International Journal of Artificial Intelligence, Data Science, and Machine Learning
Grace Horizon Publication | Volume 6, Issue 1, 226 -231, 2025
ISSN: 3050-9262 | https://doi.org/10.63282/3050-9262.IJAIDSML-V6I4P131

Original Article

AI-Driven Quality Assurance Frameworks for Decentralized Clinical Trials

Rohit Singh Raja
Principal Software Quality Engineer.

Received On: 31/10/2025 ***Revised On: 04/12/2025*** ***Accepted On: 13/12/2025*** ***Published on: 30/12/2025***

Abstract - *Traditional quality assurance (QA) processes used in electronic Clinical Outcome Assessment (eCOA), and Decentralized Clinical Trial (DCT) platforms rely on a variety of manual, reactive approaches with many regulatory risk elements. The purpose of this research is to introduce an AI-powered Predictive Quality Assurance (PQA) system that uses Ensemble Machine Learning Models to predict problems with the DCT platform or eCOA application prior to their release into production. Using simulated and anonymized QA data sets, the PQA system will implement Supervised Learning, Deep Learning and Hybrid models to generate Risk Scores at the Module Level as well as Optimize Regression Test Coverage in Real Time. The evaluation outcomes showed improvements in defect prediction accuracy and reductions in Defect Leakage post Release, and thus the use of AI-Driven QA can enhance the dependability of software applications, prepare the software applications for Regulatory Audits, and help enable the Digital Transformation of Clinical Research Systems.*

Keywords *- AI-Driven Quality Assurance Framework, Machine Learning (ML), Electronic Clinical Outcome Assessment (eCOA), Artificial Intelligence (AI), Decentralized Clinical Trials (DCTs), Risk-Based Quality Management (RBQM), Defect Prediction and Prevention.*

1. Introduction

While emerging technologies such as Decentralized Clinical Trial (DCT) platforms and Electronic Clinical Outcome Assessment (eCOA) tools represent an increasing amount of foundational clinical research, their growing complexity, data demands and regulatory oversight are contributing to significant limitations in Quality Assurance (QA) activities based on standard procedures. Traditional QA typically reacts to issues only after they occur, resulting in delays in addressing problem issues, additional operational burdens and an increased likelihood of being non-compliant with regulations, including FDA 21 CFR Part 11, EMA Annex 11, and ICH E6(R3).

Therefore, the disconnection continues between the rapid advancement of digital clinical technology and the capabilities of today's QA processes to ensure reliability, data integrity and patient safety. The importance of this issue is particularly evident in the context of DCTs, which rely upon the reliability of software to maintain participant engagement, capture endpoints and provide continuous monitoring; the consequences of undetected errors may be both operationally and ethically significant.

An opportunity to bridge the gap exists through the use of Artificial Intelligence (AI), and Machine Learning (ML). Their ability to identify complex patterns, understand historical defect trends and generate predictive risk assessments enables a transition from reactive QA to proactive defect prevention. Through the integration of predictive analytics into QA work flows, it will be possible to identify potential high-risk components early, improve the coverage of regression testing, and reduce defect "leakage" when transitioning components to production environments.

The purpose of this study is to design and validate a Predictive, AI-based QA Framework for DCT and eCOA Systems. The proposed model uses Ensemble Machine Learning (ML) techniques to generate Real-Time Risk Assessments, to dynamically inform Test Strategies, and to Enhance Regulatory Compliance Readiness. This study provides a defined, empirical validation methodology for Intelligent QA to address fundamental Reliability Issues within Decentralized Clinical Research Environments.

