Paper RM06

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

Kiran Kumar Gande, Apozeal Pharmaceuticals, Inc., Fairfield, USA

ABSTRACT

Contract manufacturing management needs proactive quality risk management for compliance, operational efficacy, and product integrity. This research investigated 7,500 batch records spread over several facilities to establish machine learning models for the predictive quality risk analysis. Four algorithms—Gradient Boosting, Random Forest, Support Vector Classifier, and Logistic Regression—were tested using stratified sampling, 10-fold cross-validation, and feature engineering to involve environmental, operational, and historical quality parameters. Gradient Boosting demonstrated the best performance with 99.9% accuracy, 99.7% precision, 99.8% recall, and an AUC-ROC of 0.999, followed by Random Forest (99.2% accuracy). Feature importance identified humidity, temperature stability, audit scores, and staff experience as major quality risk predictors. Cross-facility validation substantiated less than 2% performance loss in new environments. Economic study revealed that predictive interventions would lower quality-related costs by 35–45%, with ROI payback in 12–18 months. Results demonstrate the potential for AI to transform manufacturing management from reactive control to predictive assurance.

INTRODUCTION

Contract manufacturing organizations, or CMOs, play a key role in sustaining efficient processes while helping pharmaceutical and biotech companies meet customer demands. With time, quality risk management is becoming more complex, as original equipment manufacturers now manage distributed manufacturing networks. Simultaneously, sticking to rules, regulations, and standards governs product quality [1]. Quality assurance practices that depend on periodic audits for any future corrections are inadequate to fundamentally deal with the pervasive nature of practices in modern manufacturing [2]. Avoiding quality issues in contract manufacturing reveals many interconnected components, like as environment, process steps, operator capability, and history of past performance on specific provisions. The relationships between these factors are complex and are not straightforward to disentangle. To tackle this, advanced technologies may reveal hidden patterns and forecast quality issues [3]. Machine learning tools create the opportunity to shift quality assurance practices from waiting until after the issues arise to being preventively anticipatory. The introduction of artificial intelligence into the portfolio of production quality practices represents a shift away from older practices, like statistical process control. Intelligent vs . adaptive systems focus on objectification tracking. AI-enabled practices modify methods to utilize much larger operating data, identify complex non-linear relationships, and provide a real-time view of risks with active countermeasures [4]. The value chains of real-time quality improvement are multiplicative too compliance, cost reduction, and supply chain integrity.

LITERATURE REVIEW

TRADITIONAL METHODS OF QUALITY RISK MANAGEMENT

Typical quality risk management techniques for pharma production have included failure mode and effect analysis (FMEA) and fault tree analysis (FTA) methods [5]. While systematic, these methods are inadequate to respond to the more dynamic world of manufacturing and unique factors included in risks. ICH Q9 provides a structure for quality risk management, but not a detailed electronic reporting specifically for using predictive analytics for manufacturing control. Statistical process control (SPC) techniques are often used for quality management where control charts, and capability indices are used to help ensure process stability [7]. These conventional methods exhibit considerable shortcomings when managing intricate manufacturing systems characterized by many interconnected variables and non-linear associations. Conventional SPC techniques are reactive, detecting quality problems solely after they arise, resulting in product recalls and noncompliance with regulatory standards [8].

USES OF MACHINE LEARNING IN MANUFACTURING QUALITY

Recent advancements in machine learning have demonstrated considerable promise for high-quality applications in manufacturing. Support vector machines, random forests, and neural networks were successfully employed for defect prediction and quality categorization tasks [9]. Deep learning methods have shown considerable promise for image-related quality assessment and anomaly identification tasks [10]. Ensemble methods have shown significant success in forecasting manufacturing quality by integrating different models and minimizing the chances of overfitting. Random forest techniques have shown exceptional effectiveness in handling various data forms and offering feature importance assessment for manufacturing [11]. Gradient boosting techniques have performed exceptionally well in predictive maintenance and quality prediction situations [12].

