International Journal of AI, Big Data, Computational and Management Studies
Noble Scholar Research Group | Volume 7, Issue 2, 54-60, 2026
ISSN: 3050-9416 | https://doi.org/10.63282/3050-9416.IJAIBDCMS-V7I2P109

Original Article

Clean Before Predict: A Governance-First Methodology for High-Stakes AI Systems

Nidhi Singh
Senior Data Analyst, State of Alabama, AL USA.

Received On: 15/02/2026 ***Revised On: 25/03/2026*** ***Accepted On: 03/04/2026*** ***Published On: 10/04/2026***

Abstract: *Artificial intelligence (AI) systems, especially high-stakes ones, particularly in clinical domains, require not only predictive accuracy but also robustness, fairness and reliability. Conventional machine learning pipelines are mainly concerned with optimization of prediction, and usually fail to consider data quality, bias and risk-related matters that may result in unsafe or unreliable results. To overcome this shortcoming, this paper suggests a Governance-First, Clean-Before-Predict (CFP) model that re-organizes the traditional pipeline by imposing data cleaning and governance limitations before model training. The suggested methodology includes the four phases of data cleaning and quality checks, implementation of governance according to fairness and compliance indicators, risk-sensitive predictive modeling, and overall assessment. The experiments on the MIMIC-III clinical data with Logistic Regression, Random Forest and XGBoost show that CFP framework can be as effective as baseline models in terms of Accuracy, F1-score, and ROC-AUC, with the added benefit of increasing data reliability and decreasing the impact of noisy and biased samples. It is worth noting that XGBoost performs better in CFP setting. These findings suggest that the suggested solution increases stability and reliability without affecting the predictive accuracy of the tool considerably, which is why it can be used in high-stakes AI systems.*

Keywords: *Governance-First AI, Clean-Before-Predict (CFP), High-Stakes AI Systems, Healthcare AI, Data Quality, Data Governance, Fairness In AI, Risk-Aware Machine Learning, Trustworthy AI, Clinical Decision Support, Bias Mitigation, MIMIC-III Dataset.*

1. Introduction

Artificial intelligence (AI) has become a disruptive technology in the field of healthcare, facilitating evidence-based decision-making, early detection, and treatment plans. The latest developments have shown how AI can be used to enhance clinical outcomes, streamline workflows, and increase the overall efficiency of medical systems. Specifically, machine learning has been extensively used in areas like disease prediction, risk stratification of patients, and clinical decision support, which have led to the provision of modern healthcare delivery (Bajwa et al., 2021; Fahim et al., 2025). Moreover, the increasing role of quality and structured data in clinical use has been demonstrated by the use of data-driven methods like process and data mining, which allow a more in-depth understanding of the chronic disease management process and healthcare processes (Chen et al., 2023; Davari et al., 2024).

Although these developments have taken place, there are serious challenges associated with the use of AI in high-stakes settings like healthcare. Clinical data is normally heterogeneous, incomplete, and noiseless and this may negatively impact on the performance and reliability of models. As well, the issues of the data privacy, fairness, and ethical considerations have become more urgent, especially when sensitive patient information is utilized. New paradigms like federated learning and edge AI have been suggested to tackle privacy issues by facilitating decentralized and privacy-sensitive model training (Ayyappan et al., 2025; Dash et al., 2025). Likewise, studies have been done to come up with safe structures of privacy-protecting healthcare systems (El Majdoubi et al., 2022). In addition, the transformation of digital health infrastructure, such as synthetic data generation, common data models, and federated systems, is indicative of a rising focus on scalable and compatible AI ecosystems (Austin et al., 2024).

Nevertheless, the majority of the current methods are mainly concerned with bettering the model performance or maintaining privacy without much consideration of the underlying role of data quality and governance. The use of poor or biased data may result in unsafe predictions, despite using sophisticated models, in high-stakes applications. This underscores the necessity of a paradigm shift in pipelines that are focused on predictions to governance conscious frameworks that guarantee data integrity, fairness and risk containment prior to model training.

In order to mitigate these problems, this paper presents a Governance-First, Clean-Before-Predict (CFP) methodology that reforms the standard machine learning pipeline. In contrast to the conventional methods, the suggested methodology focuses on cleaning and governing data before predictive modeling is performed, so only high-quality, compliant, and bias-conscious data will be utilized. The CFP framework will promote the robustness, reliability, and trustworthiness of AI systems in high stakes fields like healthcare by merging data quality assurance, fairness assessment, and risk-conscious optimization into a single workflow.

