

How Multimodal Data Can Enable Scalable and Trustworthy AI in Factories of The Future

Joyjit Chatterjee, Rajendra Shankaran, Nina Dethlefs

▶ **To cite this version:**

Joyjit Chatterjee, Rajendra Shankaran, Nina Dethlefs. How Multimodal Data Can Enable Scalable and Trustworthy AI in Factories of The Future. 2024. ⟨hal-04696344⟩

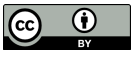
HAL Id: hal-04696344

https://hal.science/hal-04696344v1

Preprint submitted on 13 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

How Multimodal Data Can Enable Scalable and Trustworthy AI in Factories of The Future

Joyjit Chatterjee*[a], Rajendra Shankaran[b] and Nina Dethlefs[c]

[a]Reckitt Health Care UK Ltd, Dansom Lane, Hull, HU8 7DS, United Kingdom
[b]Reckitt Global Technology Hub, The Skyview 20, Raidurgam, Madhapur, Hyderabad, 500081, India
[c]University of Hull, School of Computer Science, Cottingham Road, Hull, HU6 7RX, United Kingdom

ABSTRACT

Manufacturers have increasingly turned to Artificial Intelligence (AI) to address specific problems in factories e.g. predictive maintenance, improving product quality etc. Implementing these in silos can miss critical interdependencies during root cause analysis and the interplay between various data sources across broader manufacturing operations. We provide a perspective on a holistic data model that harmonises data generated during production processes and across the supply chain enabling data-driven decisions to drive productivity, optimisation, and sustainability. Our approach is based on Overall Equipment Effectiveness (OEE) that focuses on predictive maintenance to increase availability, Continued Process Verification (CPV) and predictive stability to improve quality, and real-time insights, thereby optimising performance. The proposed multimodal model can help enhance productivity and reduce waste and rework, lending itself to sustainability imperatives. Similarly, Cost to Serve analysis, which targets inefficiencies in the distribution network and transportation, helps facilitate a reduction in costs and minimisation of carbon footprint in the supply chain.

1. Introduction

Manufacturing and production sectors contribute to one-fifth of global carbon emissions [1]. The era of Industry 4.0 has seen growing interest across industry sectors to implement sustainable manufacturing practices that can help minimise the negative environmental impact of production processes. Industry 4.0 technologies like Industrial Internet of Things (IIoT), Machine Learning (ML), and Augmented/Virtual Reality (AR/VR) etc. elevate traditional approaches based on lean, Total Productive Maintenance (TPM), Total Quality Management (TQM) and six sigma [2] to transition towards a more efficient and sustainable factory of the future.

Manufacturing traditionally uses statistical techniques like Statistical Process Control (SPC) to reduce process variability, but modern techniques like machine learning and deep learning models can solve specific problems on the shop floor e.g. predictive maintenance, optimising product quality and consistency by forecasting (and potentially correcting) variability issues before they occur by learning from patterns in historical data [3;4;5;6]. Traditional approaches have examined specific problems in isolation, neglecting the comprehensive scope of manufacturing, including R&D, sourcing, production, and distribution. Such in-isolation approaches increase the likelihood of false alarms and human bias, focusing on certain associations while overlooking novel ones. Consequently, ML methods based on data silos remain predominantly academic, with limited industrial application due to issues of scalability and trust.

Industry 3.0 (Automation) generated vast volumes of machine and production data, but in isolated systems. The advent of Industry 4.0 and digital transformation has enhanced our capacity to connect, collect, store, analyse, and visualise data. Increased interest in smart manufacturing systems for existing plants has resulted in more data from the shop floor, facilitated by new sensors, advanced processing software, affordable cloud storage, and minimal intervention in ongoing processes [7]. To optimise product quality and minimise waste, it is essential to consider the entire product development lifecycle and the collective impact and interactions of various data streams, such as environmental data, process parameters, and quality data.

