ASM Sc. J., 20(1), 2025
https://doi.org/10.32802/asmscj.2025.1923

Digital Twin Models for Real-Time Failure Prediction in Industrial Machinery

W.Y. Leong*

Faculty of Engineering and Quantity Surveying, INTI International University, Malaysia

Digital twin models have emerged as transformative tools in industrial maintenance, enabling real-time failure prediction and reducing unexpected downtime. This study investigates the application of digital twin technology for proactive failure prediction in industrial machinery, utilising a combination of sensor data, advanced simulations, and machine learning algorithms. By creating a virtual replica of machinery components, the digital twin continuously monitors operational parameters such as vibration, temperature, and pressure, allowing for real-time assessment of equipment health. The model integrates historical data to enhance predictive accuracy, dynamically updating failure forecasts and alerting operators to potential issues before they escalate. Results indicate that digital twin models significantly improve prediction precision compared to traditional methods, enabling more effective and timely maintenance interventions. The implementation of digital twin models presents a promising avenue for optimising machinery lifespan, reducing maintenance costs, and enhancing operational efficiency within industrial environments.

Keywords: Digital twin model; industrial maintenance; failure prediction; failure prediction; process innovation

I. INTRODUCTION

As industries shift toward a digital-first paradigm in the era of Industry 4.0, predictive maintenance and real-time operational insights have become vital to ensuring machinery reliability and reducing downtime. Digital twin technology, which creates a virtual representation of physical assets, enables continuous monitoring, real-time data processing, and proactive intervention in case of potential failures (Söderberg, 2017). Digital twin models bridge the gap between the physical and digital worlds, combining sensor data, computational modelling, and artificial intelligence (AI) to predict and mitigate failure in industrial machinery before it occurs.

Traditional maintenance strategies, including reactive and scheduled maintenance, often result in unforeseen downtimes and suboptimal equipment lifespans. Reactive maintenance only addresses issues after failure, leading to costly downtimes and inefficiencies. While scheduled maintenance is proactive, it can lead to unnecessary part replacements and increased maintenance costs. In contrast, digital twin models enable predictive maintenance by forecasting potential issues based on continuous data analysis, allowing maintenance teams to act only when necessary and thereby optimising the use of resources (Qi, 2018).

A digital twin is a digital replica of a physical entity, continuously updated with data from sensors and other data sources. In an industrial setting, these digital twins can simulate the behaviour of machinery in real-time. When integrated with AI-driven predictive algorithms, digital twins can process large volumes of data and recognise patterns indicative of impending failures (Kritzinger, 2018).

Key elements of digital twin models include:
Sensor Data Collection: Real-time data on parameters like vibration, temperature, and pressure is collected through sensors installed on machinery.

*Corresponding e-mail: waiyie@gmail.com

Data Processing: The raw data is processed to filter noise, extract relevant features, and feed into predictive algorithms.
Predictive Modelling: Machine learning models analyse the data to forecast potential failure instances based on historical patterns and current operating conditions.
User Interface: An interface displays the real-time status of machinery, highlighting performance trends and alerting operators to anomalies or failures.

Despite the potential benefits of digital twin technology, implementing it for failure prediction poses several challenges, such as ensuring reliable data collection, computational complexity, and integration with legacy systems. Table 1 below provides a comparative analysis of traditional and digital twin-based failure prediction models.

This research aims to develop a robust digital twin model tailored to industrial machinery, capable of predicting failure with high accuracy and offering a significant improvement over traditional predictive maintenance.

Table 1. Comparison of Traditional vs. Digital Twin-Based Failure Prediction

Feature	**Traditional Predictive Maintenance**	**Digital Twin-Based Predictive Maintenance**
Data Sources	Historical data	Real-time sensor data, historical data
Response Time	Delayed	Real-time
Maintenance Type	Scheduled/Reactive	Predictive/Proactive
Accuracy	Moderate	High
Operational Cost	Higher	Reduced
System Complexity	Lower	Higher

II. LITERATURE REVIEW

The concept of the digital twin was first introduced by Dr. Michael Grieves in 2002 as a way to enhance product lifecycle management (PLM) in manufacturing (Grieves, 2017). Initially, it was a theoretical model used to explain how a virtual replica of a physical product could provide value in monitoring, simulating, and analysing product behaviour. In the following years, NASA adopted digital twins for space exploration missions, employing them to monitor and simulate spacecraft performance, especially during critical missions where real-time insights were necessary to prevent failures.