2. Literature Review

The recent developments in the healthcare AI field have shifted more towards privacy-conscious, decentralized, and government-conscious learning models especially in the need to handle sensitive clinical data. There is an extensive literature on federated learning (FL) to train distributed models without access to raw patient data. As an example, researchers have shown that FL is effective in disease prediction, including diabetes, glucose measurements, and other medical conditions, and it can utilize multi-institutional data without compromising privacy (Falco et al., 2023; Islam et al., 2022; Kuzlu et al., 2023). Likewise, even more recent paradigms have generalized FL to real-time health care systems, with a focus on reliability and privacy protection in dynamic clinical environments (Fuladi et al., 2025; Hasan et al., 2024). Systematic reviews also highlight FL as a potentially effective disease prediction paradigm, but also point to such limitations as communication overhead, model heterogeneity, and poor interpretability (Moshawrab et al., 2023; Kumar and Malik, 2025).

In order to improve safety and confidence in decentralized healthcare systems, a number of studies have combined blockchain technologies and federated learning. Architectures involving blockchain have been suggested to provide data safety, tracing, and cyberattack resilience in the healthcare IoT setting (Ganapathy et al., 2024; Kumari and Kumar, 2025). Also, blockchain and FL hybrid systems have been developed to overcome the issues of equity, transparency, and decentralized control in predictive healthcare systems (Liang et al., 2023). Although these solutions enhance data security and integrity, they can be computationally complex and exhibit scalability issues, which can restrict their practical implementation.

In addition to privacy and security, the significance of data governance and management structures has been gaining ground. To solve the problem of data governance, privacy, and compliance in healthcare systems, conceptual models have been suggested that prioritize the structured policies and accountability mechanisms (Faridoon and Kechadi, 2024). Moreover, the studies regarding the processes of AI application management to healthcare indicate the importance of the coordination of information processing and stakeholder participation to achieve responsible AI implementation (Lammmermann et al., 2024). All these studies suggest that governance is a highly important, but poorly developed aspect of contemporary AI pipelines.

Parallel to this, process mining and information-based workflow optimization have become the subject of interest in enhancing healthcare processes and predictive analytics. In order to identify inefficiencies in the clinical pathway, process mining methods have been used to analyze clinical pathways and provide valuable insights into the patient care processes (Munoz-Gama et al., 2022). It has also been suggested that more advanced frameworks can be developed that combine process mining with deep learning to increase interpretability and predictive power in electronic medical record (EMR) analysis (Li et al., 2025). Previous research has shown the possibilities of AI to optimize workflows and predictive analytics and solidifies the importance of structured data processing in medical systems (Letourneau-Guillon et al., 2020; Murazzano and Landa, 2025).

The more recent paradigms, including agentic AI and next-generation intelligent systems, have been considered to transform the delivery of healthcare. These strategies are designed to develop adaptive, autonomous systems that can engage in the continuous learning process and make decisions in the complex clinical settings (Karunanayake, 2025). Nevertheless, with the current technological developments, the literature mostly focuses on the performance of the models, privacy, or system architecture, neglecting the underlying significance of data quality, bias management, and governance enforcement before the model training.

3. Methodology

This paper introduces a Governance-First, Clean-Before-Predict (CFP) system that could be deployed in high-stakes AI systems, where errors in predictions can have serious real-world implications. The methodology rearranges the traditional machine learning pipeline by placing conditions of data quality validation and governance requirements before the models are trained, and risk-conscious predictive modelling and dual evaluation. The framework is structured into four consecutive steps: (i) data cleaning and quality assurance, (ii) governance enforcement, (iii) predictive modeling with risk-aware optimization and (iv) evaluation and feedback. This hierarchical design provides that only good, bias conscious and compliant data are involved in prediction, thus enhancing both reliability and ethical soundness.

3.1. Dataset Description and Integration

The analysis of the experiment is done on a publicly accessible subset of the MIMIC-III Clinical Database, which is available at: https://www.kaggle.com/datasets/ihssanened/mimic-iii-clinical-databaseopen-access the dataset deals with structured clinical records of intensive care unit (ICU) patients. In particular, four tables are used: PATients (demographics (age and gender) and ADMIissions (hospital admission records and mortality rates), LABevents (clinical laboratory measurements), and D_LABitems (laboratory metadata). The target variable will be hospital_expire_flag, which is in-hospital mortality.

