[1]

Abstract

During the COVID-19 pandemic, machine learning models developed in high-resource hospitals achieved strong performance in predicting outcomes such as mortality, ICU admission, and mechanical ventilation, but their accuracy often degrades when applied to low-resource settings due to differences in patient populations, disease severity, clinical practices, and documentation quality. Low-resource hospitals also face limited patient volumes, incomplete labeled data, and strict privacy regulations (e.g., HIPAA and GDPR), which prevent centralized data sharing and hinder independent model development, creating a barrier to equitable AI deployment. To address this, we propose a federated transfer learning framework that adapts prognostic models from high-resource to low-resource hospitals without exchanging patient-level data. The approach transfers only aggregate statistics (e.g., feature means, variances, class-conditional distributions, and correlations) via a secure lightweight protocol, enabling target hospitals to align feature distributions using domain adaptation techniques and fine-tune models on small local datasets. The framework includes source model training, statistical aggregation, secure transmission, and target-side adaptation modules, ensuring no raw patient data leaves any institution. By relying on aggregate statistics, the method preserves privacy while mitigating domain shift and maintaining clinical utility across diverse healthcare environments. This scalable and privacy-preserving framework supports broader deployment of COVID-19 predictive models and provides a generalizable strategy for other medical conditions with heterogeneous healthcare settings.

Keywords Health equity, Domain adaptation, Privacy-preserving machine learning, Federated transfer learning, COVID-19 prognosis, Aggregate statistics

*Correspondence:
Diego Morales
diego.morales@gmail.com
[1] Department of Intelligent Healthcare Engineering, University of Lima, Lima, Peru
[2] Department of AI Clinical Systems, Pontifical Catholic University of Peru, Lima, Peru

Introduction

The rapid global dissemination of SARS-CoV-2 created an urgent need for clinical prognostic tools capable of identifying patients at highest risk for deterioration, mortality, and resource-intensive interventions. High-resource academic medical centers leveraged their substantial electronic health record infrastructure to develop machine learning models predicting critical COVID-19 outcomes including 30-day mortality, intensive care unit admission, and progression to invasive mechanical ventilation [1, 2]. These models incorporated diverse predictor sets spanning patient demographics, pre-existing comorbidities, vital sign trajectories, laboratory values, and chest imaging findings. Validation studies conducted within the institutions where these models were developed reported strong discriminative performance, with area under the receiver operating characteristic curve values frequently exceeding 0.85 for mortality prediction [3]. The clinical promise of these tools motivated efforts to deploy

them across broader healthcare networks to support triage decisions, resource allocation, and early intervention protocols during pandemic surges.

Subsequent investigations revealed substantial performance deterioration when prognostic models developed at tertiary care centers were applied to community hospitals, rural facilities, and international settings with different patient populations and care delivery patterns [4]. The phenomenon of dataset shift—encompassing covariate shift in patient demographics and comorbidity profiles, prior probability shift in outcome rates driven by differing admission thresholds, and concept shift arising from heterogeneous treatment protocols—systematically undermined model calibration and discrimination [5, 6]. Models trained on data from urban academic centers in high-income countries frequently overestimated mortality risk when applied to younger patient cohorts in community settings or underestimated risk in populations with different comorbidity burdens and healthcare access patterns. This generalizability failure represents a fundamental barrier to equitable deployment of AI-based prognostic tools and risks exacerbating existing healthcare disparities between well-resourced and under-resourced institutions.

Low-resource hospitals confront interconnected obstacles to developing and deploying local prognostic models. Individual facilities in rural or underserved regions may treat insufficient numbers of COVID-19 patients to achieve statistical power for reliable model training, particularly for outcomes such as mortality or mechanical ventilation that occur in a minority of hospitalized patients [7]. Electronic health record systems in these settings may lack the structured data elements, standardization, and completeness required for model development workflows designed at academic centers. Furthermore, privacy regulations including the Health Insurance Portability and Accountability Act, the General Data Protection Regulation, and emerging data localization laws in jurisdictions such as Germany, France, and China impose legal and technical barriers to the centralization of patient-level data from multiple institutions [8]. Even within single-payer health systems, governance structures and institutional review board requirements can impede data aggregation across facilities, leaving low-resource hospitals isolated from the data ecosystems necessary for robust model development.