Data-driven analytics in manufacturing not only boosts operational efficiency but also enhances innovation and responsiveness to market demands. It spans multiple stages, including Research and Development, production, quality control, and logistics. Analytics can be applied across these areas in the following ways:

- **Research and Development (R&D): Simulation and Modelling –** Predicting the performance of new products or materials through simulation, reducing the need for physical prototypes.
- **Production –**
  1. **Process Optimisation –** Analysing production processes to identify inefficiencies and optimise operations, reducing waste and lowering costs.
  2. **Predictive Maintenance –** Using sensor data to predict when equipment is likely to fail, allowing for maintenance to be scheduled before costly breakdowns occur and at the same time minimising unnecessary downtime.
- **Quality Control: Defect Analysis –** AI-powered vision systems to improve accuracy, eliminate human error and avoid product recalls.

joyjit.chatterjee@reckitt.com (J. Chatterjee*)
ORCID(s): 0000-0003-2672-3832 (J. Chatterjee*); 0000-0002-6917-5066 (N. Dethlefs)

- **Logistics: Cost-to-serve Analysis –** Determining the minimum order quantity (MOQ) to predict cost and profitability; Analysing transportation routes and modes to minimise costs and delivery time.

Achieving any of the above necessitates a comprehensive data model that integrates data from factories and the supply chain, facilitating data-driven decisions to enhance productivity, optimisation, and sustainability. Overall Equipment Effectiveness (OEE) is a Key Performance Indicator (KPI) used to measure productivity, efficiency, and manufacturing losses in modern factories [8].

$$OEE = \textbf{Availability } x \textbf{ Quality } x \textbf{ Performance}$$

Theoretically, an OEE of 100% indicates defect-free production (100% quality), at maximum speed (100% performance), with no unplanned downtime (100% availability) [9]. In practice, an OEE above 85% is considered the benchmark for world-class factories [10], reflecting minimal waste and maximum yield.

By optimising OEE as the KPI for productivity in AI-applied manufacturing data, we propose an end-to-end data model targeting predictive maintenance to enhance availability, Continued Process Verification (CPV) and predictive stability to improve quality, and real-time insights to boost performance. This perspective can aid in developing reliable and scalable AI solutions to optimise safety, quality, and efficacy, enabling real-time release and reducing waste and costs in factories of the future

2. The blessings of AI in data-driven factories

AI can contribute to a sustainable and environmentally-conscious business model in manufacturing. ML models applied to big data from factories offer numerous opportunities, such as optimising production through predictive maintenance. This approach forecasts Operations and Maintenance (O&M) requirements, preventing unexpected breakdowns and extending equipment's Remaining Useful Life (RUL). AI can also minimise discarded products due to quality issues by enhancing Quality Control/Quality Assurance (QC/QA), especially in the Consumer Packaged Goods (CPG) industry where shelf-life is limited. It can ensure high-quality products and help detect defects before reaching consumers, enabling efficient resource allocation and reducing costs and waste in raw materials, energy, water and packaging [11]. Below, we provide a comprehensive overview of predictive stability, AI-assisted CPV, predictive maintenance, and Cost to Serve analysis, which form the manufacturing data model discussed in this paper.

2.1. Predictive stability

To ensure manufactured products adhere to industry standards for safety, quality and efficacy throughout their shelf-life, stability studies are regularly conducted, particularly for CPG and pharmaceuticals [12;13]. These studies are crucial in product design to maintain physical, chemical and microbiological properties [14]. Traditionally, stability tests are performed in controlled environments (e.g. stability chambers), where multiple product batches are exposed to varying conditions (e.g. temperature, relative humidity, light) and numerous tests are conducted on active ingredients and excipients [15;16]. This complex, time-consuming and costly process can lead to regulatory compliance challenges, increased cycle times and delays in market introduction, if testing practices are flawed [17]. These issues are exacerbated in real-time and accelerated stability studies [18], which predict long-term product performance by extrapolating degradation rates under elevated stress conditions, leading to higher rates of failed studies and accumulated errors [19].

Although there is currently no alternative to laboratory stability studies for regulatory compliance, the industry has significant potential to use data-driven predictive algorithms to forecast future stability test results before they are conducted in the lab. If the final test result is predicted to be outside acceptable specification limits, potential issues in product formulations can be identified early, allowing adjustments to the composition and distribution of active ingredients and excipients. This enables formulators to recognise stability test failures in advance, indicating that the product will not meet its intended shelf-life. Such predictions help eliminate redundant tests that are likely to fail, making the traditional testing process more efficient without hindering real-world testing. Incorporating data-driven predictive insights into stability testing workflows can enhance efficiency.