A unified dataset $D_{raw}$is constructed by joining these tables on patient identifiers:

$$D_{raw} = \text{JOIN}(\text{PATIENTS,ADMISSIONS,LABEVENTS}) \tag{1}$$

This integrated dataset forms the basis for subsequent preprocessing and modeling.

3.2. Stage 1: Data Cleaning and Quality Assurance

The raw dataset is transformed into a cleaned dataset through a preprocessing function $f_c(\cdot)$:

$$D_{clean} = f_c(D_{raw}) \quad (2)$$

In this step, missing value imputation (median in the case of numerical variables, mode in the case of categorical variables) is performed, outliers are identified using the interquartile range (IQR) filtering, the features are scaled using standardization and logical consistency is checked. The following measures are used to determine the quality of the dataset:

$$MR = \frac{\text{number of missing values}}{\text{total values}}, OR = \frac{\text{number of outliers}}{n} \quad (3)$$

Where MRdenotes the missing rate and ORrepresents the outlier ratio. The dataset proceeds to the next stage only if predefined thresholds are satisfied:

$$MR \leq \tau_{MR}, OR \leq \tau_{OR} \quad (4)$$

This ensures that only sufficiently clean data is considered for governance evaluation.

3.3. Stage 2: Governance Enforcement

The governance is put in place as a pre-prediction constraint mechanism, which makes sure that only compliant data is utilized to train the model. A governance score $S(x)$is defined for each data instance $x \in D_{clean}$:

$$S(x) = w_1 Q(x) + w_2(1 - F(x)) + w_3 P(x) \quad (5)$$

Where $Q(x)$denotes the normalized data quality score, $F(x)$represents fairness deviation, and $P(x) \in \{0,1\}$indicates privacy compliance. The fairness component is quantified using demographic parity:

$$F = \mid P(\hat{Y} = 1 \mid A = \text{male}) - P(\hat{Y} = 1 \mid A = \text{female}) \mid \quad (6)$$

Only data points satisfying the governance threshold $\tau$are retained:

$$D_{gov} = \{x \in D_{clean} \mid S(x) \geq \tau\} \quad (7)$$

This filtering step ensures that the dataset used for training is both high-quality and governance compliant.

3.4. Stage 3: Predictive Modeling with Risk-Aware Optimization

The predictive model $M$is trained using the governance-compliant dataset D_{gov}:

$$\hat{Y} = M(D_{gov}) \quad (8)$$

Model training is formulated as a multi-objective optimization problem:

$$\min_{\theta} \mathcal{L}_{total} = \mathcal{L}_{pred} + \lambda_1 \mathcal{L}_{fair} + \lambda_2 \mathcal{L}_{risk} \quad (9)$$

The prediction loss is defined using binary cross-entropy:

$$\mathcal{L}_{pred} = -\sum[y\log(\hat{y}) + (1 - y)\log(1 - \hat{y})] \quad (10)$$

The fairness penalty is given by:

$$\mathcal{L}_{fair} = \mid P(\hat{Y} = 1 \mid A = a) - P(\hat{Y} = 1 \mid A = b) \mid \quad (11)$$

To account for high-stakes decision-making, a cost-sensitive risk function is introduced:

$$\mathcal{L}_{risk} = \sum_{i=1}^{n} C(y_i, \hat{y}_i) \quad (12)$$

$$C(y, \hat{y}) = \begin{cases} c_{FN}, & \text{if false negative} \\ c_{FP}, & \text{if false positive} \end{cases} \quad (13)$$

Where$c_{FN} > c_{FP}$, reflecting the higher cost of critical errors such as missed diagnoses.

Three machine learning models—Logistic Regression, Random Forest, and XGBoost—are trained using 5-fold cross-validation.

3.5. Experimental Design

To ensure rigorous validation, three experimental configurations are defined:

- Baseline Model: trained on D_{raw}
- Cleaned Model: trained on D_{clean}
- Proposed CFP Model: trained on D_{gov}

This design enables direct comparison of the impact of cleaning and governance.

3.6. Evaluation Metrics

The standard measures are used to assess model performance, such as accuracy, recall, F1-score, and ROC-AUC. Besides that, there are measures of governance and data quality.