This manuscript proposes a federated transfer learning framework that enables low-resource hospitals to adapt prognostic models originally trained at high-resource institutions using only aggregate statistical summaries of the source data distribution. The framework constrains all cross-institutional information exchange to feature-level means, standard deviations, and correlation matrices, which cannot be inverted to reconstruct individual patient records while preserving sufficient distributional information to guide domain adaptation [9]. The architecture comprises four interoperable modules—source model training, aggregate statistic extraction, a federated transfer protocol, and target model adaptation—that jointly address the privacy, communication, and data scarcity constraints characteristic of low-resource healthcare settings. The subsequent sections detail the clinical and technical background motivating this approach, present the framework architecture and component specifications, outline an evaluation strategy for assessing adaptation performance, and discuss the limitations and implementation considerations relevant to real-world deployment.

Background

COVID-19 prognostic models

Machine learning models developed during the COVID-19 pandemic targeted a consistent set of clinically meaningful outcomes including 30-day all-cause mortality, transfer to intensive care, initiation of invasive mechanical ventilation, and composite measures of severe disease progression [10]. Predictor variables spanned domains of patient demographics, pre-existing comorbidities assessed via diagnosis codes or chart review, admission vital signs and their trajectories during early hospitalization, laboratory parameters reflecting inflammatory and organ dysfunction pathways, and imaging findings from chest radiography and computed tomography [11]. Modeling approaches ranged from logistic regression and Cox proportional hazards frameworks to gradient-boosted tree ensembles and deep neural architectures, with performance assessments focusing on discrimination metrics, calibration characteristics, and net clinical benefit at clinically relevant decision thresholds. These models demonstrated that routinely collected clinical data contained substantial prognostic signal that could augment clinician judgment during pandemic conditions characterized by high patient volumes and evolving care protocols [12].

Performance drop across settings

The systematic degradation of prognostic model performance when transported across institutions has been documented across multiple COVID-19 cohorts and modeling approaches. Dataset shift manifests through several distinct mechanisms that compound to undermine model reliability in target settings. Covariate shift occurs when the joint distribution of predictor variables differs between source and target populations, as when a model trained on an older, comorbidity-burdened academic center population is applied to younger patients in a community hospital with different prevalence patterns of diabetes, hypertension, and immunosuppression [13]. Prior probability shift, also termed label shift, arises when the base rate of outcomes such as mortality or ICU admission differs substantially between institutions due to varying admission thresholds, transfer patterns, and local practice norms. Concept shift represents changes in the conditional relationship between predictors and outcomes, as when the prognostic significance of a given laboratory abnormality depends on locally available treatment modalities and clinical response capabilities [14]. These shifts interact to produce miscalibration, where models systematically overestimate or underestimate risk, potentially leading to inappropriate resource allocation decisions in target settings.

Privacy constraints

Legal frameworks governing health data privacy impose substantial constraints on the cross-institutional sharing of patient-level information. In the United States, the Health Insurance Portability and Accountability Act establishes standards for the protection of individually identifiable health information, requiring patient authorization or specific regulatory exceptions for data disclosure [15]. The European Union's General Data Protection Regulation provides even more stringent protections, classifying health data as a special category requiring explicit consent and imposing data minimization and purpose limitation principles that constrain secondary use for model development. Nation-level data localization requirements enacted in jurisdictions including China, Russia, and several European countries mandate that health data remain within national borders, creating legal obstacles to international research collaborations. Beyond formal legal constraints, ethical governance frameworks, institutional review board requirements, and public concerns about data misuse create practical barriers to data centralization that persist even when legal mechanisms for data sharing exist [16]. These constraints collectively motivate privacy-preserving approaches that derive multi-institutional insights without aggregating patient-level data.