Traditionally, laboratory records in this domain have been paper-based [20], however, in recent times data is being increasingly stored in the Laboratory Information Management System (LIMS) [21], improving data management and regulatory compliance. Despite the shift to electronic systems like LIMS and Enterprise Resource Planning (ERP), discrepancies in data formats persist due to the coexistence of paper and electronic records. While LIMS data serves the same purpose of recording historical stability testing results, inconsistencies in data formats remain, varying across different test types, analytes, and testing conditions.

To build a predictive stability model, relevant information is required from multiple resources (Bill of Materials (BOMs), LIMS data, product specification data, Transaction Data Systems (TDS) information etc.). Ongoing Stability Studies (OGS) data can be leveraged to continuously monitor products in the market and ensure their active ingredients remain reliable and within tolerance limits over time. On the face of limited/no historical data, we can leverage theoretical representations e.g. molecular fingerprints of compounds – Simplified Molecular Input Line Entry System (SMILES) [22] – to aid the learning process of ML models. Predictive stability can help save significant costs and time, accelerate speed to market, and eliminate redundant tests that are likely to finally fail in the laboratories.

2.2. Continued Process Verification (CPV)

During manufacturing, process data is continuously generated in factories. CPV identifies how variability in critical

metrics (e.g. Critical Process Parameters – CPP, Critical Quality Attributes – CQA, Critical Material Attributes – CMA) affects product quality[23]. CPV involves collecting and processing end-to-end production data to ensure critical attributes are controlled throughout manufacturing. It ensures processes remain within tolerance limits, guaranteeing optimal product quality during commercial production.

CPV is crucial for regulatory compliance, adhering to guidelines from agencies like the European Medicines Agency (EMA) and the Food and Drug Administration (FDA)[24], especially in safety-critical industries like pharmaceuticals and consumer health. It helps understand factory performance, improve OEE, reduce waste, and enable real-time release. Traditionally, CPV has used statistical techniques, such as control charts, to analyse parameter deviations, relying heavily on human expertise and historical experience. Training ML models to assist CPV supports offline periodic root cause analysis and provides real-time insights into factors affecting product quality. Implementing analytics to detect patterns gives plant operations teams actionable warnings before CPV limits indicate a problem, minimising costly production impacts. Any detected anomalies can be rectified promptly to maintain process control[25], avoiding waste from non-conforming products[26].

This process lends itself to real-time release testing (RTRT)[27], also avoiding the discard of complete batches at the end of the process[28]. Potential challenges in building AI models for CPV include a lack of structured historical big data in factories where digitisation is still at an embryonic stage, the need for ensuring continuous updates/fine-tuning of the trained ML models and limiting any negative consequences of data drift (e.g. incorporation of new process parameters, changes in the manufacturing process etc.)[23].

2.3. Predictive maintenance

In the era of Industry 4.0, predictive maintenance (or condition-based maintenance) is a vital data-driven approach for manufacturers. It continuously monitors equipment to predict machine failures, ensuring maintenance is only done when necessary, which reduces unnecessary O&M activities[29]. This maximises asset uptime, minimises costs, and boosts production[30]. Data sources like vibration signals, infrared thermography, particle & oil analysis, and visual inspections help detect and diagnose faults[31].

ML models analyse sensor data to forecast failures, using techniques such as anomaly detection and time-series analysis. By predicting issues before they cause downtime, predictive maintenance reduces costs and improves efficiency, continuously learning from historical and real-time data.

Key considerations for building predictive maintenance models include the need for high-quality historical data, challenges in validating models without actual downtime, and building trust over time as real-world inconsistencies are matched with model predictions. It is also integral to overcome the challenges in explainability faced by predictive models, especially deep learners which can provide highly accurate predictions by leveraging historical sensor data but are typically black-box.