CHALLENGES IN MONITORING MANUFACTURING CONTRACTS

Contract manufacturing presents new quality control issues stemming from geographic dispersion, differing operational norms, and restricted direct oversight of production processes [13]. The ability to monitor remotely and integrate real-time data stands out as key factors for effective quality management. The pharmaceutical industry has faced several notable quality failures in contract manufacturing agreements, highlighting the necessity for enhanced oversight protocols [14]. Regulatory bodies have increasingly emphasized the importance of supplier quality management and continuous oversight of contract manufacturing operations. The FDA guidance regarding supplier verification emphasizes risk-based approaches and continuous monitoring rather than relying solely on periodic audits.

RESEARCH GAPS

RECOGNIZED DEFICIENCIES

Current literature highlights several significant gaps in AI-driven quality risk management for controlling contract manufacturing. Initially, the focus of current research is on individual quality indicators rather than holistic multi-parameter risk assessment models. The consideration of environmental factors, operational metrics, and performance histories in combined predictive models is insufficiently researched. Secondly, a majority of published studies examines quality prediction in regulated laboratory or single-location environments, limiting applicability to complex contract manufacturing networks. The differences in manufacturing settings, machinery, and processes across various CMOs create an additional level of complexity that existing research does not fully account for. Third, current literature pays insufficient attention to the correct implementation and deployment considerations for AI-driven quality systems. The practical challenges of data integration, model maintenance, and compliance validation in operational environments require comprehensive examination.

OPPORTUNITIES FOR ENGAGEMENT

This research addresses the identified gaps by developing a comprehensive multi-parameter risk assessment framework specifically designed for contract manufacturing facilities. The described approach integrates different operational parameters, including batch characteristics, environmental conditions, workforce-related factors, and historical quality metrics into cohesive predictive models. The study contributes to the field by demonstrating successful strategies for applying AI-driven quality systems across various manufacturing environments.

METHODOLOGY

DATA COLLECTION AND PREPROCESSING

The study employed 7,500 batch production records collected from multiple contract manufacturing sites, with 16 operational parameters selected for their relevance to quality risk assessment and practical availability. Preprocessing involved standardizing continuous variables for scale consistency and checking for missing values (none were found). Outliers (0.3% of records) were retained to preserve real-world variability.

Table 1: Dataset Characteristics

Feature Category	Example Attributes	Data Type	Description
Environmental Factors	Humidity, Temperature Stability	Continuous	Real-time sensor data during production
Operational Parameters	Downtime, Throughput Rate	Continuous	Metrics tied to equipment and process efficiency
Personnel Variables	Staff Experience, Shift Timing	Categorical	Attributes related to operator or QA team
Historical Quality	Audit Score, Recall History	Categorical/Numeric	Prior performance indicators

FEATURE ENGINEERING

Feature construction emphasized operational patterns and temporal dynamics. Environmental stability metrics captured temperature and humidity fluctuations, while operational efficiency linked throughput rates with downtime. Correlation analysis revealed humidity as the strongest predictor ($r = 0.67$), followed by staff experience ($r = 0.43$) and batch volume ($r = 0.29$). Advanced engineering introduced interaction features (e.g., combined humidity–temperature effects), polynomial terms for non-linear trends, and rolling/lagged features to capture short-term shifts and cumulative process drift.

MODEL DEVELOPMENT AND TRAINING

Four algorithms were evaluated: Random Forest, Gradient Boosting, Logistic Regression, and Support Vector Classifier (SVC). Stratified sampling split data into training (70%), validation (15%), and testing (15%). Ten-fold cross-validation ensured robust performance estimates.

The predictive model was formulated as:

$$P(Risk = 1|X) = f(x_1, x_2, \ldots, x_{16}) \tag{1}$$

Feature importance in ensemble models was measured using Gini importance:

$$I_j = \sum_{t=1}^{T} p_t \cdot G_t \cdot \delta(v_t = j) \tag{2}$$

Where:

- p_t = proportion of samples reaching node t
- $G_t =$ Gini impurity reduction at node t
- $\delta(v_t = j)$ = indicator function, equal to 1 if feature j is used at node t, and 0 otherwise.