Transfer learning and domain adaptation

Transfer learning encompasses a family of techniques that leverage knowledge extracted from a source domain with abundant labeled data to improve model performance in a target domain where labeled data is scarce or absent. Fine-tuning approaches initialize a target model with parameters learned on the source task, then perform additional training on target data to adapt the decision boundary to local distributions [17]. Feature alignment methods, including correlation alignment, minimize distributional discrepancy between source and target feature representations by matching first-order and second-order statistics in a learned embedding space. Adversarial domain adaptation frameworks, such as the domain-adversarial neural network architecture, train feature extractors to produce representations that are simultaneously discriminative for the prediction task and invariant to domain membership [18]. Distribution alignment strategies offer the particular advantage in privacy-sensitive healthcare applications that they can operate on summary statistics of feature distributions rather than raw data points, enabling domain adaptation under the data sharing constraints characteristic of multi-institutional medical research [19]. These approaches have demonstrated effectiveness in clinical applications where patient populations, imaging protocols, or laboratory assays differ across institutions, though their application to COVID-19 prognosis under strict privacy constraints remains underexplored.

Framework Overview

High-level architecture

The proposed framework orchestrates a sequential workflow spanning high-resource and low-resource hospital environments while constraining all inter-institutional communication to aggregate statistical summaries. A source prognostic model is first trained within the high-resource hospital's secure computing environment using its local COVID-19 patient cohort, which may encompass thousands to tens of thousands of hospitalized cases with documented outcomes. Following source model training and validation, an aggregate statistics extraction module computes first-order and second-order distributional

summaries—feature-wise means, standard deviations, and class-conditional correlation matrices—from the model's feature space representation of the source population. These aggregate statistics, which comprise numerical vectors and matrices containing no patient-level information, are transmitted through an encrypted channel to a lightweight coordinating server that maintains no raw clinical data. An authorized low-resource hospital retrieves the source aggregate statistics pair and initializes a local model with the source parameters, then performs joint supervised fine-tuning on its small labeled dataset and distribution alignment using the received aggregate statistics. This architecture ensures that patient-level data never traverses institutional boundaries while enabling the low-resource hospital to benefit from the statistical regularities learned from the high-resource population.

Figure 1 illustrates the proposed aggregate-statistic federated transfer learning architecture, showing how source-domain prognostic knowledge can be transferred to low-resource hospitals without exchanging patient-level clinical records.

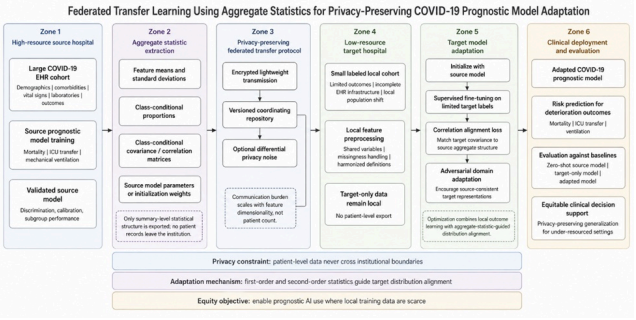

Figure 1. Privacy-preserving federated transfer learning architecture for adapting COVID-19 prognostic models from high-resource to low-resource hospitals using aggregate statistics.

Core assumptions

The framework operates under a set of explicit assumptions that define its scope of applicability. The high-resource hospital must have developed and validated a prognostic model on its local COVID-19 patient cohort and must be willing to share aggregate statistics derived from that model's feature space. The model architecture, feature preprocessing pipeline, and variable definitions must be standardized and documented to enable compatible implementation at the target hospital, though the specific model parameters remain learned from source data. The low-resource hospital must possess a labeled dataset of COVID-19 patients with documented outcomes, with the framework targeting settings where this dataset contains between 100 and 500 cases—insufficient for training a model de novo but potentially adequate for supervised fine-tuning and adaptation [20]. Outcome definitions must be sufficiently harmonized between source and target hospitals to ensure that the prediction task is clinically coherent, acknowledging that some irreducible variation in outcome adjudication may persist across institutions. Finally, the feature spaces of source and target populations must partially overlap, with shared demographic, laboratory, and clinical variables available in both settings, though the framework can accommodate differences in variable distributions and missing data patterns.