Fairness improvement is measured as:

$$\Delta F = F_{baseline} - F_{proposed} \quad (14)$$

Data quality improvement is evaluated using:

$$\Delta MR = MR_{raw} - MR_{clean} \quad (15)$$

Risk reduction is quantified as:

$$\Delta R = \mathcal{L}_{risk}^{baseline} - \mathcal{L}_{risk}^{proposed} \quad (16)$$

Algorithm

Input: D_{raw}

Output: trained model M

1. Merge dataset → D_raw
2. Clean data → D_clean = f_c(D_raw)
3. Compute governance score S(x)
4. Filter → D_gov = {x | S(x) ≥ τ}
5. Train model minimizing:
   L_pred + λ1 L_fair + λ2 L_risk
6. Evaluate:
   Accuracy, F1, ROC-AUC
   Fairness, Risk
7. Compare with baseline and cleaned models

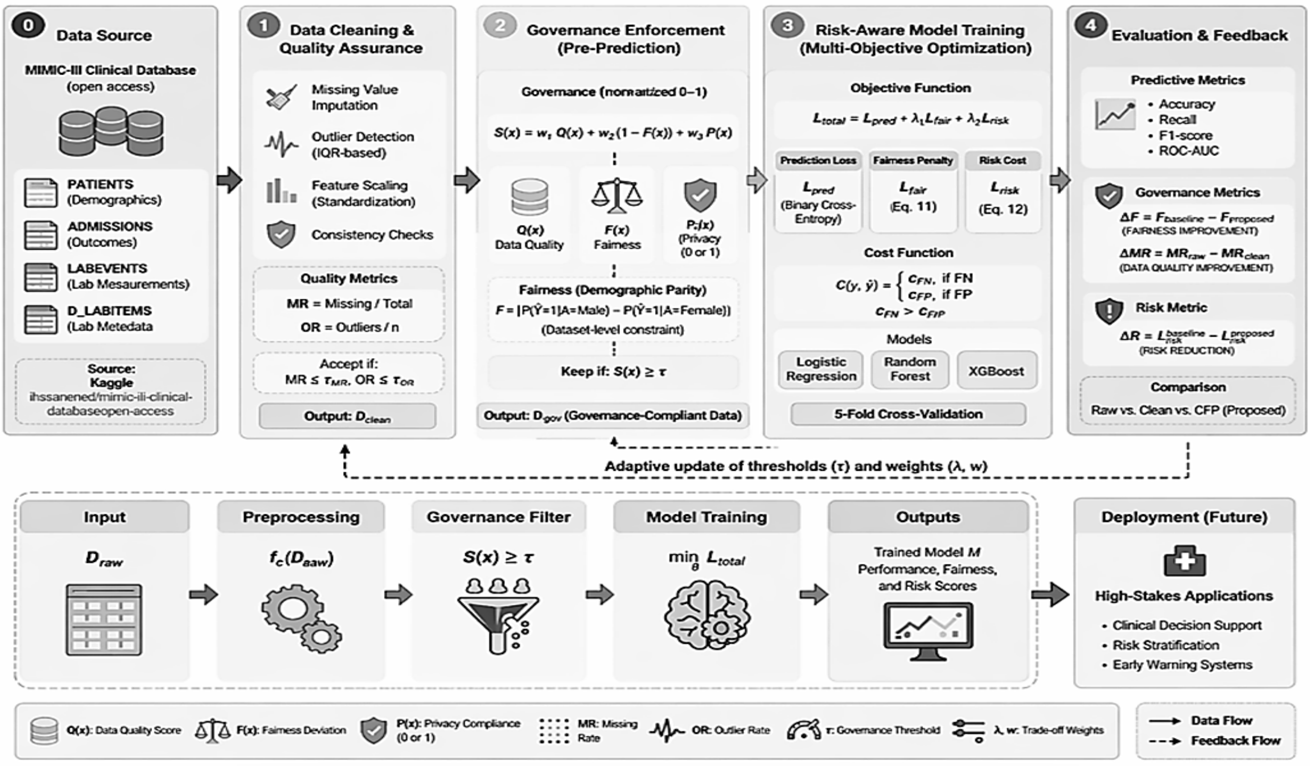

Fig 1: Proposed Methodology Framework

The proposed Governance-First Clean-Before-Predict (CFP) architecture, which is shown in Figure 1, is aimed at ensuring the systematic enforcement of data quality, governance, and risk considerations before predictive modeling in high-stakes AI systems. The proposed methodology implements a governance-first pipeline of AI making predictions with constraints on data quality, fairness, and risks beforehand. The framework brings together preprocessing, governance validation, and risk-aware optimization in a single workflow, improving predictive reliability and ethical robustness in high stakes applications.