Performance metrics included Accuracy, Precision, Recall, and F1-score:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{3}$$

ENSEMBLE STRATEGY AND HYPERPARAMETER OPTIMIZATION

A stacked ensemble was designed to combine model strengths. Base learners (tree-based, linear, kernel methods) provided diversity, while a meta-learner (regularized logistic regression) optimized the final predictions. Nested cross-validation minimized overfitting.

Table 2: 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

VALIDATION, INTERPRETABILITY, AND COMPLIANCE

Validation included temporal testing (training on past data, predicting future batches) and cross-facility evaluation to ensure model generalization across diverse sites. Statistical tests (McNemar's, bootstrap confidence intervals) confirmed significance, while Monte Carlo simulations stress-tested predictions under noise and variability. Interpretability was achieved through:

- Global: Feature importance and permutation metrics.
- Local: SHAP values for batch-level explanations.
- Surrogate Models: Simplified decision trees for non-technical stakeholders.
- Counterfactuals: Identifying variable adjustments that could shift risk outcomes.

Compliance considerations required extensive documentation of algorithms, validation protocols, and monitoring processes, aligning with FDA and ICH quality guidelines.

RESULTS AND DISCUSSION

COMPARISON OF MODEL PERFORMANCE

Table 3: Model Performance Comparison

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

The evaluation demonstrated excellent predictive accuracy across all algorithms, with ensemble models outperforming traditional methods. Gradient Boosting achieved the highest accuracy (99.9%), followed by Random Forest (99.2%), Support Vector Classifier (95.3%), and Logistic Regression (92.9%). Ensemble approaches proved most effective in capturing complex non-linear relationships.

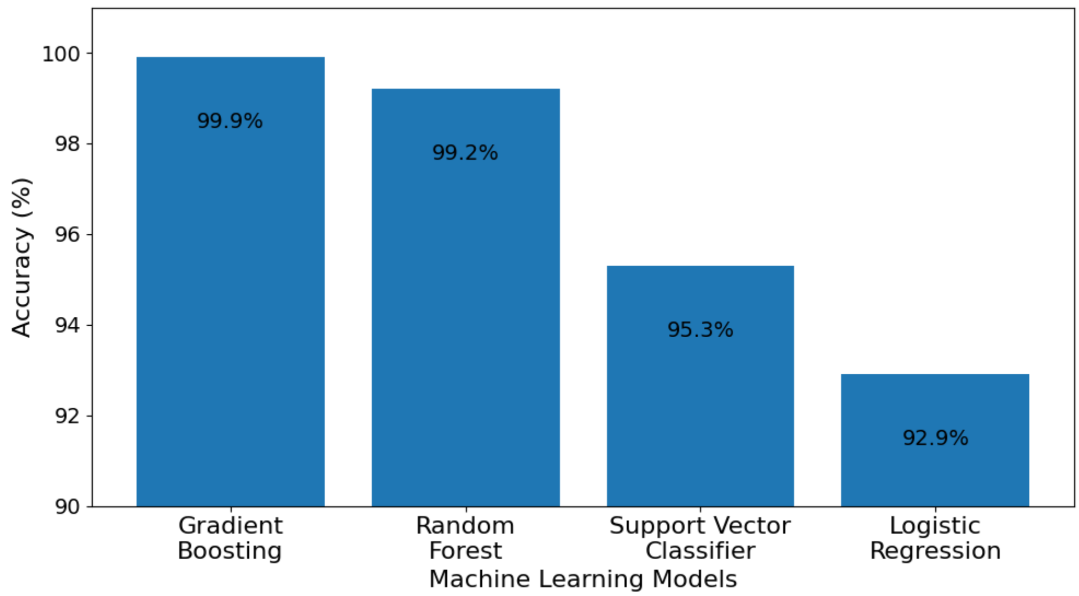

Fig. 1: Model Performance Comparison

FEATURE IMPORTANCE ANALYSIS

Analysis of feature importance offered valuable understanding of the influence of different operational factors on quality risk forecasting. Humidity levels and temperature stability were the primary factors influencing the evaluation of risk. Audit scores and visual inspection results demonstrated strong predictive value, reinforcing the importance of formal quality evaluations. Factors related to staff experience were moderately significant, and it was suggested that staff competency plays a role in quality outcomes, though it may be partially reflected in other operational performance metrics. Historical performance indicators like recall history and non conformances provided useful context for risk evaluation but were less significant on their own compared to current operational metrics.