Design principles

Three guiding design principles inform the framework architecture. Privacy preservation constitutes the foundational constraint, operationalized through the prohibition of patient-level data exchange and the exclusive use of aggregate statistics for inter-institutional communication, with optional differential privacy noise injection providing formal privacy guarantees against inference attacks on the shared statistics [21]. Communication efficiency is achieved through the compact representation of distributional information as means, standard deviations, and correlation matrices, which scale with the square of the feature dimensionality rather than the number of source patients, making transmission feasible even in bandwidth-constrained healthcare environments. Resource agnosticism ensures that the framework imposes minimal computational requirements on the target hospital, with fine-tuning and distribution alignment operations designed to execute on commodity hardware without specialized accelerators. Together, these principles enable deployment across the heterogeneous technical infrastructure landscape characteristic of global healthcare delivery systems while maintaining the privacy protections necessary for multi-institutional collaboration.

Source Model Training

High-resource hospital model

The source hospital develops a prognostic model using its comprehensive COVID-19 patient registry, which may comprise several thousand to tens of thousands of hospitalized cases with systematically adjudicated

outcomes and structured clinical data elements. Model architectures compatible with the framework span the spectrum of approaches validated for clinical prediction tasks, including regularized logistic regression for interpretable linear decision boundaries, extreme gradient boosting for modeling non-linear feature interactions and handling mixed data types, and feedforward neural networks for learning hierarchical representations from raw clinical features [22]. The feature set should encompass admission demographics, pre-existing comorbidity indicators, vital sign measurements during the first 24 hours of hospitalization, and laboratory values reflecting hematologic, renal, hepatic, and inflammatory parameters. Outcome labels comprise binary indicators for 30-day all-cause mortality, ICU transfer within 7 days of admission, and progression to invasive mechanical ventilation, with temporal anchoring to hospital admission to ensure consistent prediction horizons. Feature preprocessing includes standardization, imputation of missing values using population medians or more sophisticated multiple imputation strategies, and encoding of categorical variables.

Baseline performance

Comprehensive performance characterization of the source model on held-out source hospital data establishes the reference against which adaptation outcomes will be measured. Performance assessment encompasses discrimination metrics including area under the receiver operating characteristic curve and area under the precision-recall curve, which is particularly informative for imbalanced outcomes such as mortality and ventilation. Calibration evaluation employs the expected calibration error and visual inspection of calibration curves across the probability spectrum, as miscalibration can produce clinically consequential overestimation or underestimation of individual patient risk [23]. The Brier score provides a composite measure integrating discrimination and calibration characteristics. Stratified analyses across demographic subgroups and comorbidity categories identify potential performance disparities that may propagate to target settings. This baseline performance characterization serves three purposes: it documents the source model's capability within its development environment, it provides the initial parameters that will seed target model adaptation, and it establishes the benchmark against which zero-shot transfer and adaptation outcomes will be compared to quantify the benefit of the proposed framework.

Aggregated Statistic Extraction

First-order statistics

First-order distributional statistics capture the central tendency and dispersion of each feature within the source population, computed separately for patients experiencing each clinical outcome to preserve class-conditional information relevant to discriminative modeling. For continuous features including age, vital sign measurements, and laboratory values, the extraction module computes the mean and standard deviation within each outcome class, providing parameter estimates for the marginal distributions that guide subsequent alignment procedures [24]. For binary and categorical features, class-conditional proportions are computed, capturing the prevalence of comorbidities, medication exposures, and other discrete clinical attributes among patients who did and did not experience the outcome of interest. These first-order statistics convey substantial prognostic information: the degree of separation between class-conditional means directly informs the discriminative potential of individual features, while differences in class-conditional variances signal heteroscedasticity that may affect model calibration. Critically, means and standard deviations computed across cohorts of sufficient size provide negligible information about any individual patient, as the mapping from these aggregate parameters to specific patient-level feature values is not identifiable.

