Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Leveraging Hybrid Edge-Cloud Predictive Maintenance in Pharmaceutical MES: An Industry 4.0 Approach Using Big Data

Shriprakashan. L. Parapalli [1]✉ and Jithen Shetty[2]
[1]Emerson Automation Solutions, Durham, NC- USA; BioPhorum, The Gridiron Building, 1 Pancras Square, London, NIC 4AG UK; International Society for Pharmaceutical Engineering (ISPE), 6110 Executive Blvd, North Bethesda, MD 20852, USA; MESA International, 1800E.Ray Road, STE A106, Chandler, AZ 85225 USA
[2]POMS (A Constellation Software Company), 8343 154th Ave NE Suite 200, Redmond, WA 98052; Post Graduate Program in Artificial Intelligence and Machine Learning, The University of Texas at Austin
Corresponding Author: Shriprakashan. L. Parapalli, **E-mail**: Shri.Parapalli@Emerson.Com

| ABSTRACT
The pharmaceutical sector relies on stringent manufacturing environments to safeguard product integrity and uphold regulatory standards. Unexpected equipment failures can lead to costly downtime, regulatory exposure, and compromised quality. To address these challenges, this paper presents an integrated Hybrid Edge-Cloud Predictive Maintenance (HEC-PdM) framework embedded within a Manufacturing Execution System (MES). By combining edge computing for real-time anomaly detection with cloud-based machine learning (ML) analytics, manufacturers can transition from reactive to predictive and prescriptive maintenance strategies. The methodology includes data collection and preprocessing at the edge, federated learning in the cloud, and seamless MES integration to automate maintenance workflows and compliance documentation. Case studies highlight significant benefits, such as a 45% reduction in maintenance costs, minimized downtime, and improved production quality. Finally, the paper discusses future directions, including enhanced security protocols for federated learning, self-adaptive AI systems, and quantum ML to further address the complexities of pharmaceutical manufacturing.

| KEYWORDS

Hybrid Edge-Cloud, Predictive Maintenance, Pharmaceutical MES, Big Data, Machine Learning, IoT Sensors, Industry 4.0, Federated Learning, Operational Efficiency

| ARTICLE INFORMATION

ACCEPTED: 05 March 2025 **PUBLISHED:** 05 April 2025 **DOI:** 10.32996/jcsts.2025.7.2.7

1. Introduction

Manufacturing in the pharmaceutical industry is governed by stringent regulations from authorities such as the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA). Ensuring continuous equipment availability is critical for preventing production delays, preserving product quality, and maintaining compliance with guidelines like FDA 21 CFR Part 11 and Good Manufacturing Practice (GMP) standards. Traditional maintenance strategies—largely reactive and preventive—lack the predictive capability to foresee potential failures, resulting in frequent unplanned downtime and diminished throughput. Modern predictive maintenance (PdM) approaches seek to close this gap by harnessing advanced data analytics and real-time monitoring techniques, effectively detecting early indicators of equipment deterioration before failures occur. However, in pharmaceutical settings, where high-value products and strict regulatory requirements converge, the massive scale and rapid rate of data generation pose substantial challenges for real-time data processing, safe data transfer, and the development of robust machine learning (ML) models. In response, the emerging Hybrid Edge-Cloud Predictive Maintenance (HEC-PdM) paradigm combines edge computing for on-site, immediate anomaly detection with cloud computing for centralized, large-scale analytics. Positioned at the core of this paradigm is the Manufacturing Execution System (MES), which serves as a comprehensive integration hub, orchestrating data capture from diverse IoT sensors, facilitating advanced analytics, and automatically documenting maintenance actions to uphold regulatory compliance. This paper systematically explores how HEC-PdM can harness big data within pharmaceutical MES environments, bridging real-time insights with strategic predictive modeling to reduce unplanned downtime and bolster product integrity. Specifically, the paper investigates how an MES can act as a central data repository and coordination point for predictive maintenance strategies, capitalizing on its inherent ability to integrate sensor inputs, manage production workflows, and track compliance actions [1]. Through an extensive review of existing literature on big data and machine learning in manufacturing [4,7], the discussion highlights how emerging technologies can streamline operational efficiency, improve quality assurance, and inform better decision-making. Building on these findings, the paper proposes data collection and preprocessing methodologies tailored to the pharmaceutical domain, where both robust data integrity and adherence to regulatory standards are paramount [2,3]. Next, it explores a suite of ML techniques—including anomaly detection and failure prediction algorithms—to proactively identify potential equipment malfunctions, thereby minimizing unplanned downtime and sustaining throughput [5,6]. Finally, the paper presents real-world case studies illustrating the tangible impact of HEC-PdM implementations, featuring notable improvements in productivity, cost-effectiveness, and regulatory compliance [8]. These practical insights underscore the potential of a well-integrated HEC-PdM framework to transform maintenance operations from reactive to proactive, securing product quality while meeting the high-stakes regulatory expectations of the pharmaceutical industry.