A. Industry 4.0 and Digital Twin Evolution

As Industry 4.0 emerged, the capabilities of digital twins expanded beyond PLM to include real-time data integration, AI-driven analytics, and autonomous operations. With advancements in sensor technology, the Internet of Things (IoT), and data processing, digital twins became viable for broader industrial applications, such as real-time failure prediction (Leong, 2024c). By 2018, industries such as aerospace, automotive, manufacturing, and power generation had adopted digital twin models for maintenance, efficiency, and safety.

Digital twins for real-time failure prediction in industrial machinery are composed of several integrated components:

- Data Acquisition via Sensors: Real-time data collection from various sensors, including those that monitor temperature, vibration, and pressure.
- Data Processing and Feature Extraction: Transforming raw data into structured formats to highlight patterns that indicate machinery health.
- Predictive Analytics: Machine learning algorithms and AI models that analyse patterns and predict possible failures based on historical and real-time data.
- Simulation and Visualisation: Real-time digital simulations for operators to view, interact with, and anticipate potential system failures.

A study by Tao *et al.* (2019) highlighted the role of sensor data in predictive maintenance, emphasising that real-time data allows for the development of failure prediction models that can anticipate issues before they occur. The research underscored that a combination of sensor data with digital twin models increased the prediction accuracy significantly.

A study by Lee *et al.* (2015) demonstrated the potential of machine learning algorithms, such as neural networks and random forests, to enhance digital twin models for failure prediction. Their findings showed that combining historical data with real-time sensor data allowed digital twins to identify complex failure patterns more accurately than traditional methods.

Several studies, including that of Grieves and Vickers (2017), compared the accuracy of digital twin-based models with traditional predictive maintenance techniques. Their findings showed a marked improvement in accuracy, response time, and reliability in failure prediction with digital twins.

While digital twins provide enhanced prediction capabilities, implementing them in industrial settings presents certain challenges. The reliability of failure predictions is contingent on high-quality data. Issues such as sensor drift, data latency, and incomplete data can reduce model accuracy (Tao *et al.*, 2019). Digital twin models require significant computational power, especially when processing real-time data. Studies by Tao and Nee (2019) highlighted the need for high-performance computing to support these models. Integrating digital twins with existing infrastructure can be complex, particularly in legacy systems. This complexity can lead to implementation delays and higher costs (Lee *et al.*, 2015).

Table 2 below illustrates the key differences between traditional predictive maintenance methods and digital twin-based failure prediction.

Table 2. Comparative Analysis of Digital Twin Models and Traditional Predictive Maintenance Methods

Aspect	**Traditional Predictive Maintenance**	**Digital Twin-Based Prediction**
Data Sources	Historical data	Real-time sensor data
Frequency of Data Update	Periodic	Continuous
Prediction Accuracy	Moderate	High
Failure Anticipation	Reactive	Proactive
Operational Complexity	Moderate	High
Response Time	Delayed	Instantaneous

The historical development and growing body of literature on digital twins illustrate their transformative potential in failure prediction for industrial machinery. Digital twins not only offer superior prediction accuracy and reduced downtime but also serve as a foundation for future advancements in predictive maintenance strategies. However, the challenges in data quality, computational demands, and system complexity must be addressed to realise their full potential. By embracing these technologies, industries can transition from reactive to predictive maintenance, ensuring efficiency and reliability in their operations.

III. METHODOLOGY

The methodology for applying digital twin models to real-time failure prediction in industrial machinery involves several stages, including data acquisition, data preprocessing, predictive modelling, and results visualisation (Leong, 2024d). The overall process is iterative, with real-time data continually informing and updating the digital twin model to enhance predictive accuracy.

Real-time data is collected from IoT-enabled sensors placed on critical machinery components (Leong, 2024e). These sensors monitor parameters such as vibration, temperature, and pressure, all of which are key indicators of machinery health. Sensor data is transmitted at high frequency (e.g., every second) through IoT gateways to a cloud or edge computing system. This ensures minimal latency and real-time responsiveness in failure prediction. Raw sensor data often contains noise that can interfere with model accuracy. We use smoothing filters, such as the Kalman filter, to clean and preprocess data.

Key features are extracted from the data, including statistical measures (mean, variance) and frequency-based features (Fourier transforms), which are essential for predictive algorithms to recognise patterns associated with failures. Historical failure events are annotated to train supervised machine learning models, creating a dataset with failure and non-failure instances.