Second-order statistics

Second-order statistics encode pairwise linear dependencies among features through class-conditional correlation or covariance matrices computed from the source population feature vectors. These matrices capture clinically informative co-variation patterns, including the expected correlation between inflammatory markers such as C-reactive protein and procalcitonin, the inverse relationship between oxygenation parameters and respiratory rate among patients with respiratory failure, and the clustering of metabolic derangements in patients with multi-organ dysfunction [25]. Correlation alignment uses these second-order statistics to define a target distribution toward which the target hospital's feature representations should be transformed, formalized as the minimization of the Frobenius norm between the source and target feature covariance matrices. This alignment procedure can be implemented in either the original feature space, preserving

interpretability of individual variables, or in a learned representation space produced by a shared feature extractor, enabling alignment of non-linear feature transformations while still guided by source aggregate statistics. The extracted second-order statistics, encoded as symmetric matrices with dimensionality quadratic in the number of features, remain vastly more compact than patient-level datasets and resist inversion to individual-level data, particularly when computed over sufficiently large source cohorts.

Table 1 clarifies how different aggregate statistics contribute distinct forms of transferable knowledge while maintaining the manuscript's central prohibition on patient-level data exchange.

Table 1. Functional Contribution of Aggregate Statistics to Privacy-Preserving Domain Adaptation

Aggregate statistic shared by source hospital	Statistical information transferred	Adaptation function in the target hospital	Privacy-preserving rationale
Feature-wise means	Central tendency of each clinical predictor in the source population	Guides normalization, distribution matching, and identification of covariate shift	A mean summarizes th cohort and do not identify individual patients wher computed ove sufficiently larg samples
Feature-wise standard deviations	Dispersion and scale of each predictor	Supports feature rescaling and alignment of marginal variability	Variance information is aggregate-lev and not direct invertible to individual records
Class-conditional means	Predictor separation between outcome-positive and outcome-negative groups	Helps preserve discriminative structure during target fine-tuning	Transfers outcome-relevant signa without exposi individual labeled case
Class-conditional proportions	Outcome-stratified prevalence of binary or categorical variables	Supports alignment of comorbidity and exposure patterns across hospitals	Shares prevalence patterns rathe than patient identities
Correlation or covariance matrices	Pairwise dependency structure among clinical features	Enables correlation alignment and second-order distribution matching	Matrix summaries ar underdetermin relative to patient-level records
Source model parameters or initialization weights	Learned predictive structure from the high-resource domain	Provides starting point for target fine-tuning and reduces sample requirements	Parameters encode task-relevant patter without requiri raw source records
Differentially private noisy statistics	Perturbed aggregate summaries with formal privacy protection	Balances statistical utility with stronger disclosure protection	Noise limits inference abo any single source patier

Federated Transfer Protocol

Communication architecture

The federated transfer protocol defines a lightweight communication infrastructure mediating aggregate statistic exchange between source and target hospitals without persistent connectivity. A neutral coordinating server maintains a repository of uploaded feature means, standard deviations, class-conditional proportions, and correlation matrices, versioned and timestamped to allow target hospitals to select statistics matching their local context [26]. Target hospitals authenticate, download the desired package, and initiate adaptation within their own secure environments. This decoupled architecture accommodates asynchronous institutional workflows.

Privacy guarantees

Privacy derives from the mathematical irreversibility of aggregate-to-individual mappings, with feature means and standard deviations over large cohorts conveying distributional information asymptotically independent of any single record [27]. A differential privacy mechanism can add calibrated Gaussian noise to transmitted statistics, with privacy budget epsilon controlling the fidelity-protection trade-off. Correlation matrices scale with feature count rather than patient count, and reconstruction attacks face fundamental underdetermination when patients outnumber features, providing layered structural, formal, and dimensional protections.

Adversarial domain alignment

Correlation alignment penalizes the Frobenius distance between target feature covariance and the received source covariance matrix, encouraging representations with matched second-order structure [28]. A domain-adversarial neural network augments the prognostic model with a domain classifier trained adversarially using source aggregate statistics, employing a gradient reversal layer to produce domain-invariant representations. This hybrid approach addresses both marginal and complex representational shifts without accessing individual source records.