2.4. Computer vision for Quality Control

ML-based vision systems use computer vision and deep learning algorithms to inspect product labels and packaging[32]. These systems efficiently and accurately validate product codes, expiration dates, batch numbers, and other essential details. They can also be leveraged for recognising optical defects in manufactured products e.g. cracks, tears, stains etc. By comparing captured images to predefined patterns or templates, they can identify discrepancies or errors in real-time. This automated inspection process enhances quality assurance (QA), minimises human error, and boosts production efficiency.

For quality control, ML-based vision systems rely on various data sources, such as high-resolution images captured by industrial cameras, scanned documents of label designs, and barcode/QR code information. They also use optical character recognition (OCR) outputs to extract text from labels, product databases to verify product details, historical inspection records to monitor past quality checks, and manufacturing batch data to confirm expected label content. These data sources are essential for training ML models and providing reference information for real-time quality inspections in the factory.

2.5. Cost to Serve (CTS) analysis

After product manufacturing, order fulfilment becomes crucial, with Cost to Serve (CTS) analysis playing a key role in assessing the total expenses of delivering goods and services to consumers, while also managing profitability[33;34].

In the CPG sector, CTS analysis evaluates all costs associated with product delivery to customers. Implementing ML models to optimise CTS can significantly boost efficiency and profitability. The main goal is to reduce overall CTS while maintaining or improving service quality and customer satisfaction. To accomplish this objective, several steps can be taken by leveraging data from the Warehouse Management System (WMS) and Transportation Management System (TMS):

- Linear programming or genetic algorithms can be employed to determine the most profitable product mix for different channels and customers.
- To improve profitability using an ML model that analyses cost-to-serve and determines the minimum order quantity (MOQ), regression models can be combined to predict cost and profitability, and then optimisation techniques can be applied to identify the ideal MOQ.
- Algorithms such as genetic algorithms or particle swarm optimisation can be implemented to find the most cost-effective delivery routes. Additionally, clustering algorithms like K-means or hierarchical clustering can be used to group delivery locations before applying routing algorithms.

Table 1
Description of key data sources in the manufacturing data model

Data Type	Description	Typical Sources	Typical Modality
Process data	Equipment settings (e.g. flow rate, feed rate, temperature, pressure etc.) used to monitor & control process performance	Historical data from historians; real-time tags from equipment Programmable Logic Controllers (PLCs); in-line sensors and IoT devices	Structured data; streaming data
Quality data	Raw material attributes / analysis, results of in-process test and post-batch tests	ERP, LIMS, Execution Systems, Quality management systems, vendor provided CofA, chromatography systems	Structured data
SME know-how	Feedback loop from SMEs applying domain-knowledge to the ML models (e.g. during feature engineering; reward and penalty assigned to the model during reinforcement learning etc.)	Adhoc	
Other data sources	Out of Specification (OOS)/ Out of Trend (OOT) data; directed graphs representing manufacturing process flow; formulation composition data; write-off wastes data etc.	Text files, images, audio, reports from in-process inspection systems	Structured or unstructured data
Environmental data	Room temperature, humidity, pressure etc.	Environmental Management System (EMS)/ Building Management System (BMS)	Structured data
Digital images	Photos of products and labels; digital scans of approved label designs		Unstructured data
Warehouse data	Costs associated with storing products, including rent, utilities, and labour; costs for picking, packing, and preparing products for shipment	Warehouse Management System (WMS)	Structured data
Transportation data	Expenses related to shipping products from manufacturing facilities to warehouses, and from warehouses to customers	Transportation Management System (TMS)	Structured data

- Reinforcement Learning (RL) techniques can be leveraged for continuously improving decision-making and enhancing profitability by learning from past actions (e.g. historical shipment delivery routes and delivery dates for different order quantities) and outcomes (e.g. total transportation costs). RL algorithms can then suggest optimal future adjustments to product delivery planning to maximise rewards (e.g. cost savings and Full Truck Load (FTL)).

3. Multimodal manufacturing data model

Manufacturing data encompasses qualitative and quantitative information, time-series data, controllable and uncontrollable attributes, present and historical data, measured and simulated data, observable data and process state variables[35]. Recently, the AI community has increasingly developed multimodal models, particularly with the rise of Generative AI, to learn from diverse data sources such as text, images, audio, structured information, 3D models and computational graphs[36]. These models capture complex information effectively by integrating multiple data strands.