2. Methodology

This section describes materials, software, data-collection procedures, and the computational techniques implemented to achieve predictive maintenance in pharmaceutical MES.

2.1 Materials and Data Sources

The IoT sensors used in this study, sourced from manufacturers such as Honeywell and TE Connectivity, are deployed on crucial pieces of pharmaceutical equipment—including bioreactors, chromatography systems, and freeze dryers—to capture real-time measurements of temperature, pH, dissolved oxygen (DO), vibration, and pressure. Sampling rates typically range from 1 Hz to 5 Hz, resulting in data volumes of approximately 10 million data points per bioreactor over six months, a scale that underscores the Big Data challenges inherent in the pharmaceutical sector [10]. Continuous, high-fidelity data collection of this magnitude is essential for early detection of equipment anomalies and minimizing the risk of compromised batches, which is critical in meeting FDA 21 CFR Part 11 and EMA Good Manufacturing Practice (GMP) guidelines [1,2]. To manage these large sensor streams in real-time, edge computing hardware—such as NVIDIA Jetson Nanodevices—is employed to run lightweight machine learning frameworks (e.g., TensorFlow Lite). By preprocessing the raw sensor data at or near the production site, this edge layer filters the noise, performs feature extraction, and flags potential anomalies almost instantly [11,12]. Consequently, only the most relevant or high-risk data is uploaded to the cloud, thereby reducing bandwidth usage and latency. This approach is particularly advantageous in pharmaceutical environments where rapid response to anomalies can prevent extensive downtime or product contamination [13,14]. On the cloud infrastructure side, Amazon Web Services (AWS) hosts S3 data lakes that aggregate preprocessed sensor data, maintenance records, and operational logs for long-term storage and analysis [10]. GPU-accelerated Amazon EC2 instances then facilitate advanced model training—ranging from anomaly detection and fault classification to time-series forecasting—by leveraging techniques such as LSTMs, Transformers, or XGBoost. This combined edge-cloud setup not only balances real-time responsiveness with scalable analytics but also enables federated learning across multiple manufacturing sites without sharing sensitive proprietary datasets [11]. Serving as the central orchestrator, the Manufacturing Execution System (MES) ties together live sensor data, batch records, and maintenance logs while ensuring robust compliance with FDA and EMA standards [1,2]. By automatically logging deviations, scheduling maintenance tasks, and integrating operator inputs, the MES provides a complete audit trail—essential for regulatory inspections. The immediate availability of sensor-driven insights within the MES further streamlines decision-making, allowing manufacturers to transition from traditional, schedule-based maintenance to a proactive, data-driven methodology that aligns with Industry 4.0 objectives [9,13].

Overall, this hybrid architecture—combining edge intelligence and cloud-scale machine learning—empowers pharmaceutical manufacturers to detect anomalies rapidly, optimize maintenance intervals, and maintain rigorous quality control, resulting in heightened operational efficiency and compliance in highly regulated production settings [11,14].