Target Model Adaptation

Fine-Tuning with Regularization

Target model adaptation initializes with source parameters and performs supervised fine-tuning on the limited labeled dataset using an objective combining binary cross-entropy, L2 weight decay, and elastic weight consolidation to prevent catastrophic forgetting [29]. Early stopping on a held-out target validation subset and a reduced learning rate constrain parameter updates. For gradient-boosted tree models, adaptation proceeds through additional boosting rounds with source predictions as an offset, learning residual corrections from local data patterns.

Distribution alignment

Distribution alignment is integrated into fine-tuning by adding correlation alignment loss to the supervised objective with a tunable weighting hyperparameter. A domain-adversarial classifier head distinguishes source-consistent from target-deviant representations, with reversed gradients encouraging domain invariance. This joint optimization enables the target model to exploit both limited outcome labels and distributional knowledge from source statistics, yielding representations predictive of clinical outcomes and robust to distributional shift.

Evaluation Strategy

Adaptation metrics

Discrimination is assessed via area under the receiver operating characteristic and precision-recall curves, while calibration is measured through expected calibration error. The primary comparison contrasts the adapted model against zero-shot source model application to the target population. A secondary comparison evaluates against a target-only baseline trained without source knowledge, quantifying the marginal transfer benefit. Net reclassification improvement and integrated discrimination improvement assess clinical consequence at relevant thresholds.

Sample efficiency

Systematic evaluation across target sample sizes of 50, 100, 200, and 500 labeled patients characterizes adaptation sample efficiency. Performance trajectories for the adapted model, zero-shot source model, and target-only baseline are compared, with the gap between curves representing the efficiency benefit of aggregate statistic transfer. The diminishing returns profile identifies minimum data requirements and guides resource-limited hospitals in implementation decisions.

Sensitivity analysis

Robustness to source data quality is assessed by simulating degradation through missing data, measurement error, and outcome misclassification. Distribution shift magnitude is quantified via Kullback-Leibler divergence and total variation distance, enabling stratified analysis of adaptation benefit. Scenarios with discordant outcome definitions between hospitals examine transferability limits when local bed capacity or triage protocols fundamentally alter endpoint interpretation.

Table 2 provides an evaluation framework that links statistical performance, privacy preservation, sample efficiency, and clinical deployment readiness in low-resource hospital settings.

Table 2. Evaluation Logic for Demonstrating Transfer Benefit, Privacy Utility, and Deployment Readiness

Evaluation dimension	Core question addressed	Recommended comparison	Prin indic
Zero-shot transportability	How poorly does the source model perform when directly applied to the target hospital?	Source model without adaptation vs. adapted model	AUF AUF expe calib error, sc
Target-only feasibility	Can the low-resource hospital train an adequate model independently?	Target-only model vs. adapted model	Perfor at 50 200, a labeled
Calibration reliability	Does adaptation produce clinically usable risk estimates rather than only ranking patients?	Adapted model vs. zero-shot and target-only baselines	Calib curve calib interce slo
Clinical utility	Does the adapted model improve decision-making at relevant thresholds?	Decision-curve analysis across risk thresholds	Net b n reclass improv integ discrim improv
Robustness to data quality variation	Does the method remain useful under incomplete or noisy target data?	Simulated missingness, measurement error, and outcome misclassification scenarios	Perfor degra curve sens pro
Robustness to distribution shift	Does adaptation help under different magnitudes of source-target mismatch?	Stratification by KL divergence, total variation distance, or covariance discrepancy	Transf as a fu of s magr
Privacy-utility tradeoff	How much predictive value is lost when stronger privacy protection is imposed?	Aggregate statistics with and without differential privacy noise	Utility l privacy va
Equity impact	Does the framework reduce performance gaps affecting under-resourced hospitals?	Adapted model performance across hospital types or patient subgroups	Subg AUF calibr false-n rate, positiv

Limitations

Technical limitations

First-order and second-order statistics may fail to capture multi-modal distributions, non-linear interactions, or higher-order dependencies influencing prognostic behavior. Distribution matching assumptions may not hold when optimal decision boundaries differ due to treatment effect heterogeneity. Source hospital cooperation is required, and institutions may decline participation due to competitive concerns or residual privacy risks. The framework does not address multi-source federated scenarios or mechanisms for reconciling heterogeneous statistics across populations.