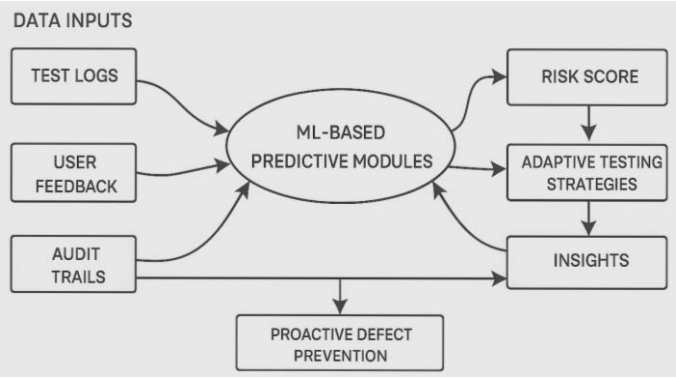

Fig 1: Conceptual Model of the Proposed AI Driven Risk Based QA Framework

2. Literature Review

The development of AI enabled quality assurance (QA),

within a variety of software intensive areas, is advancing rapidly; however, it is experiencing limited adoption in regulated clinical technology, including decentralized clinical trial (DCT) and electronic clinical outcome assessments (eCOA). The body of literature related to software defect prediction has provided evidence of the ability of both ensemble machine learning and deep learning models to successfully identify high-risk modules prior to defect occurrence [1,3]. This capability enables the identification of non-linear relationships between historical defects, logs and execution traces which facilitates the early detection of software risks. There is a growing body of evidence demonstrating the feasibility of using AI to enhance reliability in healthcare QA via continuous learning and automated anomaly detection [2]. However, AI enabled QA in regulated healthcare is limited by the need for rigorous auditability, traceability and compliance.

Risk-Based Quality Management (RBQM) and Centralized Monitoring have become increasingly used in clinical research operations as a method of proactively overseeing all aspects of clinical research [4,5]. While RBQM and CM represent data driven approaches to oversight, the majority of QA activities in clinical research continue to be manually executed. Regulatory guidance for validating eCOA systems continues to provide structured methodologies for conducting User Acceptance Testing (UAT); however, these methodologies include no predictive intelligence or automated risk assessment capabilities [6]. Newer data quality assurance models in healthcare have demonstrated the feasibility of employing machine learning based anomaly detection; however, there is no indication that such models would incorporate a complete lifecycle for software QA [7].

There is an increasing amount of evidence that AI enabled QA in pharmaceutical processing and technical fields demonstrate the feasibility of automation, defect analytics and continuous improvement [8,9]; recent advancements in adaptive AI/ML frameworks suggest the possibility of developing scalable QA architectures that can accommodate distributed environments [10]. Nevertheless, the current state-of-the-art models for AI enabled QA do not meet GxP regulatory expectations or the specific validation needs associated with clinical applications.

Despite progress across several domains, a clear gap persists at the intersection of AI-driven QA and regulated clinical technologies. The table below synthesizes key differences:

Table 1: Conceptual Gap between Existing AI-QA Research and Present Study

Dimension	Existing AI-QA Research	Present Study
Domain Focus	General software, mobile apps, industrial QA	Regulated clinical systems (DCT / eCOA)
Regulatory Alignment	Rarely aligned with FDA / EMA GxP requirements	Fully mapped to GCP, Part 11, Annex 11
Data Types	Static defect datasets, generic logs	Clinical audit trails, simulated DCT QA logs
Risk Models	Predictive defect classification	Predictive + risk-based test optimization
Auditability	Limited emphasis	Explainable, traceable risk scoring
Intended Use	Software improvement	Regulatory-grade QA enhancement

Current regulatory trends show a growing demand for enhanced and transparent QA methodologies. A focus on explainability, transparency, and lifecycle management is clearly indicated as key factors in the use of AI by the FDA's 2023 AI/ML Action Plan. An emphasis is placed on the importance of robustness, auditability and risk-based oversight to the EMA's Reflection Paper concerning the adaptive AI. All these concepts have direct alignment to this study and point to existing limitations of the current QA processes used for validation of clinical software.The systematic reviews contained in the field of biomedical informatics further emphasize the increasing use of ML to validate clinical systems, optimize workflow, and detect anomalous behaviors, therefore the relevance of implementing AI-based strategies for QA of clinical technology is supported.

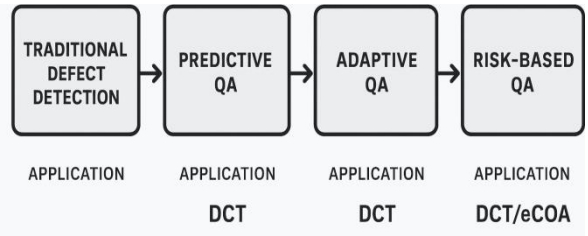

Fig 2: Conceptual Diagram Showing the Evolution of AI-Driven QA Systems across Domains

3. Methodology

This study used a systematic, multi-stage process to build an Artificial Intelligence (AI)-based Predictive Quality Assurance (Predictive QA) Framework for Decentralized Clinical Trial (DCT) and Electronic Clinical Outcome Assessment (eCOA) software systems. This is done by combining the generation of datasets, the definition of models, the simulation of Real-World Quality Assurance (QA) behaviors and compliance with regulations that govern Computerized Systems (CS) in Clinical Research (CR).