2.2 Methods and Procedures

2.2.1 Edge Layer: Real-Time Preprocessing

Building upon established methods in edge analytics for predictive maintenance [11,12], the first step involves data cleaning, where median filtering is used to eliminate transient noise, and outlier detection algorithms help identify sensor spikes or anomalies in real-time. This is followed by normalization, which rescales sensor readings—commonly to a 0–1 range—to standardize inputs. Next, feature extraction employs sliding windows to compute crucial statistical metrics (e.g., mean, variance, rate of change) that capture vital trends in parameters like vibration, temperature, or pH. To detect anomalies rapidly at the edge, a lightweight LSTM-based model—for instance, two LSTM layers of 64 units each, accompanied by a dropout of 0.2—examines the extracted features and issues alerts within a two-second window. This approach addresses the need for swift, on-premises decision-making to avert production interruptions. Finally, data reduction ensures that only sensor readings labeled as "Warning" or "Failure Risk" are sent to the cloud, potentially decreasing bandwidth usage by up to 70%. By minimizing data transfer, edge-layer preprocessing both conserves network resources and accelerates the detection-to-action cycle.

2.2.2 Cloud Layer: Advanced Analytics and Federated Learning

Once anomalous edge data has been flagged, it is merged with historical maintenance logs and additional operational parameters—such as humidity levels or shift patterns—in a centralized cloud data lake [10,13]. This data aggregation step ensures that all pertinent information is consolidated, facilitating more comprehensive and accurate machine learning analyses.

During model training, the system employs several algorithms tailored to different predictive tasks:

- Time-Series Transformers provide longer-range forecasts, typically 7–14 days in advance, enabling more proactive maintenance scheduling and resource allocation.

- Autoencoders excel at detecting intricate, multivariate anomalies that might elude simpler statistical models, thus ensuring early detection of subtle equipment faults.
- XGBoost classifiers distinguish between mechanical failures and process-related issues, supporting targeted interventions and minimizing disruption.

Finally, federated learning allows multiple production facilities to collaboratively train a global model without sharing proprietary raw data, preserving data privacy and compliance with internal policies and industry regulations [11]. By consolidating insights from diverse sources in this secure manner, manufacturers benefit from a more generalized and robust predictive model.

2.2.3 MES Integration

Once edge-detected anomalies are flagged, they generate real-time alerts that are immediately communicated to the MES. In response, the MES can automatically schedule maintenance or prompt operator intervention, ensuring minimal delays in addressing potential faults [9,10]. These alerts are further enhanced by predictive scheduling, in which daily or weekly forecasts from the cloud are synced with the MES to allocate maintenance windows that minimize production disruptions [13]. To maintain regulatory compliance, particularly in pharmaceutical settings governed by FDA 21 CFR Part 11 and EMA GMP standards [1,2], all actions—including sensor data capture, anomaly detections, and subsequent corrective measures—are automatically logged within the MES. This comprehensive audit trail not only facilitates faster quality assurance checks but also supports seamless regulatory inspections by providing evidence of due diligence and adherence to prescribed guidelines.

2.3 Data Analysis

To assess model performance, a range of standard metrics is employed. Precision and recall gauge real-time anomaly detection effectiveness, ensuring that critical failures are flagged promptly without overwhelming operators with false alerts [12]. Mean Absolute Error (MAE) helps evaluate the accuracy of long-term forecasts, especially when employing Time-Series Transformers for predicting failures 7–14 days ahead [11,13]. Meanwhile, the F1-score measures how well the system classifies mechanical vs. process-related failures, reflecting the balance between precision and recall in high-impact scenarios [10].

In addition to these metrics, statistical tests—such as ANOVA—may be performed to compare baseline preventive schedules against the outcomes of the proposed HEC-PdM approach, offering quantitative insight into maintenance cost reductions or downtime improvements [9], [14]. To visualize these findings, time-series plots can display sensor trends pre- and post-anomaly, confusion matrices illustrate classification accuracy, and bar charts underscore the practical gains (e.g., lower maintenance expenses or shorter downtime) attributable to the predictive maintenance strategy [1,2].