Clinical limitations

Resource-dependent endpoints such as ICU admission and ventilation are intrinsically linked to local capacity and practice norms that vary across health systems. Treatment availability modifies the relationship between clinical features and outcomes in ways aggregate statistics cannot capture. Patient populations differ in unmeasured social

determinants, health-seeking behaviors, and environmental exposures not represented in structured records. These irreducible variations establish a transferability ceiling requiring acknowledgment during clinical deployment.

Conclusion

The proposed framework provides a systematic approach for adapting COVID-19 prognostic models from high-resource to low-resource hospitals through exclusive use of aggregate statistics, integrating source model training, distributional statistic extraction, federated transmission, and regularized target adaptation with distribution alignment. This design reconciles model generalizability with patient privacy protections.

Key advantages include a privacy-preserving architecture eliminating patient-level data exchange, low communication overhead from compact summary statistics, and accessibility for hospitals lacking independent model development capacity. Compatibility with diverse model architectures and commodity hardware lowers adoption barriers across heterogeneous healthcare infrastructures.

Important limitations include the prerequisite of source hospital cooperation, constraints on transferable distributional relationships from low-order statistics, and irreducible transferability ceilings from divergent outcome definitions and treatment contexts. These limitations necessitate context-specific validation before clinical deployment.

Implementation across international consortia, national health systems, and regional exchanges could expand prognostic model access to currently underserved facilities. The approach generalizes beyond COVID-19 to other conditions with heterogeneous populations and treatment paradigms. Prospective evaluation, multi-source extension, and clinical decision support integration represent essential directions for translating this framework into measurable patient outcome improvements.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 09 Mar 2025 Revised: 16 May 2025 Accepted: 18 Jul 2025
Published online: 20 January 2026

References

Vaid A, Jaladanki SK, Xu J, Teng S, Kumar A, 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.

Dayan I, Roth HR, Zhong A, Harouni A, Gentili A, et al. Federated learning for predicting clinical outcomes in patients with COVID-19. Nat Med. 2021;27(10):1735-43.

Wichmann RM, Robiati Bigoto MA, Chiavegatto Filho AD. Federated learning for COVID-19 mortality prediction in a multicentric sample of 21 hospitals. PLoS Comput Biol. 2025;21(11):e1013695.

Peng L, Luo G, Walker A, Zaiman Z, Jones EK, et al. Evaluation of federated learning variations for COVID-19 diagnosis using chest radiographs from 42 US and European hospitals. J Am Med Inform Assoc. 2023;30(1):54-63.

Cheng FY, Joshi H, Tandon P, Freeman R, Reich DL, et al. Using machine learning to predict ICU transfer in hospitalized COVID-19 patients. J Clin Med. 2020;9(6):1668.

Parchure P, Joshi H, Dharmarajan K, Freeman R, Reich DL, et al. Development and validation of a machine learning-based prediction model for near-term in-hospital mortality among patients with COVID-19. BMJ Support Palliat Care. 2020;12(e3):e424-e431.

Yadaw AS, Li YC, Bose S, Iyengar R, Bunyavanich S, et al. Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. Lancet Digit Health. 2020;2(10):e516-e525.

Li X, Ge P, Zhu J, Li H, Graham J, et al. Deep learning prediction of likelihood of ICU admission and mortality in COVID-19 patients using clinical variables. PeerJ. 2020;8:e10337.

van Klaveren D, Rekkas A, Alsma J, Verdonschot RJ, Koning DT, et al. COVID outcome prediction in the emergency department (COPE): using retrospective Dutch hospital data to develop simple and valid models for predicting mortality and need for intensive care unit admission in patients who present at the emergency department with suspected COVID-19. BMJ Open. 2021;11(9):e051468.

Cummings BC, Ansari S, Motyka JR, Wang G, Medlin RP Jr, et al. Predicting intensive care transfers and other unforeseen events: analytic model validation study and comparison to existing methods. JMIR Med Inform. 2021;9(4):e25066.