Humans learn by engaging multiple senses (e.g. vision and hearing with reading text), as multimodal enrichment aids sensory and motor regions in cross-modal perceptual processing[37]. Viewing the human brain as a knowledge graph interlinking various senses to create comprehensive understanding, a multi-sensory AI model similarly integrates diverse data types from the natural environment, reflecting their cross-modal relationships. Cross-modal learning enhances AI models by providing holistic intelligence, reducing bias, and making traditionally opaque models (e.g. neural networks) more accurate, scalable, and transparent.

However, leveraging such big data in manufacturing in practice continues to be at a nascent stage. Figure 1 illustrates a comprehensive manufacturing data model for digital transformation in factories of the future, which is

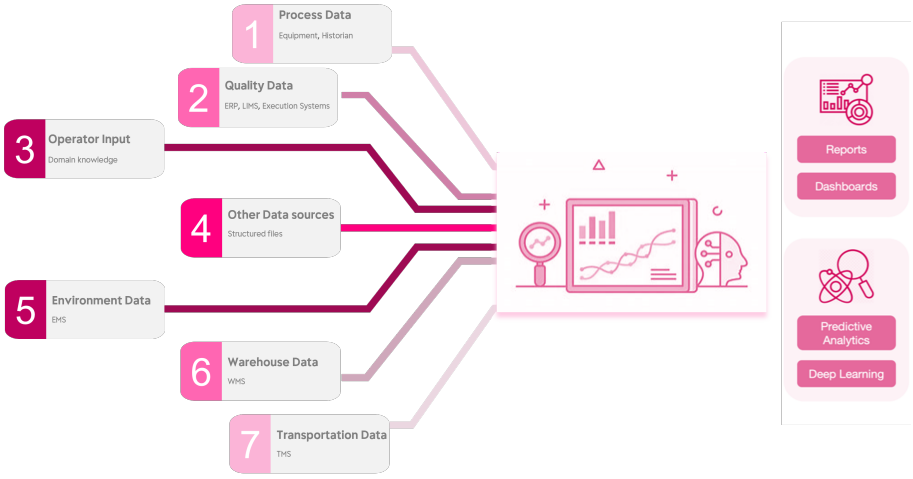


Figure 1: Summary of typical data sources and their modalities in the manufacturing data model

both generic and scalable across various products and manufacturing stages (e.g. granulation, coating, compression, packaging). This is akin to creating a comprehensive digital twin of the entire manufacturing process, from new product development in R&D through to logistics and supply chain, where finished products are produced and delivered to end consumers. Both structured and unstructured data sources throughout the factory can be utilised in this model.

Table 1 summarises typical data sources and modalities. When building a multimodal data model, a key factor is identifying the relevant attributes to map and integrate data from various sources, e.g. production order dates and batch/order numbers. It is also crucial to establish a data hierarchy that clearly links different stages of the manufacturing process.

Operator input is central to the manufacturing data model. Consistent interdependencies between different data types, ML models, and production problems are rare due to the domain-specific nature of manufacturing issues. Data complexity varies from factory to factory and requires insights from subject matter experts (SMEs) for effective decision-making[7]. Real-world validation of ML model insights is essential – SMEs, such as process engineers, can assess the model using their domain knowledge. Improving AI models with factory data is a long-term, iterative process that needs SME input to eliminate parameters that may be mathematically significant but irrelevant in practice.

Furthermore, the model can be applied to real-time data to guide the manufacturing process, with outputs validated to ensure critical parameters remain within the optimal range during production. Human oversight, or a *human-in-the-loop*, is a vital component of the manufacturing data model.

4. Discussion

Once a foundational manufacturing data model is established, the possibilities for analysis and decision support are manifold. Both traditional ML models (e.g. linear and logistic regression, Support Vector Machines (SVM), decision trees, random forests, eXtreme Gradient Boosting (XGBoost), Naïve-Bayes, etc.) and deep learning models (e.g. Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), transformers, Generative Adversarial Networks (GAN), etc.) can be utilised. These models handle classification tasks (e.g. predicting worker safety from real-time video streams), regression (e.g. predicting shelf life during stability testing), root cause analysis (e.g. identifying causes of waste during CPV using causal inference methods like the Peter-Clark algorithm or Causal Forests), and optimisation (e.g. improving delivery routes with genetic algorithms or Deep Q-Learning for CTS analysis).