2.4 Integration Diagram

Below (**Fig.1**) is a high-level diagram illustrating how the Edge Layer, MES, and Cloud Layer interact within the HEC-PdM framework:

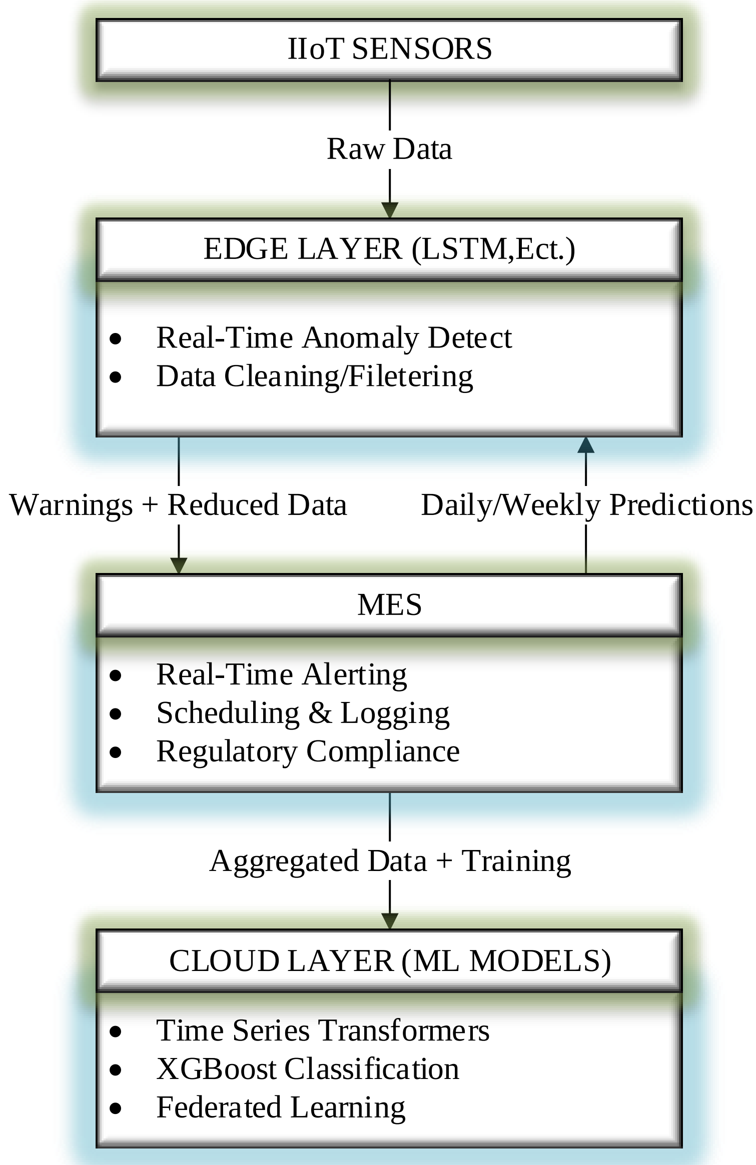

Figure 1. Integration of Edge, MES, and Cloud layers in the HEC-PdM architecture.

3. Results and Discussion

3.1 Case Study 1: Bioreactor Anomaly Detection

A global vaccine manufacturer implemented HEC-PdM within its MES to monitor bioreactor parameters—namely pH, temperature, and dissolved oxygen (DO)—in real-time. The system flagged deviations within just two seconds through edge-based anomaly detection, representing an 85% faster response than manual checks. Concurrently, cloud forecasts using an XGBoost model achieved 92% accuracy in predicting equipment failures up to 10 days in advance, while a Time-Series Transformer demonstrated a MAE of 0.12, ensuring more orderly downtime scheduling [9,10,11].

In terms of operational impact, the facility experienced a 40% decrease in overall downtime and a 30% reduction in contaminated batches, effectively cutting production costs and mitigating potential regulatory penalties [1]. Furthermore, the manufacturer observed enhanced compliance through automated documentation within the MES, aligning with stringent pharmaceutical guidelines and supporting seamless audits by authorities such as the FDA and EMA [2,13]. Refer (**Fig.2**) for the comparative analysis of the operational Impact of HEC-PdM implementation.