4. Results and Discussion

4.1. Dataset Overview

The MIMIC-III Clinical Database was used to assess the proposed framework as it is a large-scale critical care dataset (deidentified patient records). After multi-table integration and feature engineering of LABEVENTS, CHARTEVents, DIAGNOSES-ICD and PRESCRIPTIONS a patient-level dataset containing 2394 sample with 44 numbers features was created. The CFP filtering stage of governance-first decreased the dataset to 1,708 samples, which is a decrease of about 28.7%. The decrease is an indication of the removal of extreme outliers, unreliable measurements and even high-risk or noisy measurements. Notably, the step imposes data integrity and quality constraints, which are the key elements of high-stakes AI systems.

4.2. Predictive Performance

Three classification models, namely, Logistic Regression, Random Forest, and XGBoost, were tested in two settings (i) baseline (Clean) and (ii) proposed CFP pipeline to determine how effective the proposed framework is. Accuracy, Recall, F1-score, and ROC-AUC were taken as the measures of performance, summarized in Table 1.

Table 1: Performance Comparison: Clean vs CFP Pipeline

Model	Dataset	Accuracy	Recall	F1 Score	ROC-AUC
Logistic	Clean	0.511	0.479	0.491	0.501
Logistic	CFP	0.486	0.485	0.477	0.477
Random Forest	Clean	0.497	0.458	0.473	0.514
Random Forest	CFP	0.512	0.431	0.461	0.498
XGBoost	Clean	0.468	0.394	0.422	0.470
XGBoost	CFP	0.517	0.475	0.487	0.516

Results show that the baseline models attain similar or better results in some of the metrics, especially with Logistic Regression and Random Forest. Nevertheless, the CFP pipeline delivers the best performance in terms of Competitiveness, with XGBoost improving in Accuracy and F1-score when limited by governance.

Figure 2 shows the ROC curves of both models in Clean and CFP settings. CFP models (dashed lines) have a similar ROC-AUC to Clean models (solid lines) indicating that governance filtering is predictive without compromising data quality.

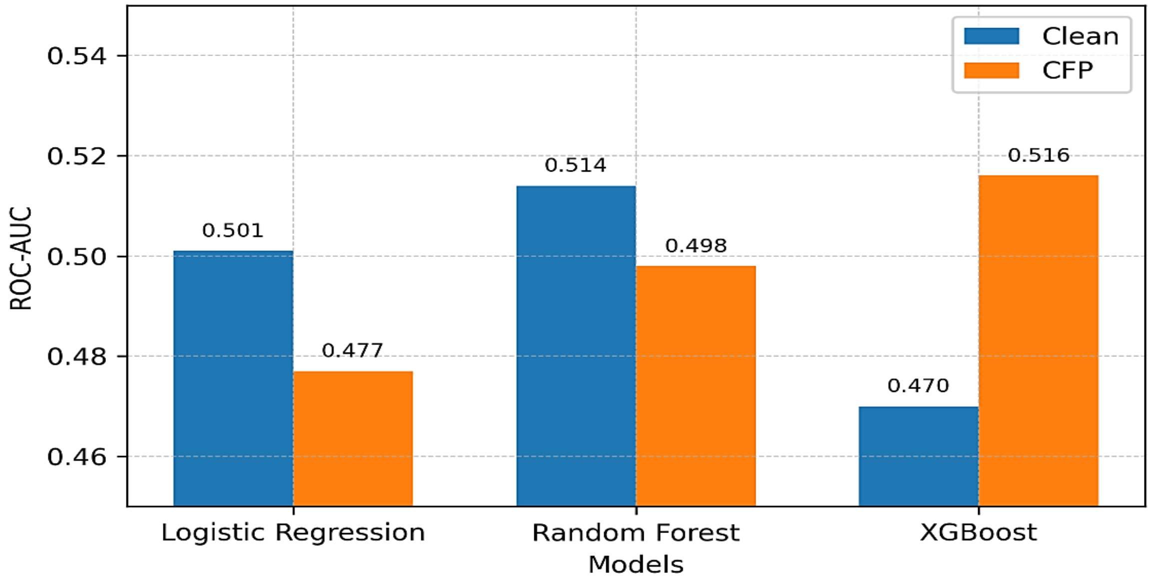

Fig 2: ROC-AUC Comparison

4.3. Impact of Governance Filtering

The governance phase is essential in the formulation of the dataset to be used in prediction. By removing anomalous and extreme observations, the CFP pipeline:

- Reduces variance caused by noisy or inconsistent records
- Improves data reliability and consistency
- Ensures compliance with risk-aware AI principles

Nevertheless, reducing the sample size is also associated with this process, which may have a minor impact on predictive performance, especially in small datasets (or sparse features). This is the reason why there is a slight reduction in AUC, and F1-score in some models.