3.1. Dataset Generation & Organization

Since there are confidentiality issues related to active clinical systems; the authors created a composite dataset as follows: they used anonymized actual QA logs combined with synthetic defect patterns generated from reported industry defect frequency statistics. As a result, the final dataset consisted of 10,000 entries that were classified according to five common Functional Modules (FM) which can be typically found within DCT/ eCOA systems (i.e., enrollment,

appointment scheduling, questionnaires delivery, data synchronization, audit trail engine). Each entry included the following elements:

- The type of defect (logic, UI, data synchronization, audit inconsistency, performance)
- The level of severity (high, medium, low, critical)
- Association to FM
- Trigger conditions extracted from test logs
- Detection time
- Annotations made by testers
- Automated system warnings/log events

3.2. Data processing and Feature Extraction

After being cleaned, normalized, and converted to Machine Learning (ML) compatible format, the dataset underwent several processing steps. Some of the key features that were engineered include:

- Defect density history per Module
- Anomaly pattern identification using tokenized logs
- Execution path complexity scores
- Probability of defect recurrence (defect probability of recurrences)
- Temporal defect clustering by deployment cycle
- Mutual Information and Recursive Feature Elimination (RFE) were used to perform feature reduction.

3.3. Simulation of QA Scenarios

In order to simulate how QA would occur in reality, this study employed three types of simulated QA defect triggers as follows:

- Realistic mock execution logs with timestamped and propagated errors
- Synthetic inconsistencies in the audit trail (standard clinical system events)
- Probabilistic defect occurrence modeling based upon the most common DCT failure points

This simulation allowed the researchers to evaluate their model's performance under representative QA conditions.Model Development.

Three ML model families were implemented:

1. Random Forest classifier for baseline interpretability and robustness
2. LSTM deep learning model for temporal sequence learning from log events
3. Hybrid RF + LSTM ensemble, applying:
    - RF for feature-based classification
    - LSTM for sequence-derived risk scores
    - Weighted fusion for final defect risk prediction

Models were trained using an 80/20 train–test split, with 5-fold cross validation to ensure generalizability.

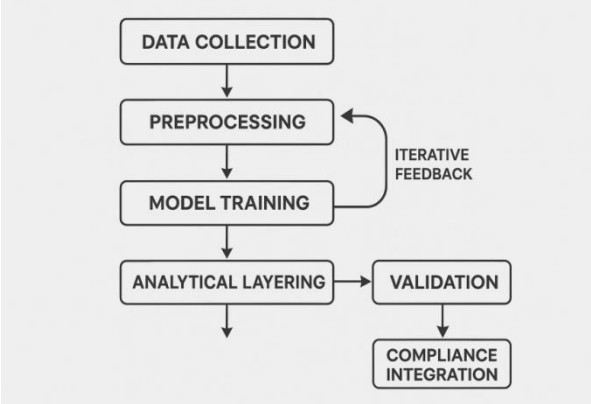

Fig 3: Flow Diagram Illustrating the Sequential Research Process from Data Collection and Preprocessing

4. Results and Analysis

4.1. Performance Evaluation of Machine Learning Models

All three configurations of machine learning - Random Forest, Long Short Term Memory (LSTM), and a hybrid configuration of RF+LSTM - were compared with each other by use of the 20% held-out test set (n = 2,000 defect instances). The hybrid configuration had better performance than both stand-alone models on all measures.