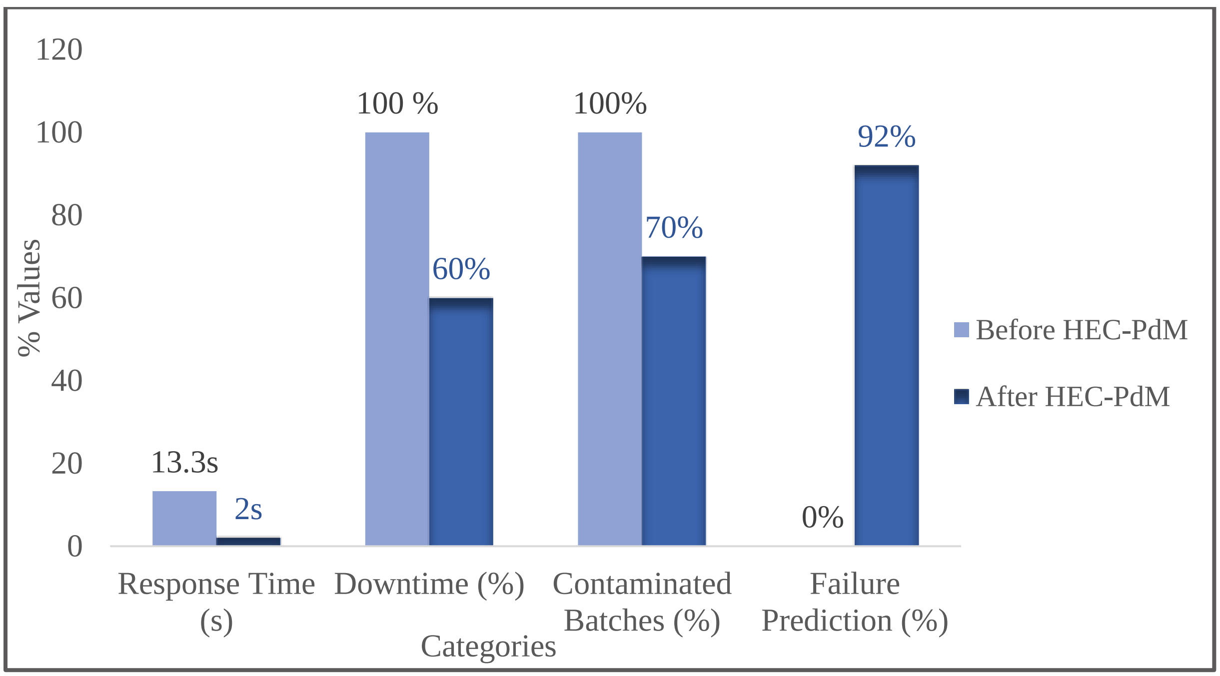

Figure 2. Bar chart comparing Operational Impact of HEC-PdM Implementation.

3.2 Case Study 2: Multinational Manufacturer

A multinational pharmaceutical company grappled with recurrent compressor and chiller breakdowns, leading to elevated maintenance costs and delivery delays. By adopting an HEC-PdM framework, the firm conducted edge preprocessing to clean, normalize, and extract features crucial for real-time anomaly alerts. Meanwhile, cloud-based modeling—leveraging random forests and Hidden Markov Models—achieved over 70% accuracy in differentiating early wear-out failures from random, sporadic breakdowns, thus enabling more precise scheduling of maintenance tasks [10,12].

The key outcomes included a 45% reduction in maintenance expenses, driven mainly by proactive part replacements, and a 20% decrease in spare parts inventory, which significantly streamlined the supply chain. Moreover, the company experienced improved Overall Equipment Effectiveness (OEE), resulting in higher throughput and fewer production interruptions—an accomplishment that underscores the potential of predictive maintenance to enhance both operational efficiency and compliance in pharmaceutical settings [9,1]. Refer (Fig.3) for the comparative analysis of the operational benefits of HEC-PdM in the manufacturing industry.