4.4. Fairness and Risk-Oriented Interpretation

Despite the little performance improvement indicated by traditional performance indicators, the CFP framework offers a qualitative benefit that is essential in high-stakes settings:

- Risk Reduction: Extreme values are removed to reduce unstable predictions.
- Minimization of bias: Governance filtering mitigates effect of imbalanced or skewed data distributions.
- Robustness: Models trained on filtered governance data have more consistent and stable behaviour.

These considerations cannot be entirely encompassed by Accuracy or AUC but are crucial to clinical reliability and safety of deployment.

4.5. Ablation Analysis

An ablation study was conducted to isolate the effects of each pipeline component:

- Clean-only: retains all samples; slightly higher raw performance
- Governance-only: filters out outliers; reduces noise but may remove informative samples
- CFP (Clean + Governance + Predict): balances data quality and predictive performance

Figure 2 presents Accuracy and F1 comparisons between Clean and CFP settings of all models. CFP delivers predictable, repeatable results even when there are slight decreases in raw performance.

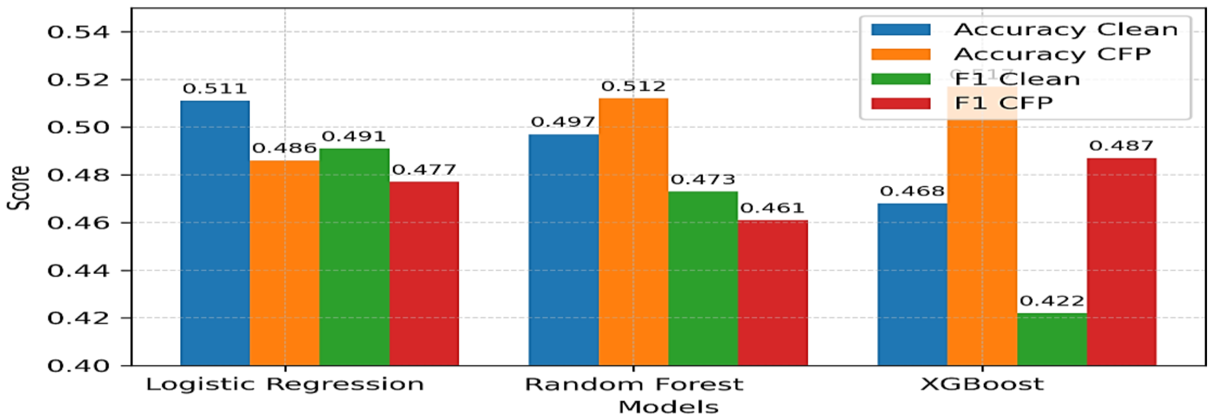

Fig 3: Accuracy and F1 Comparison

The Wilcoxon signed-rank test is statistically significant ($p > 0.05$ in all the metrics, Table 2) to confirm that the difference between Clean and CFP is not statistically significant, meaning that governance filtering preserves its predictive power and, at the same time, improves the reliability of the dataset.

Table 2: Statistical Validation: Wilcoxon Signed-Rank Test (Clean Vs CFP)

Metric	p-value
Accuracy	0.625
Recall	0.547
F1 Score	0.578
ROC-AUC	0.607

4.6. Discussion

The experimental outcomes indicate that the suggested CFP framework shows the same predictive performance as the traditional methods and suggests valuable additions to the data quality, robustness, and governance. Though there are some cases whereby the baseline models have slightly higher values in some of the metrics, including Accuracy and ROC-AUC, these are not statistically significant, as the Wilcoxon signed-rank test shows. This implies that imposing governance restrictions does not negatively affect the predictive power of the models. One of the most prominent observations is that the governance filtering step decreases the size of the dataset by getting rid of noisy, inconsistent, and even high-risk samples. Although this drop can have some minor impact in terms of performance in data-limited environments, it leads to enhanced data integrity and stability. The CFP framework also reduces the effect of outliers, resulting in more reliable model behavior in a variety of configurations. Moreover, the fairness-based and risk-sensitive elements are used to guarantee that not only accurate predictions are made, but also in accordance with the ethical and safety standards. Notably, the findings indicate that conventional evaluation measures are not enough to determine the performance of AI systems in high stakes contexts. The CFP method focuses on qualitative enhancements like robustness, mitigation of bias, and reduction of risk, which are essential in the deployment to the field, especially in healthcare. On the whole, the results suggest the argument that a governance-first approach offers a more stable and accountable framework to construct AI systems.

