

AI-Driven Quality Risk Management for Contract Manufacturing Oversight: An ML Approach

Kiran Kumar Gande

*Senior Director-Regulatory Affairs and
Pharmacovigilance, Quagen Pharmaceuticals*

New Jersey, USA

Agenda

- Background & Context
- Motivation for the Study
- Problem Statement
- Research Objectives
- Related Work & Research Gaps
- Dataset & Feature Engineering
- Methodology & Training Strategy
- Results & Performance Analysis
- Discussion & Practical Implications
- Limitations, Future Work & Conclusion
- References

Background & Context

- Contract Manufacturing Organizations (CMOs) are critical to pharma supply chains
- Quality oversight is increasingly complex due to:
 - Distributed facilities
 - Limited direct control
 - Regulatory scrutiny
- Traditional quality assurance relies on:
 - Periodic audits
 - Reactive controls
- These approaches struggle with multi-factor, non-linear risk patterns

Motivation for the Study

- Quality failures often emerge from interacting factors, not single causes
- Environmental, operational, and personnel variables jointly influence risk
- Existing methods detect issues after deviations occur
- Machine learning enables:
 - Early risk detection
 - Proactive intervention
 - Continuous, data-driven oversight

Problem Statement

- Current quality risk management is:
 - Audit-driven
 - Fragmented
 - Largely reactive
- Limited use of predictive intelligence across CMOs
- Key question:
 - How can multi-parameter manufacturing data be leveraged to predict quality risk *before* failures occur?

Research Objectives

Develop an AI-based predictive framework for quality risk assessment

Evaluate multiple ML algorithms for performance and robustness

Identify key operational and environmental risk drivers

Validate model generalizability across facilities

Assess practical and economic implications of predictive quality systems

Related Work & Research Gaps

Existing work focuses on:

- FMEA, FTA, SPC (reactive methods)
- Single-parameter quality indicators
- Single-site manufacturing environments

Key gaps identified:

- Lack of holistic, multi-parameter risk models
- Limited cross-facility validation
- Insufficient discussion on deployment and regulatory compliance

Dataset & Feature Engineering

- **Dataset:**

- 7,500 batch production records collected across multiple contract manufacturing facilities
- 16 operationally relevant parameters
- Realistic class distribution: ~14.5% high-risk and ~85.5% low-risk batches

- **Feature Groups:**

- Environmental: humidity, temperature stability
- Operational: downtime, throughput rate
- Personnel: staff experience, shift timing
- Historical quality: audit scores, recall history

- **Feature Engineering Highlights:**

- Standardization of continuous variables; outliers retained to preserve real-world behavior
- Interaction features (e.g., humidity $\times$ temperature) and non-linear transformations
- Lagged and rolling features to capture short-term variability
- Correlation analysis identified humidity and staff experience as key predictors

Methodology & Training Strategy

- Algorithms evaluated:
 - Gradient Boosting
 - Random Forest
 - SVC
 - Logistic Regression
- Stratified split:
 - 70% train / 15% validation / 15% test
- 10-fold cross-validation
- Nested CV for hyperparameter optimization
- Stacked ensemble with logistic regression meta-learner

Table 1: Hyperparameter Optimization

Algorithm	Key Hyperparameters	Values Used
Gradient Boosting	Learning Rate, Max Depth, Subsample	0.05, 6, 0.8
Random Forest	Number of Trees, Max Depth, Min Samples Leaf	100, 10, 2
Support Vector Classifier	Kernel Type, C (Regularization), Gamma	RBF, 1.0, 'scale'
Logistic Regression	Penalty, Solver, C (Regularization)	l2, liblinear, 1.0

Results & Performance Analysis

- Gradient Boosting:
 - 99.9% accuracy
 - AUC-ROC = 0.999
- Random Forest:
 - 99.2% accuracy
- Ensemble models outperform linear approaches
- High precision and recall for high-risk batch detection

Table 2: Model Evaluation

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC
Gradient Boosting	99.9	99.7	99.8	99.75	0.999
Random Forest	99.2	98.9	99.1	99.00	0.998
Support Vector Classifier	95.3	94.5	95.0	94.75	0.982
Logistic Regression	92.9	91.2	92.1	91.65	0.957