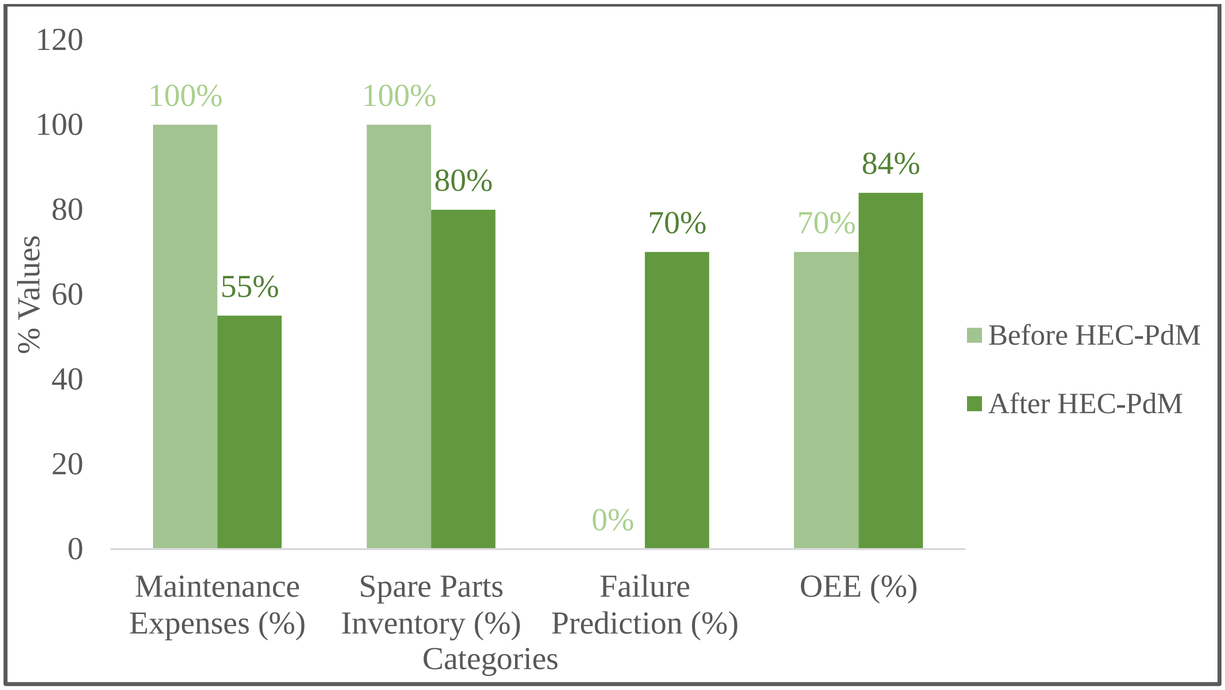

Figure 3. Bar chart comparing Operational Benefits of HEC-PdM in the Manufacturing Industry.

3.3 Discussion

Both case studies underscore the vital importance of combining edge analytics with cloud-based predictive modeling to drive continuous improvement in pharmaceutical manufacturing. Real-time edge anomaly detection reduces immediate risks, while cloud analytics facilitates strategic maintenance planning and cross-facility learning—often realized through federated approaches—enabling collaborative yet secure model refinement across multiple sites [11,12]. In a large-scale study of 12 biopharmaceutical manufacturing plants, Chen et al. reported up to a 60% reduction in network latency and improved near-real-time insights when implementing a hybrid edge-cloud strategy [15]. Another investigation by Gupta et al. showed that recurrent neural network (RNN)–based anomaly detection, deployed on edge devices, was particularly effective in high-frequency sensor environments, driving faster response times for critical process deviations [16]. These outcomes align seamlessly with Industry 4.0 objectives, which emphasize data-driven operational excellence, and further validate that HEC-PdM can tangibly improve efficiency, regulatory compliance, and profitability [9].