Castro VM, McCoy TH, Perlis RH. Assessment of the performance consistency of an adverse outcome prediction tool for patients hospitalized with COVID-19. JAMA Netw Open. 2021;4(7):e2118413.

Munera N, Garcia-Gallo E, Gonzalez Á, Zea J, Fuentes YV, et al. A novel model to predict severe COVID-19 and mortality using an artificial intelligence algorithm to interpret chest radiographs and clinical variables. ERJ Open Res. 2022;8(2).

Wendland P, Schmitt V, Zimmermann J, Häger L, Göpel S, et al. Machine learning models for predicting severe COVID-19 outcomes in hospitals. Inform Med Unlocked. 2023;37:101188.

Guo LL, Pfohl SR, Fries J, Johnson AE, Posada J, et al. Evaluation of domain generalization and adaptation on improving model robustness to temporal dataset shift in clinical medicine. Sci Rep. 2022;12(1):2726.

Mutnuri MK, Stelfox HT, Forkert ND, Lee J. Using domain adaptation and inductive transfer learning to improve patient outcome prediction in the intensive care unit: retrospective observational study. J Med Internet Res. 2024;26:e52730.

Pfisterer F, Harbron C, Jansen G, Xu T. Evaluating domain generalization for survival analysis in clinical studies. In: Conference on Health, Inference, and Learning. PMLR; 2022. p. 32-47.

Lee GH, Shin SY. Federated learning on clinical benchmark data: performance assessment. J Med Internet Res. 2020;22(10):e20891.

Tahir N, Jung CR, Lee SD, Azizah N, Ho WC, et al. Federated learning-based model for predicting mortality: systematic review and meta-analysis. J Med Internet Res. 2025;27:e65708.

Pati S, Kumar S, Varma A, Edwards B, Lu C, et al. Privacy preservation for federated learning in health care. Patterns. 2024;5(7).

Zhang F, Kreuter D, Chen Y, Dittmer S, Tull S, et al. Recent methodological advances in federated learning for healthcare. Patterns. 2024;5(6).

Shiri I, Salimi Y, Sirjani N, Razeghi B, Bagherieh S, et al. Differential privacy preserved federated learning for prognostic modeling in COVID 19 patients using large multi institutional chest CT dataset. Med Phys. 2024;51(7):4736-47.

Maurer MM, Pfitzner B, van de Water RP, Faraj L, Riepe C, et al. Privacy preserving federated learning for 90-day mortality prediction in colorectal surgery: a multicenter retrospective development and comparison study. Int J Surg. 2025;111(12):9065-74.

Yang X, Wan D, Han G, Zhang W, Tang W. Personalized federated learning with hierarchical reweighting for multi-center clinical prediction. Comput Methods Programs Biomed. 2025;109015.

Haripriya R, Khare N, Pandey M. Privacy-preserving federated learning for collaborative medical data mining in multi-institutional settings. Sci Rep. 2025;15(1):12482.

Zhao H, Sui D, Wang Y, Ma L, Wang L. Privacy-preserving federated learning framework for multi-source electronic health records prognosis prediction. Sensors (Basel). 2025;25(8):2374.

Sharma S, Guleria K. A collaborative privacy preserved federated learning framework for pneumonia detection using diverse chest X-ray data silos. Int J Math Eng Manag Sci. 2025;10(2):464.

Darzi E, Sijtsema NM, van Ooijen PM. A comparative study of federated learning methods for COVID-19 detection. Sci Rep. 2024;14(1):3944.

Qayyum A, Ahmad K, Ahsan MA, Al-Fuqaha A, Qadir J. Collaborative federated learning for healthcare: multi-modal COVID-19 diagnosis at the edge. IEEE Open J Comput Soc. 2022;3:172-84.

Zhang AS, Li NF. A two-stage federated transfer learning framework in medical images classification on limited data: a COVID-19 case study. In: Proceedings of the Future Technologies Conference. Cham: Springer Int Publ; 2022. p. 198-216.