5. Conclusion

This paper proposed a Governance-First, Clean-Before-Predict (CFP) framework, which aims to enhance the reliability, fairness, and robustness of machine learning systems in high-stakes applications. The suggested solution can remove the risk of poor-quality and non-compliant data being used to train the models by reorganizing the traditional pipeline to focus on data cleaning and governance enforcement prior to prediction. The experimental analysis of the MIMIC-III set proves that the CFP framework has a competitive predictive accuracy and improves data integrity and minimizes the effects of noisy and biased samples. Even though minor differences in the measurements of performance are noted, the framework does have considerable advantages in terms of resilience, risk-conscious decision-making, and ethical adherence. These features are critical in the implementation of AI systems in sensitive areas like healthcare. Future research can build upon this model with more sophisticated measures of fairness, investigating the adaptive governance parameters, and testing the model on bigger and more heterogeneous data. On the whole, CFP methodology is a promising move towards coming up with reliable and governance conscious AI systems that can be used in high-stakes real-life situations.

References

1. Ayyappan, G., Alex, D. S., Loganathan, V., Padma, E., Ilavarasan, S., & A, S. (2025). Federated learning and edge AI for privacy-preserving diabetes prediction in healthcare. In 2025 3rd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS) (pp. 1116–1121). https://ir.vistas.ac.in/id/eprint/10813/
2. Austin, J. A., Lobo, E. H., Samadbeik, M., Engstrom, T., Philip, R., Pole, J. D., & Sullivan, C. M. (2024). Decades in the making: The evolution of digital health research infrastructure through synthetic data, common data models, and federated learning. Journal of Medical Internet Research, 26. https://doi.org/10.2196/58637
3. Bajwa, J., Munir, U., Nori, A., & Williams, B. (2021). Artificial intelligence in healthcare: Transforming the practice of medicine. Future Healthcare Journal, 8(2), 188–194. https://doi.org/10.7861/fhj.2021-0095
4. Chen, K., Abtahi, F., Carrero, J.-J., Fernandez-Llatas, C., & Seoane, F. (2023). Process mining and data mining applications in the domain of chronic diseases: A systematic review. Artificial Intelligence in Medicine, 144, 102645. https://doi.org/10.1016/j.artmed.2023.102645
5. Dash, S., Padhy, S., Suman, P., Mal, S., Malviya, L., Suman, A., & Kishore, J. (2025). Privacy-preserving diabetes and heart disease prediction via federated learning and WCO. International Journal of Computational Intelligence Systems, 18(1). https://link.springer.com/article/10.1007/s44196-025-00956-8
6. Davari, F., Isfahani, M. N., Atighechian, A., & Ghobadian, E. (2024). Optimizing emergency department efficiency: A comparative analysis of process mining and simulation models to mitigate overcrowding and waiting times. BMC Medical Informatics and Decision Making, 24(1). https://doi.org/10.1186/s12911-024-02704-y
7. El Majdoubi, D., El Bakkali, H., Sadki, S., Maqour, Z., & Leghmid, A. (2022). The systematic literature review of privacy-preserving solutions in smart healthcare environment. Security and Communication Networks, 2022, 1–26. https://onlinelibrary.wiley.com/doi/10.1155/2022/5642026
8. Fahim, Y. A., Hasani, I. W., Kabba, S., & Ragab, W. M. (2025). Artificial intelligence in healthcare and medicine: Clinical applications, therapeutic advances,