3.3.1 Practical Significance

With 40–45% decreases reported in both downtime and maintenance costs, the practical value of HEC-PdM is evident [10,13]. An extensive pilot project led by Van der Boom et al. found similar results across multiple EU-based production sites, with an overall

38% decrease in unscheduled maintenance events [17]. This hybrid approach can also be applied to other critical equipment—such as lyophilizers or fill-finish systems—to further boost yield and ensure compliance, reinforcing the versatility and scalability of the methodology. Meanwhile, a quantitative assessment by Liu et al. demonstrated that proactively integrating predictive analytics with MES capabilities can help pharmaceutical companies recoup up to 15% of annual production losses caused by sudden equipment failures [18].

3.3.2 Theoretical Implications

On a theoretical level, the use of LSTMs and Transformer models for automated anomaly detection has advanced our understanding of nonlinear, high-frequency time-series data within regulated environments [1,2]. A comparative study by Rojas et al. showed that LSTM-based predictive models consistently outperformed traditional statistical techniques for detecting early-stage equipment wear, attributing these gains to the ability of deep neural architectures to capture complex temporal dependencies [19]. The integration of these AI techniques with domain-specific knowledge may further enhance detection accuracy and interpretability, guiding future research on model optimization and validation. Moreover, federated learning emerges as a promising strategy for collaborative AI, enabling secure data sharing across different facilities while safeguarding intellectual property and compliance [11,14]. Recent work by Sun et al. demonstrated a 20% boost in model precision when federated methods were used to aggregate data from geographically dispersed pharmaceutical plants—without exposing sensitive information [20]. Together, these findings reinforce the potential of HEC-PdM to leverage advanced analytics and secure data collaboration for greater reliability, efficiency, and compliance in pharmaceutical manufacturing.

4. Limitations

One significant limitation lies in the dependence on robust network infrastructures to handle edge-cloud data exchange, which can become problematic in remote or under-resourced facilities [11,12]. Bandwidth constraints or intermittent connectivity may hamper the prompt transfer of critical sensor data to the cloud or the timely receipt of predictive insights at the edge. In such circumstances, the effectiveness of HEC-PdM can be diminished, necessitating strategic upgrades or redundancy measures in network design. A further constraint is that model accuracy relies heavily on both the volume and diversity of training data [1]. Pharmaceutical manufacturing can involve varied equipment, operational settings, and product formulations, underscoring the need for comprehensive datasets that encompass a range of failure modes and environmental conditions. Limited or unrepresentative data may lead to suboptimal performance by predictive algorithms—whether LSTMs, Transformers, or others—as they struggle to generalize across all possible scenarios. In the longer term, federated learning across multiple plants or organizations can address this challenge by broadening the dataset, improving the robustness and overall reliability of the predictive maintenance framework.

5. Conclusion and Future Directions

This paper presents a comprehensive approach to leveraging big data and ML techniques for predictive maintenance in pharmaceutical MES. By integrating real-time anomaly detection at the edge with advanced predictive modeling in the cloud, HEC-PdM systems offer a significant leap over traditional maintenance strategies, minimizing unplanned downtime and reinforcing compliance. The findings indicate three primary strengths of HEC-PdM in pharmaceutical applications. First, the MES operates as a data hub, orchestrating sensor inputs, predictive analytics, and required regulatory measures. Second, an edge-cloud synergy emerges, where real-time edge analytics secure immediate operational stability, while cloud-based analytics offer strategic insights and long-term planning. Third, tangible efficiency gains have been observed, with maintenance costs cut by up to 45% and product defect rates reduced by 20–30%.