Clustering techniques, such as K-Means or Fuzzy k-means, can be employed when historical data is limited (e.g. to identify product families with frequent failures during QC/QA). Generative AI techniques can summarise complex analytical insights into actionable reports for SMEs, utilising Large Language Models (LLMs) to create detailed factory reports based on model insights.

Our experience suggests that simpler ML models are often sufficient for tasks related to OEE optimisation in factories, thanks to the availability of structured data in most large factories. These simpler models can provide clearer and more transparent outcomes. For more complex scenarios involving unstructured data (e.g. computer vision models for worker safety or packaging defect identification), deep learning models may be used. However, these black-box models can be challenging to explain to business stakeholders. When using such models, it is crucial to employ Explainable AI methods (such as Shapley values (SHAP) analysis or attention mechanisms) and integrate domain-specific information (e.g., through vector databases and knowledge graphs) to address any safety and business impact in manufacturing.

5. Conclusion

In this golden age of AI and abundant data in factories, there has never been a better opportunity to enhance manufacturing through multimodal data enrichment. Traditional approaches to leverage data in silos for AI-driven decision making in factories have faced a long-standing gap in trust and scalability. We have outlined the benefits of a comprehensive multimodal manufacturing data model, which involves creating a digital twin of the entire manufacturing process to enable cross-modal learning. This approach can significantly boost the adoption of AI, supporting continuous improvements in the supply chain.

We emphasise that AI will not replace SMEs but will assist them in making smarter decisions in future factories. AI can improve OEE, reduce waste, and support real-time release through AI-assisted CPV and predictive stability. By utilising all available historical data to its fullest potential, AI can help identify critical associations and issues, acting as an enabler for more intelligent decision-making.

Acknowledgements

We would like to thank colleagues at Reckitt for their support.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. It is declared that Joyjit Chatterjee and Rajendra Shankaran are employed at Reckitt.

References

[1] World Economic Forum, 2022. URL: https://www.weforum.org/impact/carbon-footprint-manufacturing-industry/.

[2] A. Kedar, R. Lakhe, V. Deshpande, P. Washimkar, M. Wakhare, A comparative review of tqm, tpm and related organisational performance improvement programs, in: 2008 First International Conference on Emerging Trends in Engineering and Technology, 2008, pp. 725–730. doi:10.1109/ICETET.2008.133.

[3] B. Hrnjica, S. Softic, Explainable AI in Manufacturing: A Predictive Maintenance Case Study, in: B. Lalic, V. Majstorovic, U. Marjanovic, G. von Cieminski, D. Romero (Eds.), IFIP International Conference on Advances in Production Management Systems (APMS), volume AICT-592 of *Advances in Production Management Systems. Towards Smart and Digital Manufacturing*, Springer International Publishing, Novi Sad, Serbia, 2020, pp. 66–73. URL: https://inria.hal.science/hal-03635701. doi:10.1007/978-3-030-57997-5_8, part 2: Digital Transformation Approaches in Production Management.

[4] T. Yang, X. Yi, S. Lu, K. H. Johansson, T. Chai, Intelligent manufacturing for the process industry driven by industrial artificial intelligence, Engineering 7 (2021) 1224–1230.

[5] L. K. Vora, A. D. Gholap, K. Jetha, R. R. Thakur, H. K. Solanki, V. P. Chavda, Artificial intelligence in pharmaceutical technology and drug delivery design, Pharmaceutics 15 (2023) 1916.

[6] S. Patel, M. Patel, M. Kulkarni, M. S. Patel, De-interact: A machine-learning-based predictive tool for the drug-excipient interaction study during product development—validation through paracetamol and vanillin as a case study, International Journal of Pharmaceutics 637 (2023) 122839.