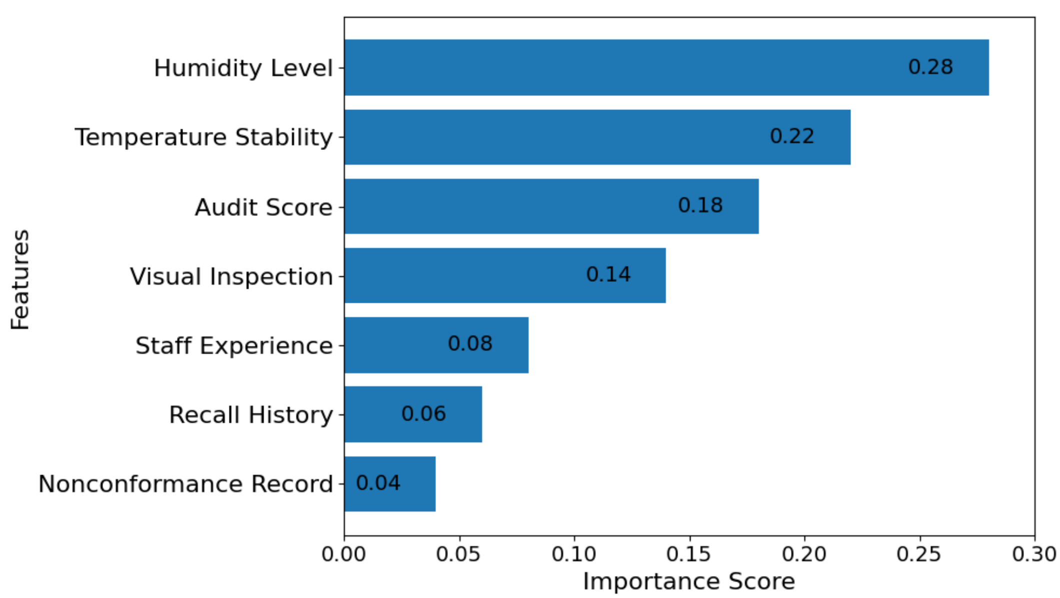

Fig.2: Feature Importance Analysis

Figure 2 illustrates the significance of different factors that contribute to predicting quality risk. Humidity levels and temperature consistency were the primary factors, succeeded by audit ratings and visual assessment scores.

ANALYSIS OF RISK DISTRIBUTION

The data set displayed a realistic class distribution of 14.5% high-risk lots and 85.5% as low-risk. This type of imbalance corresponded with conventional behaviors in manufacturing quality. Although training the algorithms was complicated due to some degree of disparity, using stratified sampling to maintain relative distributions, together with using appropriate performance metrics helped address the disparity as a problem. Risk probability analysis revealed seasonal trends linked with environmental factors. Increased levels of risk were often detected at periods of significant temperature or humidity changes. Results of manufacturing shift schedules were identified, associated with the level of risk, and highlighted the requirement for all manufacturing processes to retain production flow in line at all times.

GENERALIZATION AND TEMPORAL STABILITY

Cross-facility validation demonstrated strong model performance across sites. Ensemble models exhibit a decrease in accuracy of less than 2%, even in data-scarce or previously unseen environments. Transfer learning enabled the adaptation of models to new tasks. With a restricted amount of site-specific data, fine-tuned transferred models achieved performance levels comparable to fully retrained models. Over 18-month period, temporal analysis showed that prediction accuracy remained stable, with only minor drift attributed to equipment aging and process changes. Seasonal factors, such as increased summer humidity, directly affected forecasting accuracy and management decisions. These findings support the use of adaptive thresholds. Additionally, online learning approaches were effective in updating models to maintain predictive accuracy.