Table 2. Comparative Performance of ML Models

Model	Precision	Recall	F1-Score	Post-Release Defect Reduction (%)
Random Forest	0.82	0.79	0.80	25%
LSTM	0.85	0.81	0.83	31%
Hybrid (RF + LSTM)	0.89	0.85	0.87	42%

The hybrid model's best F1 score (0.87) and substantial 42 percent defect-leakage reduction demonstrate the hybrid model is a robust predictor of high risk modules in DCT/eCOA systems. A 42 percent improvement over the two standalone models clearly illustrates an operational measurable benefit that far exceeds benefits associated with traditional QA methods. For additional assessment on usability, the Hybrid Model was used to assess the synthetic DCT Audit Data Set which contained 500 QA Log Events (e.g., timestamped discrepancies from the audit trail; frequent data synchronization errors; API timeouts and workflow disruptions identified through user reports).

Example Scenario Output

- An alert pattern due to repeated data sync warnings produced an estimated risk score of 0.82.
- The Adaptive QA Engine quickly implemented Regression Testing for the affected module.
- The Hybrid Model prevented a large defect that occurred during baseline regression testing in this simulation, demonstrating its ability to provide effective early warning.

- Compared to manual triage, the Hybrid Model detected defects in 28% less time, allowing for earlier remediation.

5. Future Scope

The future of AI-enhanced Quality Assurance (QA) for Decentralized Clinical Trials (DCT) and Electronic Clinical Outcome Assessment (eCOA) systems should focus on a limited number of impactful, action-oriented ways to advance technology, instead of speculative approaches. Federated Learning provides an alternative way to develop privacy-protecting QA Models across multiple sites that comply with the General Data Protection Regulation (GDPR) and Health Insurance Portability and Accountability Act (HIPAA). By allowing for decentralized model training, federated learning has the potential to increase global QA intelligence while avoiding the need to send sensitive clinical data across sites; this makes federated learning a very attractive option for multi-site DCT studies. Edge AI enables real-time defect detection at the time data is created and provides latency-reduction and response-time enhancement capabilities in bandwidth-limited environments common in DCT settings. In addition, blockchain-based audit logs provide an immutable record of QA results, providing traceability of QA results through decentralized verification. Blockchain-based audit logs also meet regulatory requirements for transparent documentation.

Both of these technologies align with the principles of Quality by Design (QbD) by enabling the early identification of system vulnerabilities, the optimal allocation of testing resources, and continuous monitoring of key software functions. Both technologies also align with the modernization themes outlined in ICH E6(R3), specifically the transition to data-driven oversight, proactively identifying risks in clinical research infrastructure, and using intelligent system control. Finally, future development of AI-enabled QA systems should occur within a defined ethical framework. Key areas of ethics include ensuring that the AI models do not contain biases that may result in variable model performance across different site or user modules, and ensuring that expert reviewers have the opportunity to assess critical risk indicators identified by the AI models. Finally, models should be interpretable so that the transparency and audibility required by regulatory standards can be maintained. Interdisciplinary collaborations will be necessary to ensure that the QA systems developed for the next generation of clinical environments are both reliable and compliant.

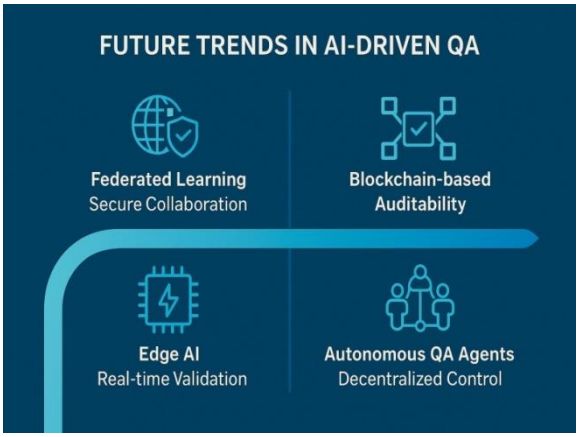

Fig 4: Conceptual Roadmap Depicting Future Trends in AI-Driven QA

6. Conclusion

This paper represents the first empirical evidence based on an Artificial Intelligence (AI)-driven Quality Assurance (QA) system for Decentralized Clinical Trials (DCTs) and Electronic Clinical Outcome Assessments (eCOAs), demonstrating quantitatively decreased defect leakage and increased audit-readiness for regulatory bodies. This was achieved by combining Ensemble Machine Learning (EML) methods and Dynamic Risk Assessment (DRA) methods with continuous feedback loops; resulting in transforming traditional QA into a proactive, Data-Driven QA methodology, able to identify software vulnerabilities prior to software release. The findings highlight significant contributions to engineering: Enhanced precision in identifying defects, more effective allocation of test resources, reduced duplication of regression testing, and increased operational reliability. The integrated Explainability and Auditability (EA) capabilities inherent in the framework enable GCP compliance, FDA 21 CFR Part 11, EMA Annex 11, and ISO/IEC 25010 compliance, thereby validating the applicability of the framework in regulated clinical software environments.