[7] D. Weichert, P. Link, A. Stoll, S. Rüping, S. Ihlenfeldt, S. Wrobel, A review of machine learning for the optimization of production processes, The International Journal of Advanced Manufacturing Technology 104 (2019) 1889 – 1902.

[8] L. D. C. Ng Corrales, M. Lambán, M. Hernandez Korner, J. Royo, Overall equipment effectiveness: Systematic literature review and overview of different approaches, Applied Sciences 10 (2020) 6469.

[9] Fiix, What is oee?: Oee calculations, 2023. URL: https://fiixsoftware.com/maintenance-metrics/overall-equipment-effectiveness/.

[10] Vorne Industries, World-class oee, 2024. URL: https://www.oee.com/world-class-oee/.

[11] M. Hassan, N. A. B. A. Wahab, R. B. M. Nor, The role of artificial intelligence in waste reduction in the beverage industry: A comprehensive strategy for enhanced sustainability and efficiency, AI, IoT and the Fourth Industrial Revolution Review 13 (2023) 1–8.

[12] N. S. Sanjay Bajaj, Dinesh Singla, Stability testing of pharmaceutical products, Journal of Applied Pharmaceutical Science (2012) 129–138.

[13] K. Huynh-Ba, M. Dong, Stability studies and testing of pharmaceuticals: An overview, LCGC North America 38 (2020) 325.

[14] H. Khan, M. Ali, A. Ahuja, J. Ali, Stability testing of pharmaceutical products - comparison of stability testing guidelines, Current Pharmaceutical Analysis 6 (2010) 142–150.

[15] S. Mani, Stability studies: A review, Asian Journal of Pharmaceutical Analysis and Medicinal Chemistry 1 (2013) 184.

[16] H. Veith, M. Zaeh, C. Luebbert, N. Rodríguez-Hornedo, G. Sadowski, Stability of pharmaceutical co-crystals at humid conditions can be predicted, Pharmaceutics 13 (2021).

[17] J. M. Hawkins, The importance of the laboratory to the pharmaceutical business, Journal of Automated Methods & Management in Chemistry 22 (2000) 47–52.

[18] R. Magari, Assessing shelf life using real-time and accelerated stability tests, BioPharm International 16 (2003) 36–48.

[19] K. De Vore, A six-sigma approach to stability testing, Journal of Pharmaceutical and Biomedical Analysis 47 (2008) 413–421.

[20] S. Dubey, H. Jangala, Laboratory information and management system: A tool to increase laboratory productivity, Clinical Research and Regulatory Affairs 29 (2012).

[21] T. S. Picker, Digitalization in laboratories of the pharmaceutical industry, Solid State Development and Processing of Pharmaceutical Molecules (2021) 397–420.

[22] D. Weininger, Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules, J. Chem. Inf. Comput. Sci. 28 (1988) 31–36.

[23] K. R. Konnaiyan, Artificial intelligence (ai) based continued process verification (cpv), 2023. URL: https://ispe.org/pharmaceutical-engineering/ispeak/artificial-intelligence-ai-based-continued-process-verification.

[24] D. Singh, S. Singh, S. Bajaj, Regulatory Guidelines on Stability Testing and Trending of Requirements, 2018, pp. 1–30. doi:10.1007/978-1-4939-7686-7_1.

[25] Perceptive Engineering Ltd., Continuous process verification, 2024. URL: https://www.perceptiveapc.com/industries/pharmaceutical/solutions/cpv/.

[26] E. Hanley, Continuous & continued process verification, 2016. URL: https://www.pharmout.net/wp-content/uploads/2016/07/2015_GMP_Validation_Forum_D2.T4.2.1-Eoin-Hanley-Continuous-Continued-Process-Verification.pdf.

[27] D. Markl, M. Warman, M. Dumarey, E.-L. Bergman, S. Folestad, Z. Shi, L. F. Manley, D. J. Goodwin, J. A. Zeitler, Review of real-time release testing of pharmaceutical tablets: State-of-the art, challenges and future perspective, International Journal of Pharmaceutics 582 (2020) 119353.

[28] A. D. Bois, Real-time release testing in pharmaceutical manufacturing, 2024. URL: https://www.news-medical.net/life-sciences/Real-Time-Release-Testing-in-Pharmaceutical-Manufacturing.aspx.