Results & Performance Analysis

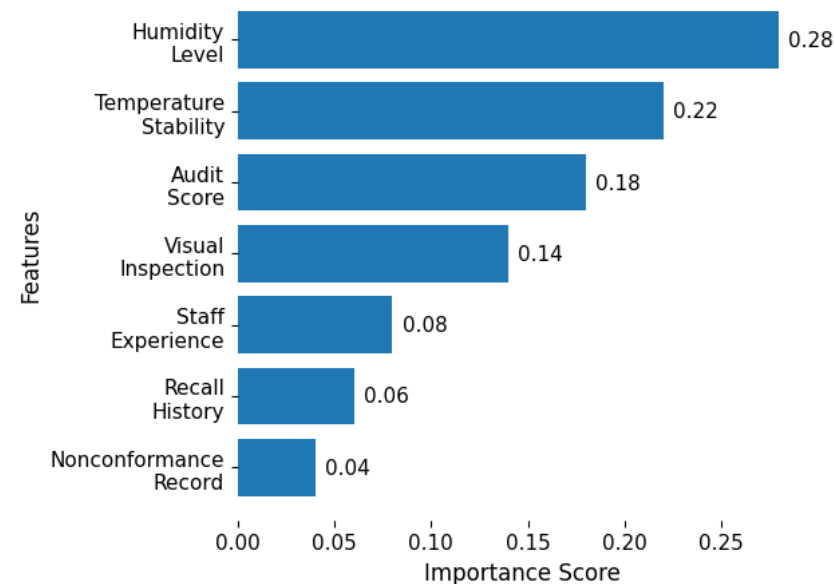


Figure 1: Feature Importance Analysis

Environmental factors, especially humidity and temperature stability, are the dominant predictors of quality risk. Audit and inspection scores also play a significant role, while personnel and historical indicators contribute to a lesser extent. This highlights the importance of real-time operational conditions in risk assessment.

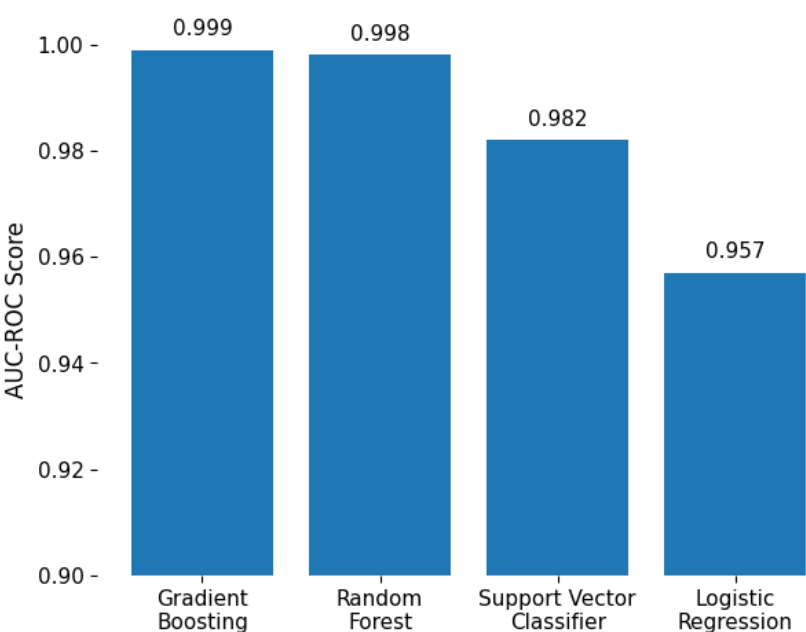


Figure 2: AUC-ROC Model Performance Comparison

Ensemble models demonstrate the strongest ability to distinguish between high- and low-risk batches. Gradient Boosting and Random Forest achieve near-perfect AUC values, outperforming traditional models. These results confirm the effectiveness of ensemble approaches for quality risk prediction.