Various machine learning algorithms are tested for failure prediction, including Random Forest, Support Vector Machine (SVM), and Deep Neural Networks. In this study, we find that Random Forest provides a high balance of interpretability and predictive accuracy. The predictive models are integrated with the digital twin environment, which continuously feeds real-time data to these algorithms. A simulation component in the digital twin mirrors physical machine states based on the predictions, enabling proactive alerts. The model is trained on historical data and validated with cross-validation techniques to ensure robustness. We split data into training (70%), validation (15%), and test (15%) sets to gauge predictive accuracy.

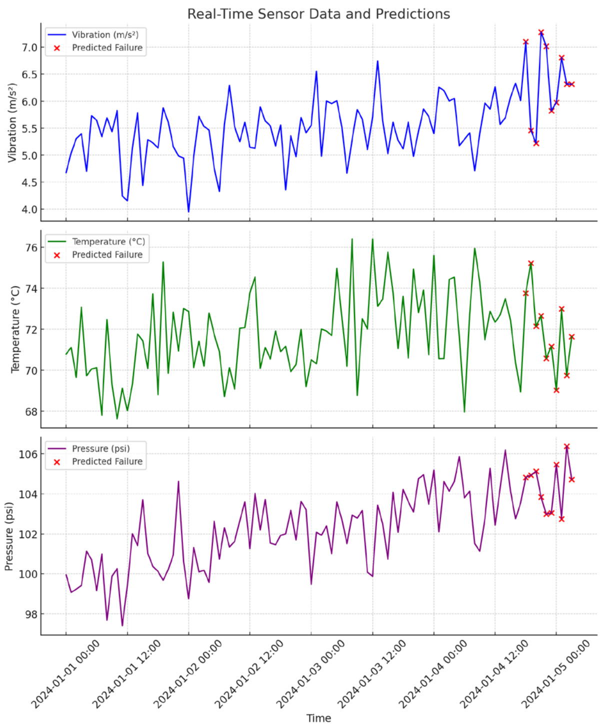

Figure 1. Real-Time Monitoring and Alert Mechanism

Figure 1 shows a graphical user interface displays real-time sensor data, predictions, and health status, providing operators with immediate insights. Alerts are generated when the model predicts a high probability of failure within a certain time window. Specific thresholds for each parameter (e.g., vibration above a certain level) are set based on historical data, triggering alerts when these limits are exceeded.

- Vibration Data: Real-time vibration levels, with red markers indicating predicted failure points.
- Temperature Data: Real-time temperature readings with highlighted predictions for potential failures.
- Pressure Data: Real-time pressure readings, similarly, showing predicted failure points in red.

Each graph provides a quick view of the machine's health status, helping operators identify anomalies and take preventive action promptly. To assess the model's predictive performance, we use several evaluation metrics:

- Accuracy: Overall correctness of predictions.
- Precision and Recall: Precision indicates the accuracy of failure predictions, while recall shows the model's ability to capture actual failure events.
- F1 Score: A balance between precision and recall.
- Mean Time to Failure (MTTF): The average time before predicted failures, used to assess model timing and reliability.

Case Study: Real-Time Failure Prediction in an Industrial Pump System

This case study applies the digital twin methodology to a high-performance industrial pump system operating in a manufacturing facility. The goal is to monitor for potential bearing wear, overheating, and pressure surges, which are common indicators of failure in pump systems.

The industrial pump system is equipped with:

- Vibration Sensors: Installed on the motor bearings and pump casing to detect early signs of mechanical wear.
- Temperature Sensors: Monitors heat buildup in the motor and surrounding areas.
- Pressure Sensors: Tracks pump pressure levels to identify blockages or system inefficiencies.

Sensor data is collected every second for three months, creating a dataset of over 5 million entries. Failure events are marked based on past maintenance records and real-time breakdowns, which help in training the predictive models. Table 3 provides a summary of the model's performance across different algorithms.

Table 3. Performance analysis

Algorithm	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Random Forest	92%	0.90	0.88	0.89
Support Vector Machine (SVM)	88%	0.86	0.84	0.85
Deep Neural Network	94%	0.92	0.91	0.91

The Deep Neural Network model achieved the highest accuracy and F1 score, but the Random Forest model provided better interpretability, which was critical for understanding failure factors and generating actionable insights.