[29] M. Pech, J. Vrchota, J. Bednář, Predictive maintenance and intelligent sensors in smart factory: Review, Sensors 21 (2021).
[30] Spotfire, Predictive maintenance: Enhancing equipment reliability and reducing downtime, 2024. URL: https://www.spotfire.com/glossary/what-is-predictive-maintenance.
[31] P. Girdhar, C. Scheffer, 1 - predictive maintenance techniques: Part 1 predictive maintenance basics, in: P. Girdhar, C. Scheffer (Eds.), Practical Machinery Vibration Analysis and Predictive Maintenance, Newnes, Oxford, 2004, pp. 1–10. URL: https://www.sciencedirect.com/science/article/pii/B9780750662758500011. doi:https://doi.org/10.1016/B978-075066275-8/50001-1.
[32] A. Akundi, M. Reyna, A machine vision based automated quality control system for product dimensional analysis, Procedia Computer Science 185 (2021) 127–134. Big Data, IoT, and AI for a Smarter Future.
[33] R. Guerreiro, S. Bio, E. Merschmann, Cost-to-serve measurement and customer profitability analysis, International Journal of Logistics Management, The 19 (2008) 389–407.
[34] The Supply Chain Consulting Group Ltd., Understanding cost-to-serve in the supply chain operations, 2024. URL: https://sccgltd.com/understanding-cost-to-serve-in-the-supply-chain-operations/#:~:text=Cost%2Dto%2Dserve%20is%20an,which%20to%20base%20strategic%20decisions.
[35] A. Ondracka, A. Gasset, X. García-Ortega, D. Hubmayr, J. B. van Wijngaarden, J. L. Montesinos-Seguí, F. Valero, T. Manzano, Cpv of the future: Ai-powered continued process verification for bioreactor processes, PDA Journal of Pharmaceutical Science and Technology (2022).
[36] Gemini Team Google et al., Gemini: A family of highly capable multimodal models, 2024. URL: https://arxiv.org/abs/2312.11805. arXiv:2312.11805.
[37] B. Mathias, K. von Kriegstein, Enriched learning: Behavior, brain, and computation, Trends in Cognitive Sciences 27 (2023) 81–97. Open Access via the Elsevier Agreement Funder: German Research Foundation: KR 3735/3-1,MA 9552/1-1.

About the authors



Joyjit Chatterjee is IT & Digital Data Scientist – Continuous Improvement/Senior Business Analyst at Reckitt, UK. A Forbes 30 Under 30 Lister, he holds a PhD in Computer Science from the University of Hull, UK. Joyjit has also received the Green Talents Award from German government. His work has been covered by Forbes, World Economic Forum, Times of India etc. Joyjit has been endorsed as a Global Talent by UKRI and received recognition for his past research from Innovate UK. He regularly features in high-profile events as invited speaker and panellist alongside Chief Information Officers and Senior Vice Presidents at Fortune 500 companies.



Rajendra Shankaran is a seasoned professional with extensive cross-industry experience, currently spearheading data and analytics initiatives in the manufacturing domain for a leading Fast-Moving Consumer Goods (FMCG) company. His expertise spans Industry 4.0 technologies, including Industrial Internet of Things (IIoT), Augmented/Virtual Reality (AR/VR), and Artificial Intelligence/Machine Learning (AI/ML), where he has pioneered innovative solutions to enhance manufacturing processes. Rajendra excels at bridging the gap between Information Technology (IT) and Operational Technology (OT) and has been instrumental in fostering innovation and efficiency across the production landscape and driving digital transformation and smart manufacturing practices.



Nina Dethlefs is a Reader in Computer Science at the University of Hull, UK, where she leads and conducts research into natural language processing, applied machine learning and wider artificial intelligence. She is the founding head of the Big Data Analytics research group, and Aura Centre for Doctoral Training Theme lead for "Big data, sensors and digitalisation for the offshore environment". Nina has a PhD in Computational Linguistics from the University of Bremen, Germany. She has served on many scientific committees for leading conferences and journals and has been area chair for ACL, and COLING.