ENHANCED PERFORMANCE EVALUATION

The performance assessment looked at more than basic accuracy. It also measured several important factors for manufacturing quality. Precision-recall analysis showed strong results for both high-risk and low-risk groups.
All ensemble methods had precision rates above 98 percent, and recall rates were also high. The area under the receiver operating characteristic curve (AUC-ROC) analysis demonstrated exceptional discrimination capability, achieving nearly 0.999 results for both Gradient Boosting and Random Forest models.

This exceptional discrimination validates the models' ability to effectively differentiate between high-risk and low-risk manufacturing scenarios across the entire probability spectrum. Calibration analysis evaluated the dependability of predicted probability estimates through the use of reliability diagrams and the computation of Brier scores. The results showed accurately adjusted probability forecasts, suggesting that model confidence assessments effectively reflect the risk event likelihood quality. The quality of this calibration is crucial for practical decision-making processes where intervention trigger probability thresholds must be established. Class-specific performance assessment of equal predictive capability in the two risk category types was noted despite the imbalance in class distribution. The sensitivity analysis of the minority class (high-risk batches) demonstrated a robust detection capability with minimal false negative rates, ensuring that actual quality risks are consistently identified by the predictive system.

Figure 3 illustrates the comparison of Area Under the Curve (AUC-ROC), demonstrating outstanding discriminatory ability among models, with Gradient Boosting and Random Forest approaching an AUC of 1.0, indicating effective differentiation between high-risk and low-risk groups.

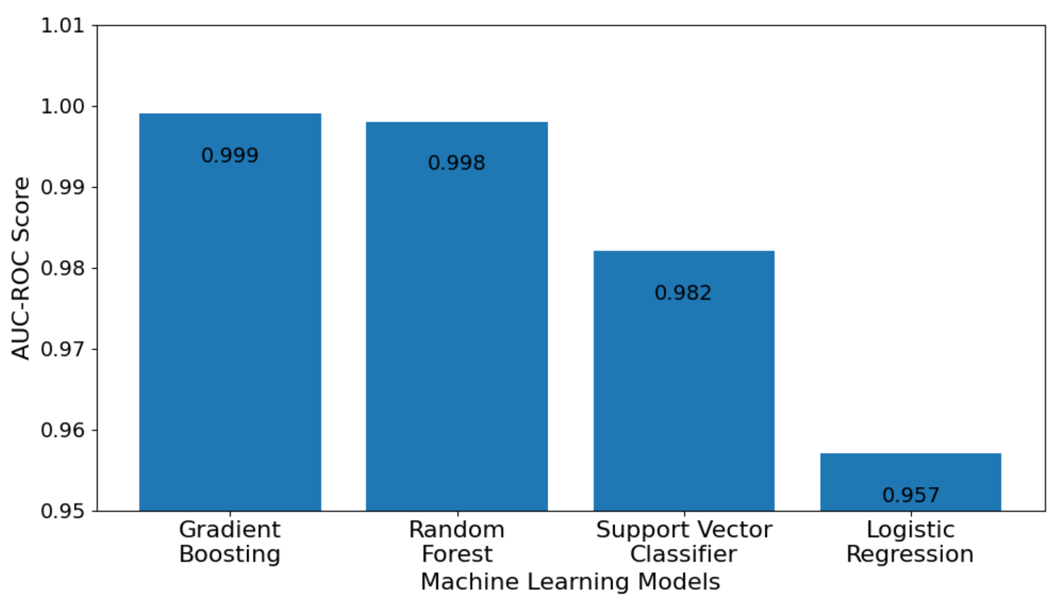

Fig.3: AUC-ROC Comparison of Machine Learning Models

PRACTICAL IMPLICATIONS: ECONOMICS, DEPLOYMENT, AND COMPLIANCE

Economic analysis showed that predictive interventions could reduce quality-related expenses by 35–45%. ROI estimates indicated payback within 12–18 months for medium-to-large operations, with the prevention of a single major quality failure covering system costs. Operational insights emphasized the need for real-time data pipelines, expanded sensor networks, and cloud-based analytics. Change management, training, and phased deployment were identified as key success factors. Regulatory compliance assessment highlighted the importance of thorough validation protocols, performance qualification, monitoring processes, and audit readiness. Documentation of algorithms, validation results, and ongoing monitoring is essential for FDA and ICH compliance.