Discussion & Practical Implications

- Machine learning models effectively capture complex, multi-factor quality risk patterns in contract manufacturing
- Ensemble methods outperform traditional approaches by identifying non-linear interactions across environmental, operational, and quality variables
- Predictive quality risk assessment enables earlier intervention, reducing reliance on reactive audits
- Findings support a shift toward continuous, data-driven quality oversight across distributed manufacturing sites

Limitations, Future Work & Conclusion

- **Limitations**

- Analysis is based on batch-level data with complete sensor availability
- Further validation is needed across additional product types and facilities

- **Future Work**

- Real-time and online learning models
- Integration with MES and ERP systems
- Advanced analytics for multi-modal quality data

- **Conclusion**

- AI-driven quality risk management improves compliance, efficiency, and decision-making in contract manufacturing

References

1. Z. Xu, “Considerations on Regulatory Quality Control in Pharmaceutical Industry,” *Advances in economics, business and management research/Advances in Economics, Business and Management Research*, Jan. 2022, <https://doi.org/10.2991/aebmr.k.220306.047>
2. S. Wang, J. Yang, B. Yang, D. Li, and L. Kang, “An Intelligent Quality Control Method for Manufacturing Processes Based on a Human–Cyber–Physical Knowledge Graph,” *Engineering*, vol. 41, p. 242, Jul. 2024, <https://doi.org/10.1016/j.eng.2024.03.022>
3. H. Tercan and T. Meisen, “Machine learning and deep learning based predictive quality in manufacturing: a systematic review,” *Journal of Intelligent Manufacturing*, vol. 33, no. 7. Springer Science+Business Media, p. 1879, May 28, 2022. <https://doi.org/10.1007/s10845-022-01963-8>
4. J. P. Nelson, J. B. Biddle, and P. Shapira, “Applications and Societal Implications of Artificial Intelligence in Manufacturing: A Systematic Review,” *arXiv (Cornell University)*. Cornell University, Jan. 01, 2023. <https://doi.org/10.48550/arxiv.2308.02025>
5. J. Fabiś-Domagała, M. Domagała, and H. Momeni, “A Matrix FMEA Analysis of Variable Delivery Vane Pumps,” *Energies*, vol. 14, no. 6, p. 1741, Mar. 2021, <https://doi.org/10.3390/en14061741>
6. J. Jayamani, C. C. Janardan, S. V. Appan, K. Kathamuthu, and M. E. Ahmed, “A Practical Tool for Risk Management in Clinical Laboratories,” *Cureus*, Dec. 2022, <https://doi.org/10.7759/cureus.32774>
7. J. E. Agolla, “Smart Manufacturing: Quality Control Perspectives,” in *IntechOpen eBooks*, IntechOpen, 2021. <https://doi.org/10.5772/intechopen.95143>

References

8. Y. Ren, Y. Liu, T. Ji, and X. Xu, “AI Agents and Agentic AI-Navigating a Plethora of Concepts for Future Manufacturing,” 2025, <https://doi.org/10.48550/ARXIV.2507.01376>
9. A. Mitra, “Predictive analytics in quality assurance for assembly processes: lessons learned from a case study at an industry 4.0 demonstration cell.” Jan. 2021. Accessed: Sep. 02, 2025. <https://www.sciencedirect.com/science/article/pii/S2212827121010064>
10. S. Sundaram and A. Zeid, “Artificial Intelligence-Based Smart Quality Inspection for Manufacturing,” *Micromachines*, vol. 14, no. 3, p. 570, Feb. 2023, <https://doi.org/10.3390/mi14030570>
11. P. Ganguly and I. Mukherjee, “Enhancing Retail Sales Forecasting with Optimized Machine Learning Models,” p. 884, Oct. 2024, <https://doi.org/10.1109/icses63445.2024.10762950>
12. X. Zhang et al., “Data-Driven Loan Default Prediction: A Machine Learning Approach for Enhancing Business Process Management,” *Systems*, vol. 13, no. 7, p. 581, Jul. 2025, <https://doi.org/10.3390/systems13070581>
13. V. Hegiste, T. Legler, and M. Ruskowski, “Application of federated learning in manufacturing,” *arXiv (Cornell University)*, Jan. 2022, <https://doi.org/10.48550/arXiv.2208.04664>
14. G. H. Tiwaskar, A. G. Pise, S. Pise, and P. Bhardwaj, “Strategic Analysis of Warning Letters Issued by United State Food and Drug Administration to Pharmaceutical Companies Globally,” *International Journal of Current Research and Review*, p. 118, Jan. 2021, <https://doi.org/10.31782/ijcrr.2021.sp253>