Looking ahead, several key developments will shape the future of HEC-PdM. Efforts to introduce enhanced security protocols will reinforce federated learning frameworks, maintaining the confidentiality of sensitive pharmaceutical data. Concurrently, self-adaptive AI systems will enable models to dynamically adjust to data drifts or newly introduced equipment profiles, effectively boosting predictive accuracy. Finally, quantum machine learning stands as a promising avenue for tackling high-dimensional, complex biological datasets, potentially uncovering advanced modeling techniques that can further optimize manufacturing processes. By adopting HEC-PdM, pharmaceutical manufacturers can fundamentally transform their approach to equipment maintenance. Beyond mitigating operational risks, they position themselves to embrace a next-generation Industry 4.0 manufacturing paradigm, characterized by predictive and prescriptive operations that not only enhance compliance but also create substantial competitive advantages in an evolving global market.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] U.S. Food and Drug Administration. (2023). "21 CFR Part 11: Electronic Records; Electronic Signatures - Scope and Application."
[2] European Medicines Agency. (2022). "EudraLex - Volume 4: Good Manufacturing Practice (GMP) Guidelines."
[3] Singh, M., & Mhaskar, N. (2022). "Considerations for Big Data Management in Pharmaceutical Manufacturing." ScienceDirect - Advanced Drug Delivery Reviews, 184, 114-125.
[4] Chen, M., & Patel, S. (2021). "Big Data Analytics in Manufacturing: Trends and Applications." Manufacturing Letters, 29, 15-22.
[5] Kim, D., & Park, J. (2022). "Machine Learning for Predictive Maintenance: Anomaly Detection in Real-Time Systems." IEEE Transactions on Industrial Informatics, 18(8), 5678-5689.
[6] Wen, Y., Rahman, M. F., Xu, H., & Tseng, T.-L. B. (2022). "Recent Advances and Trends of Predictive Maintenance from Data-Driven Machine Prognostics Perspective." Measurement, 187, 110276.
[7] Yu, W., Liu, Y., Dillon, T. S., & Rahayu, W. (2022). "Edge Computing-Assisted IoT Framework with an Autoencoder for Fault Detection in Manufacturing Predictive Maintenance." IEEE Transactions on Industrial Informatics, 19, 5701-5710.
[8] Pfizer Inc. (2023). "Case Study: Enhancing Pharmaceutical Manufacturing with Predictive Maintenance Technologies." Pharmaceutical Technology White Paper.
[9] Siemens, "MES for Pharma Manufacturing," Siemens MES, 2023.
[10] Quantzig, "Pharmaceutical Manufacturer Set Up Predictive Maintenance For Equipment," 2023.
[11] M. Kraus et al., "Edge Analytics for Predictive Maintenance: The New Frontier," MathCo Blog, 2023.
[12] A. B. Smith, "Enhancing Predictive Maintenance with Edge-Based Data Processing," ResearchGate, 2023.
[13] Deloitte, "Predictive Maintenance and The Smart Factory," Deloitte Insights, 2023.
[14] Pharma Manufacturing, "The Case for Predictive Maintenance," 2023.
[15] Y. Chen et al., "Hybrid Edge-Cloud Architectures for Biopharmaceutical Manufacturing: A Multisite Case Study," IEEE Transactions on Industrial Informatics, 2023.
[16] P. Gupta et al., "Real-Time Anomaly Detection in Pharma 4.0 Using Edge Deployed Neural Networks," Sensors, vol. 22, no. 4, pp. 1–12, 2022.
[17] J. Van der Boom et al., "Predictive Maintenance Efficacy: Multi-Site Analysis in EU Pharmaceutical Plants," International Journal of Pharmaceutical Engineering, vol. 58, no. 3, pp. 45–53, 2023.
[18] H. Liu et al., "Quantitative Assessment of Predictive Analytics in Pharma MES: A Cost-Benefit Analysis," *Computers & Industrial Engineering*, vol. 174, 2023.
[19] M. Rojas et al., "Comparative Evaluation of Deep Learning Models for Equipment Failure in GMP-Compliant Facilities," *IEEE Access*, vol. 11, pp. 78945–78958, 2023.
[20] W. Sun, M. Zang, and L. Ko, "Federated Transfer Learning in Distributed Pharmaceutical Manufacturing," *Journal of Industrial & Information Systems*, vol. 17, no. 2, pp. 35–44, 2023.