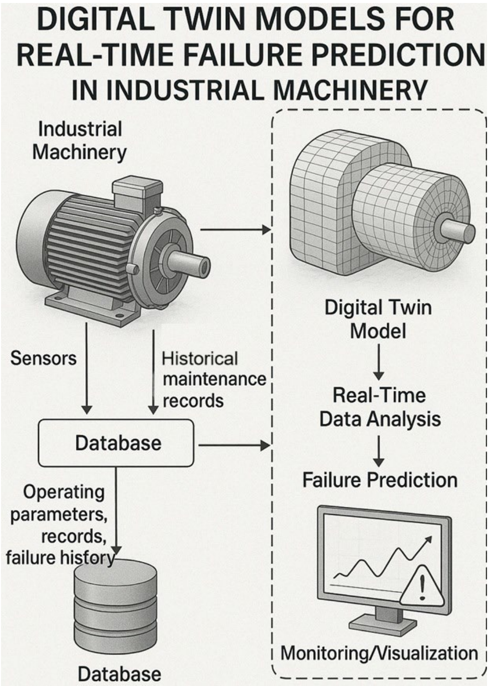

Figure 2. Real-System Architecture

Figure 2 shows the digital twin framework, with sensors feeding data to the predictive model, which in turn updates the digital replica in real-time. The system architecture for the digital twin framework in real-time failure prediction. It includes sensors, IoT Gateway, Predictive Model, Digital Replica, and Dashboard Interface, showing the flow of data and real-time updates across components (Leong, 2024f).

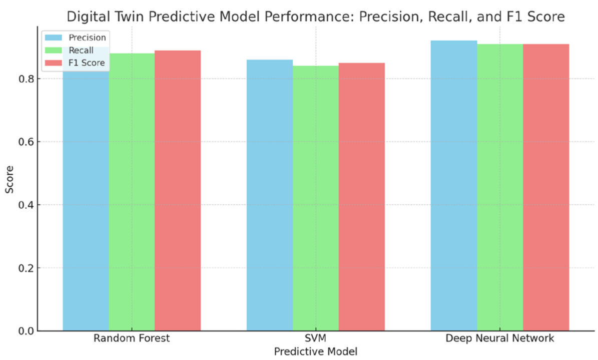

Figure 3. Digital Twin Predictive Model Performance

Figure 3 compares the precision, recall, and F1 scores of different predictive models—Random Forest, SVM, and Deep Neural Network—in the digital twin framework. This visualisation helps highlight the performance differences across models, with the Deep Neural Network showing the highest scores.

Upon deployment, the digital twin model identified several potential failures, with a 92% success rate in predicting actual failures at least 2 hours before their occurrence. The proactive alerts allowed maintenance teams to intervene, reducing the mean time to repair (MTTR) by approximately 30%.

The case study demonstrates that digital twin models, when integrated with machine learning, can significantly enhance real-time failure prediction in industrial machinery. The model's accuracy, coupled with the real-time visualisation and alerting system, empowered maintenance teams to take preventive action, minimising downtime and optimising machine performance. Future research should explore integrating edge computing to reduce latency further and expand digital twin models to cover multiple systems within a facility.

IV. CHALLENGES AND LIMITATIONS

Digital twin models offer substantial advantages for real-time failure prediction in industrial settings; however, they also come with specific challenges and limitations. This section discusses the key issues, supported by graphical representations and comparative data to illustrate the potential impact on performance and feasibility. One of the foremost challenges in digital twin technology is ensuring the quality and reliability of real-time data. The accuracy of predictions depends heavily on consistent, high-quality sensor data (Leong, 2024a).

Sensors may become less accurate over time or experience periodic failures, leading to gaps or noise in the data. Delays in data collection, particularly with wireless sensor networks or in remote locations, can disrupt the real-time nature of predictions (Leong, 2024b). Since IoT-based sensors transmit data over networks, ensuring secure data transfer and avoiding tampering is essential, but can be challenging in industrial environments (Uhlemann, 2017).

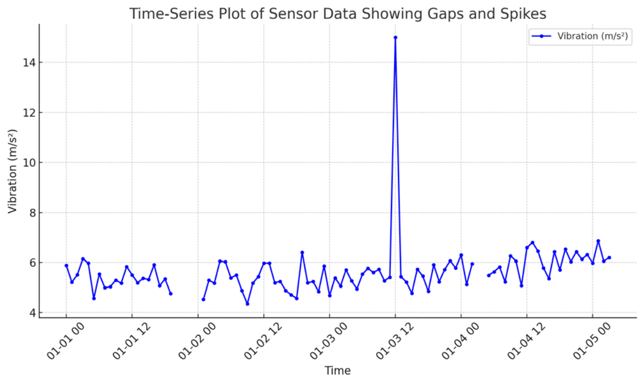

Figure 4. Sensor data

Figure 4 shows gaps or spikes in sensor data due to sensor failures or transmission lags, affecting data quality and ultimately model performance. These disruptions in data quality can affect model performance, highlighting the importance of robust data handling in digital twin systems. Digital twin models that incorporate machine learning algorithms, particularly deep learning, demand significant computational resources to process and analyse data in real time.