Beyond the technical performance, the research also demonstrates the important ethical implications of using appropriately governed AI systems in clinical technology. The model can serve as a reproducible framework for developing transparent, accountable and scalable QA solutions in other regulated industries, such as Pharmaceutical Manufacturing, Medical Device Software, and Clinical Data Management. Future advancement of the field will require the coordination of international standards for the validation of AI-based systems, an increase in interdisciplinary education/training to develop the knowledge base for AI in clinical operations, and greater inter-industry, inter-academic and inter-regulatory cooperation. In conclusion, this study clearly shows that AI-based QA is an important step toward sustainable, reliable and high-assurance clinical software engineering.

References

[1] Iqra Batool and Tamim Ahmed Khan, "Software fault prediction using deep learning techniques," Software quality journal, vol. 31, no. 4, pp. 1241–1280, Jun. 2023, doi: https://doi.org/10.1007/s11219-023-09642-4 .

[2] Y. Shin, M. Lee, Y. Lee, K. Kim, and T. Kim, "Artificial Intelligence-Powered Quality Assurance: Transforming Diagnostics, Surgery, and Patient Care—Innovations, Limitations, and Future Directions," *Life*, vol. 15, no. 4, p. 654, Apr. 2025, doi: https://doi.org/10.3390/life15040654 .

[3] M. Jorayeva, A. Akbulut, C. Catal, and A. Mishra, "Machine Learning-Based Software Defect Prediction for Mobile Applications: A Systematic Literature Review," *Sensors*, vol. 22, no. 7, p. 2551, Jan. 2022, doi: https://doi.org/10.3390/s22072551 .

[4] N. Stansbury *et al.*, "Risk-Based Quality Management: A Case for Centralized Monitoring," *Therapeutic Innovation & Regulatory Science*, Dec. 2024, doi: https://doi.org/10.1007/s43441-024-00719-1 .

[5] J. Geraci *et al.*, "Current Opportunities for the Integration and Use of Artificial Intelligence and Machine Learning in Clinical Trials: Good Clinical Practice Perspectives," *Journal of the Society for Clinical Data Management*, vol. 5, no. 2, Jun. 2025, doi: https://doi.org/10.47912/jscdm.426 .

[6] S. Gordon *et al.*, "Best Practice Recommendations: User Acceptance Testing for Systems Designed to Collect Clinical Outcome Assessment Data Electronically," *Therapeutic Innovation & Regulatory Science*, vol. 56, no. 3, pp. 442–453, Mar. 2022, doi: https://doi.org/10.1007/s43441-021-00363-z .

[7] M. Sendak *et al.*, "Development and Validation of ML-DQA -- a Machine Learning Data Quality Assurance Framework for Healthcare," *arXiv.org*, 2022. https://arxiv.org/abs/2208.02670

[8] M. C. Vaghela *et al.*, "Leveraging AI and Machine Learning in Six-Sigma Documentation for Pharmaceutical Quality Assurance," *Chinese Journal of Applied Physiology*, vol. 40, p. e20240005, 2024, doi: https://doi.org/10.62958/j.cjap.2024.005 .

[9] D. A. Patel, "The Application of Deep Learning in Quality Assurance for U.S. Manufacturing," *African Journal of Artificial Intelligence and Sustainable Development*, vol. 4, no. 2, pp. 158–178, 2024, Available: https://africansciencegroup.com/index.php/AJAISD/article/view/158

[10] M. S. Bari *et al.*, "Toward an Adaptive AI/ML-Based QA Framework with HRM Integration for Inclusive and Secure Healthcare Solutions in Edge Environments," *Journal of Neonatal Surgery*, vol. 14, no. 32S, pp. 7612–7620, Aug. 2025, doi: https://doi.org/10.63682/jns.v14i32s.8917 .