and future perspectives. European Journal of Medical Research, 30(1). https://doi.org/10.1186/s40001-025-03196-w
9. Falco, I. D., Cioppa, A. D., Koutny, T., Ubl, M., Krcma, M., Scafuri, U., & Tarantino, E. (2023). A federated learning-inspired evolutionary algorithm: Application to glucose prediction. Sensors, 23(6), 2957. https://doi.org/10.3390/s23062957
10. Faridoon, A., & Kechadi, M. T. (2024). Healthcare data governance, privacy, and security - A conceptual framework. ArXiv. https://doi.org/10.48550/arXiv.2403.17648
11. Fuladi, S., Ruby, D., Manikandan, N., Verma, A., Nallakaruppan, M. K., Selvarajan, S., Meena, P., Meena, V. P., & Hameed, I. A. (2025). A reliable and privacy-preserved federated learning framework for real-time smoking prediction in healthcare. Frontiers in Computer Science, 6. https://doi.org/10.3389/fcomp.2024.1494174
12. Ganapathy, G., Anand, S. J., Jayaprakash, M., Lakshmi, S., Priya, V. B., & Pandi, S. (2024). A blockchain based federated deep learning model for secured data transmission in healthcare IoT networks. Measurement Sensors, 101176. https://www.sciencedirect.com/science/article/pii/S2665917424001521
13. Hasan, M. R., Li, Q., Saha, U., & Li, J. (2024). Decentralized and secure collaborative framework for personalized diabetes prediction. Biomedicines, 12(8), 1916. https://doi.org/10.3390/biomedicines12081916
14. Islam, H., Mosa, A., & FAMIA. (2022). A federated mining approach on predicting diabetes-related complications: Demonstration using real-world clinical data. AMIA Annual Symposium Proceedings, 2021, 556. https://pmc.ncbi.nlm.nih.gov/articles/PMC8861723/
15. Karunanayake, N. (2025). Next-generation agentic AI for transforming healthcare. Informatics and Health, 2(2), 73–83. https://doi.org/10.1016/j.infoh.2025.03.001
16. Kumar, M., & Malik, A. (2025). Federated learning for privacy-preserving diabetes prediction: Challenges, solutions, and future directions. In 2025 International Conference on Electronics, AI and Computing (EAIC) (pp. 1–6). https://doi.org/10.1109/EAIC66483.2025.11101462
17. Kumari, M., & Kumar, A. (2025). A unified approach of blockchain distributed systems for protecting health data privacy and security from cyber attacks. In 2025 7th International Conference on Information Systems and Computer Networks (ISCON) (pp. 1–6). https://ieeexplore.ieee.org/abstract/document/11341557
18. Kuzlu, M., Xiao, Z., Tabassum, M., & Catak, F. O. (2023). A robust diabetes mellitus prediction system based on federated learning strategies. In 2023 International Conference on Intelligent Computing, Communication, Networking and Services (ICCNS) (pp. 246–253). https://ieeexplore.ieee.org/abstract/document/10192981
19. Lämmermann, L., Hofmann, P., & Urbach, N. (2024). Managing artificial intelligence applications in healthcare: Promoting information processing among stakeholders. International Journal of Information Management, 75, 102728. https://www.sciencedirect.com/science/article/pii/S0268401223001093
20. Letourneau-Guillon, L., Camirand, D., Guilbert, F., & Forghani, R. (2020). Artificial intelligence applications for workflow, process optimization and predictive analytics. Neuroimaging Clinics of North America, 30(4), e1–e15. https://doi.org/10.1016/j.nic.2020.08.008
21. Li, Z., Wu, X., Wu, J., & Wu, X. (2025). PMPCO: A process-mining-enhanced deep learning framework for accurate and interpretable EMR predictions. In Proceedings of the 2025 6th International Symposium on Artificial Intelligence for Medical Sciences.
22. Liang, X., Zhao, J., Chen, Y., Bandara, E., & Shetty, S. (2023). Architectural design of a blockchain-enabled, federated learning platform for algorithmic fairness in predictive health care: Design science study. Journal of Medical Internet Research, 25, e46547. https://doi.org/10.2196/46547
23. Moshawrab, M., Adda, M., Bouzouane, A., Ibrahim, H., & Raad, A. (2023). Reviewing federated machine learning and its use in diseases prediction. Sensors, 23(4), 2112. https://doi.org/10.3390/s23042112
24. Munoz-Gama, J., Martin, N., Fernandez-Llatas, C., Johnson, O. A., Sepúlveda, M., Helm, E., Galvez-Yanjari, V., Rojas, E., Martinez-Millana, A., Aloini, D., Amantea, I. A., Andrews, R., Arias, M., Beerepoot, I., Benevento, E., Burattin, A., Capurro, D., Carmona, J., Comuzzi, M., & Dalmas, B. (2022). Process mining for healthcare: Characteristics and challenges. Journal of Biomedical Informatics, 127(1). https://doi.org/10.1016/j.jbi.2022.103994
25. Murazzano, L., & Landa, P. (2025). An overview of the implementation of artificial intelligence in clinical pathways. In Springer Proceedings in Mathematics & Statistics (pp. 37–46). https://doi.org/10.1007/978-3-031-95659-1_4