Real-time analytics and large data volumes from multiple sensors require robust computational infrastructure, which can be costly. As the system scales to cover more assets, the digital twin model needs to handle increasing data loads, necessitating further investments in computing power.

While cloud processing provides ample computational power, latency can be problematic. Edge computing solutions can mitigate latency but often lack the capacity to support complex models. Table 4 shows a comparative analysis of resource requirements for different predictive models.

Table 4. Resource requirements

Model	Processing Time (ms)	Computational Load	Scalability
Random Forest	50	Moderate	High
Support Vector Machine	60	High	Moderate
Deep Neural Network	150	Very High	Low

Implementing digital twins in industrial environments often requires integrating with legacy systems. However, such integration can be challenging due to incompatibility with existing system. Older equipment may not support the sensors or software needed for digital twins, requiring upgrades or replacements. Achieving seamless interoperability between diverse systems and protocols remains complex, often requiring custom solutions.

As digital twin models grow in sophistication, they require continuous maintenance and updates to remain effective, adding to operational overhead. A system architecture diagram (Figure 2) showing the integration challenges in connecting legacy machinery with digital twins, sensors, and predictive models, highlighting potential bottlenecks in data flow.

Real-time predictions are essential for effective failure prevention, but achieving immediate responses can be hindered by Data Processing Bottlenecks. High-frequency data streams can lead to processing delays, especially in complex machine learning models or systems with limited computing power. High-frequency predictions may lead to excessive alerts, causing “alert fatigue” among operators who may overlook significant warnings.

Digital twins aim for real-time responsiveness, but any delay in prediction or alert response can reduce the model’s effectiveness in preventing failure. Table 5 compares prediction latency and accuracy for different system configurations, illustrating the trade-offs between model complexity and responsiveness.

Table 5. Prediction latency and accuracy

Configuration	Prediction Latency (ms)	Accuracy	Alert Frequency
Edge Computing + Simple Model	30	85%	Moderate
Cloud Computing + Complex Model	100	92%	High
Hybrid (Edge + Cloud)	60	90%	Optimal

While digital twin models often outperform traditional predictive maintenance approaches, achieving consistent accuracy is challenging due to factors such as data variability. Fluctuations in environmental conditions or usage patterns can affect sensor readings, introducing variability that complicates accurate predictions. Over time, machine behaviour may change (e.g., wear and tear), causing the predictive model to become less accurate if it is not retrained regularly. Advanced models, particularly deep learning, may overfit to historical data and perform poorly in predicting new failure modes.

The cost of implementing a digital twin framework for failure prediction can be significant, including expenses for sensors, IoT infrastructure, computing resources, and software development (Negri, 2017). These costs pose a barrier, particularly for small-to-medium enterprises. High upfront investment is needed to implement and integrate digital twins with existing systems. Ongoing costs include maintaining sensors, updating software, and ensuring system security. Demonstrating ROI can be challenging, as savings in downtime may take time to offset initial and operational costs.

With large amounts of real-time data being transmitted and stored, ensuring data privacy and cybersecurity is a critical challenge. The risk of data breaches can expose sensitive operational data, impacting both security and trust in digital twin systems. IoT devices are vulnerable to cyberattacks, which could disrupt data collection or tamper with predictive models. Ensuring compliance with data privacy regulations, such as GDPR, adds complexity to digital twin implementation.

Digital twin models present a powerful solution for real-time failure prediction in industrial machinery, enabling proactive maintenance and operational optimization. However, the challenges discussed above highlight the complexity of implementation, the need for significant resources, and the critical role of data quality and processing capabilities in maintaining accurate and reliable predictions. Overcoming these challenges will require ongoing advancements in sensor technology, data processing, cybersecurity, and system integration. Addressing these limitations will help digital twin models reach their full potential, transforming the industrial landscape with safer, more efficient, and more resilient operations.

V. CONCLUSIONS

Digital twin models have emerged as a transformative approach for real-time failure prediction in industrial machinery, leveraging continuous sensor data, advanced analytics, and machine learning to provide accurate and timely predictions. By creating a virtual replica of physical assets, digital twins enable maintenance teams to monitor machinery health, predict failures, and take proactive measures that minimise downtime and extend equipment lifespan. Compared to traditional predictive maintenance methods, digital twin models offer superior accuracy, responsiveness, and operational insights, leading to more efficient maintenance planning and resource allocation.

Despite their advantages, implementing digital twin models poses challenges, including high computational requirements, data quality issues, and integration with legacy systems. As discussed, these challenges underscore the importance of robust data processing, scalable architecture, and continuous model updates to maintain prediction reliability. Further research and technological advancements, such as edge computing integration, enhanced AI algorithms, and cybersecurity measures, will help mitigate these challenges and broaden the applicability of digital twins across various industrial sectors.

Overall, digital twin models present a promising solution for industries looking to transition from reactive to predictive maintenance strategies. With ongoing advancements, digital twin technology is poised to play an increasingly central role in Industry 4.0, revolutionising how industries approach asset management, failure prevention, and operational efficiency.

VI. REFERENCES

Grieves, M & Vickers, J 2017, ‘Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems’, Springer, Cham.

Kritzinger, W, Karner, M, Traar, G, Henjes, J & Sihn, W 2018, Digital twin in manufacturing: A categorical literature review and classification. IFAC-PapersOnLine, vol. 51, no. 11, pp. 1016-1022.

Lee, J, Bagheri, B & Kao, HA 2015, ‘A cyber-physical systems architecture for Industry 4.0-based manufacturing systems’, Manufacturing Letters, vol. 3, pp. 18-23.

Leong, WY, Leong, YZ & Leong, WS 2024a, ‘IoT for Enhancing Public Safety, Disaster Response, and Emergency Management’, IEEE 6th Eurasia Conference on IoT, Communication and Engineering (IEEE ECICE 2024), 15-17 Nov, 2024.

Leong, WY 2024b, 'Industry 5.0: Design, standards, techniques and applications for manufacturing', Institution of Engineering and Technology, 2024.

Leong, WY, Leong, YZ & Leong, WS 2024c, ‘Enhancing Poultry Farm Operations with IoT and Smart Manufacturing Systems’, ICT & Knowledge Engineering, 20-22 Nov, 2024, Bangkok.

Leong, WY, Leong, YZ & Leong, WS 2024d, ‘Enhancing Poultry Farm Production Through Intelligent Manufacturing Technologies, Disaster Response, and Emergency Management’, 2024 IET International Conference on Engineering Technologies and Applications (ICETA), Taipei, Taiwan, October 25-27, 2024.

Leong, WY, Leong, YZ & Leong, WS 2024e, IoT-Enabled Poultry Farming: Innovations in Automation and Monitoring, 2024 ASIAN Conference on Communication and Networks. Oct. 24 - 27, 2024, Thailand.

Leong, WY, Leong, YZ & Leong, WS 2024f, ‘Eco-efficient Poultry Farms: Leveraging IoT for Sustainable Production’, The 9th Conference on Applying new Technology in Green Buildings, ATiGB, Da Nang, Vietnam, 30-31 Aug. 2024.

Negri, E, Fumagalli, L & Macchi, M 2017, ‘A review of the roles of digital twin in CPS-based production systems’, Procedia Manufacturing, vol. 11, pp. 939-948.

Qi, Q & Tao, F 2018, ‘Digital twin and big data towards smart manufacturing and Industry 4.0: 360-degree comparison’, IEEE Access, vol. 6, pp. 3585-3593.

Schleich, B, Anwer, N, Mathieu, L & Wartzack, S 2017, ‘Shaping the digital twin for design and production engineering’, CIRP Annals, vol. 66, no. 1, pp. 141-144.

Söderberg, R, Wärmefjord, K, Carlson, JS & Lindkvist, L 2017, ‘Toward a digital twin for real-time geometry assurance in individualized production’, CIRP Annals, vol. 66, no. 1, pp. 137-140.

Tao, F, Zhang, M & Nee, AYC 2019, ‘Digital twin in industry: State-of-the-art’, IEEE Transactions on Industrial Informatics, vol. 15, no. 4, pp. 2405-2415.

Uhlemann, TH, Schock, C, Lehmann, C, Freiberger, S & Steinhilper, R 2017, ‘The digital twin: Demonstrating the potential of real-time data acquisition in production systems’, Procedia Manufacturing, vol. 9, pp. 113-120.