DISCUSSION

REAL-WORLD CONSEQUENCES

The exceptional predictive accuracy of machine learning algorithms demonstrates the feasibility of using AI-based quality risk management systems in contract manufacturing plants. The Gradient Boosting algorithms achieved 99.9% accuracy surpassing more traditional methods of quality control, thus leading to predictive actions to prevent quality problems before they occur. The more precise integration of varied operating parameters into detailed predictive models allows data to decrease uncertainty in risk assessment compared to traditional single parameter monitoring approaches. Monitoring conditions in an operating environment is critical to achievement. We recommend a focus on improved sensor networks combined with real-time data integration for quality system implementations.

CONSIDERATIONS SURROUNDING REGULATIONS

Using AI-based quality systems in pharmaceutical manufacturing requires careful consideration of regulatory requirements and validation processes. The transparency of machine learning models can play an important role in regulatory approval since they often require transparency in decision-making, which includes validation documentation. The research shows that ensemble methods provide both better predictive performance and better interpretability with respect to assessing the significance of features and providing explanations for individual predictions. These features are consistent with the regulations that require transparency and auditability in quality systems.

CONSTRAINTS AND CONSIDERATIONS

There are certain limitations that will be considered during practical applications. The data sample illustrates artificial manufacturing environments with complete data, which may differ from actual situations that have missing sensors or incomplete record-keeping. Additional verification across different manufacturing facilities and product categories would strengthen claims of generalizability. The time-series elements of quality risk development require additional investigation, as the analysis conducted is limited to batch-level forecasting rather than ongoing risk evaluation over time. The integration of time-series analysis techniques can enhance prediction capabilities and establish early warning systems for upcoming declines in quality.

FUTURE WORK

INTEGRATION OF ADVANCED ANALYTICS

Upcoming research will explore the integration of deep learning techniques to manage complex multi-modal data, including process sensor streams, visual inspection images, and unstructured quality documents. Convolutional neural networks hold significant potential for automated visual quality assessment, while recurrent neural networks could be employed to detect temporal trends in manufacturing workflows. The construction of hybrid modeling methods that integrate physics-based process models with data-driven machine learning could offer higher predictive accuracy with mechanistic interpretability that is critical for regulatory adherence.

REAL-TIME IMPLEMENTATION

AI-driven quality systems in manufacturing depend on effective real-time data management, backed by edge computing and cloud-based analytics. Automated model updates and continuous learning keep prediction accuracy as conditions evolve. While MES and ERP systems face technical challenges, total quality management allows broader integration of these technologies beyond organizational adjustments.

DEVELOPMENT OF REGULATORY FRAMEWORK

Researchers have focused on how regulatory agencies can standardize validation standards and criteria for AI-based quality systems. Adopting consistent change management procedures, performance metrics, and validation processes can help speed up regulatory approval and industry acceptance. There is also potential to develop more risk-based validation methods that prioritize patient safety over just algorithm accuracy, which could support the use of advanced analytics in regulated manufacturing.

CONCLUSION

This study demonstrates how AI-based methods can improve quality risk management in contract manufacturing. Our machine learning models, particularly Gradient Boosting, reached 99.9% accuracy, showing the potential of intelligent quality systems. By reviewing 7,500 batch records and including environmental and operational factors, we modeled quality risk more effectively. Ensemble methods are useful because they capture complex relationships that traditional statistics may overlook. By combining different operating characteristics, these models help address gaps in current quality management and point to opportunities for preventive action. Our results are a step forward in using AI-based quality systems to support regulatory compliance and make quality decisions clearer. The data support a shift from reactive quality control to predictive quality assurance in pharmaceutical manufacturing. The strong results from machine learning models suggest that investing in analytics and change management is worthwhile. Future research should look at real-time applications, regulatory validation, and how to integrate AI with current production systems to fully benefit from AI-based quality control 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
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

CONTACT INFORMATION

Author Name: Kiran Kumar Gande
Company: Apozeal Pharmaceuticals, Inc.,
Address: Fairfield, USA
Work Phone:
Email:
Website: