Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA
Published: February 18, 2025

DOI: 10.46254/AN15.20250186

Robust Predictive Maintenance Framework Based on Federated Learning

Chao-Lung Yang, Hwai-Chen Qiu and Cheng-Ye Yang
Graduate Institute of Intelligent Manufacturing Technology
National Taiwan University of Science and Technology (NTUST)
Taipei, Taiwan
clyang@mail.ntust.edu.tw, d11251004@mail.ntust.edu.tw, m11101t07@mail.ntust.edu.tw

Abstract

This paper proposes a predictive maintenance framework based on Federated Learning to address the challenges of data privacy and efficiency in Industry 4.0 manufacturing. The framework employs a combined network called a one-dimensional convolutional neural network (1DCNN) and a bidirectional long short-term memory (BiLSTM) network (1DCNN-BiLSTM) model for predictions, enabling cross-client collaborative training without sharing raw data. Using the Microsoft Azure Predictive Maintenance dataset, our experimental results demonstrate that Federated Learning can achieve 81% prediction accuracy, representing a 27% improvement over individual models. The federated model maintains at least 71% prediction performance even with 50% reduced training data or 25% client participation while showing excellent generalization ability on unseen data. These results validate the framework's effectiveness in balancing data privacy, efficiency, and model performance in real-world applications.

Keywords
Industry 4.0, Predictive Maintenance, Federated Learning, Data Efficiency, Data Privacy

1. Introduction
In the manufacturing industry, predictive maintenance has emerged as a crucial strategy for reducing operational costs and enhancing competitive advantage. However, implementing effective predictive maintenance faces significant challenges, particularly in data acquisition and privacy protection. The performance of predictive maintenance systems heavily depends on large volumes of high-quality historical data, yet individual organizations often struggle with limited failure records (Wan, Tang et al. 2016, Budach, Feuerpfeil et al. 2022). While centralized cloud computing could address data limitations through cross-firm integration, it raises serious concerns about data security and privacy, as equipment data often contains sensitive business information that companies are reluctant to share (Weber, Xu et al. 2016).

This research seeks to establish a predictive maintenance framework that balances data privacy, data volume, and model performance. We propose combining federated learning with a supervised fault prediction model called one-dimensional convolutional neural network (1DCNN) and a bidirectional long short-term memory (BiLSTM) network (combined as 1DCNN-BiLSTM) to address these challenges (Ahn, Lee et al. 2023). The framework aims to facilitate cross-enterprise collaboration without data sharing, enabling model training with relatively limited data while maintaining model robustness and stability.

We investigate the following research questions:

RQ1. Is the federated learning model superior to local client models in terms of average predictive performance?
RQ2. Does the federated learning model outperform client models under reduced training data?
RQ3. Is the federated learning model resilient to reductions in participating clients?
RQ4. Can the federated model maintain predictive performance on unseen data?

The paper is organized as follows: Section 2 reviews relevant literature on predictive maintenance and federated learning. Section 3 details the methodology, including the federated learning framework and the model architecture. Section 4 presents the data collection and preprocessing steps. Section 5 discusses the experimental results, and Section 6 concludes with findings and future research directions.

2. Literature Review

Predictive maintenance (PdM) has emerged as a critical component of Industry 4.0, focusing on minimizing equipment downtime by predicting failures before they occur (Hashemian 2010). Traditional PdM faces challenges in data acquisition and privacy protection, which has led to the exploration of federated learning (FL) as a potential solution (Ge, Li et al. 2022). FL enables multiple data owners to train a shared model collaboratively without sharing raw data, thereby preserving privacy while improving model performance.

Recent studies have made significant progress in federated predictive maintenance (Fed-PdM). Table 1 summarizes key contributions in this field:

Table 1. Related literature on federated predictive maintenance

Algorithm	Model Architecture	Key Contributions	Results
Aggregation strategy (Shubyn, Kostrzewa et al. 2023)	FL-based LSTM	Weighted aggregation based on local model performance	Improved energy consumption prediction in AGVs
Time-series anomaly detection (Ahn, Lee et al. 2023)	FL-based 1D CNN-BiLSTM	Dynamic grouping method for heterogeneous data	Superior predictive accuracy under FlexCFL|
Reduce model parameter transmission (Liu, Garg et al. 2020)	FL-based AMCNN-LSTM	Top-k gradient compression mechanism	Reduced communication costs

Based on these foundations listed in Table 1, Shubyn's model aggregation strategy shows promising improvement on prediction (Shubyn, Kostrzewa et al. 2023) while FL-based 1D CNN-BiLSTM model shows the superior performance (Ahn, Lee et al. 2023). Therefore, our research leverages Shubyn's work and Ahn's 1D CNN-BiLSTM architecture and extend the previous work by focusing on FL's robustness under data volume fluctuations - a critical consideration for practical implementations.

For empirical validation, we utilized the Microsoft Azure Predictive Maintenance dataset, which contains time-series data from multiple devices. While previous studies have primarily applied this dataset to centralized machine learning research, we leverage it to evaluate FL's predictive performance and robustness in a distributed setting(Cardoso and Ferreira 2020).

The literature review reveals several gaps that our research addresses:

1. Limited exploration of FL performance under varying data volumes
2. Need for robust evaluation of model performance with reduced client participation
3. Lack of comprehensive analysis on model generalization to unseen data

3. Methods

3.1 Research Framework & Hypotheses

This research utilizes federated learning to address predictive maintenance challenges, with the following key assumptions:

- Equipment data contains proprietary information that should not be shared
- Dataset contains fault patterns predictive of failures
- Fault patterns from diverse environments may reveal unknown failure modes
- Devices can communicate in real-time for model updates

Figure 1 shows the framework comprising clients (IoT devices collecting real-time data) and a server (cloud server for model aggregation).

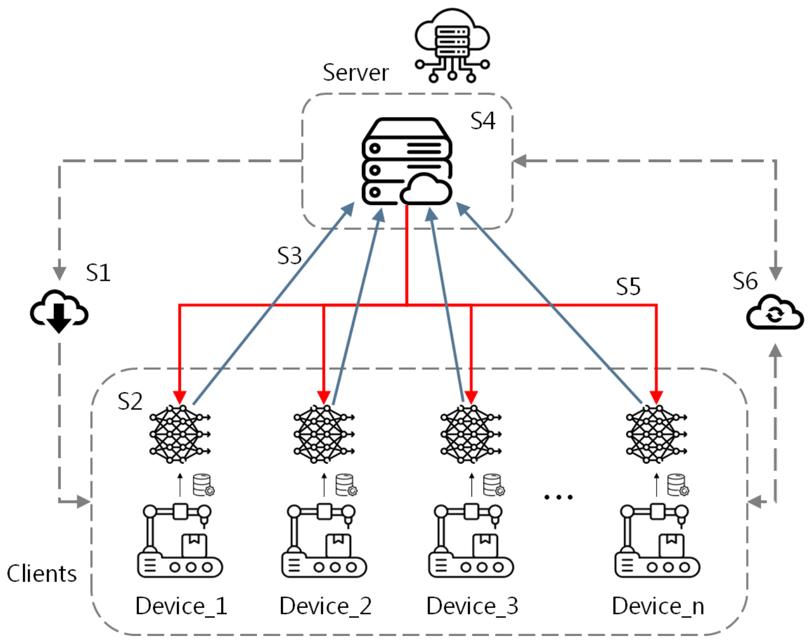

Figure 1. Research Framework Diagram

The training workflow proceeds based on FL network structure are listed below:

1. Server initializes global model
2. Clients train local models using local datasets
3. Local models are sent to server
4. Server aggregates models to build new global model
5. Updated global model is distributed to clients
6. Process repeats until convergence

3.2 Model Aggregation Strategy

We adopt an aggregation approach inspired by Shubyn, Kostrzewa et al. (2023), which favors high-performance local models. Unlike Shubyn's approach, which uses MSE for performance evaluation, we incorporate three metrics listed below to determine aggregation weight. BCE measures model proximity to true labels, while Accuracy and F1-Score evaluate classification performance, with F1-Score specifically addressing imbalanced datasets. The weighting process, shown in Equations 3-1 and 3-2, calculates each local model's impact based on its evaluation results, allowing clients with higher BCE, Accuracy, or F1-Score to have greater influence on the global model. The three metrics for determining local model influence (Inf_i): 1) Binary Cross Entropy (BCE), 2) Accuracy, and 3) F1-Score.

$$Inf_i = \begin{cases} \frac{1-\frac{ME_i}{ME_{sum}}}{n-1}, \ if\ ME\ is\ Binary\ Cross\ Entropy \\ \frac{ME_i}{ME_{sum}}, \ if\ ME\ is\ Accuracy,\ F1-Score \end{cases} \quad (3\text{-}1)$$

$$G_w = \sum_{i=1}^{n} w_i \times Inf_i \quad (3\text{-}2)$$

Each round of aggregation in the federated learning cycle is termed an iteration, with multiple rounds facilitating model convergence to the final federated model.

3.3 Predictive Model Architecture

The 1D CNN-BiLSTM model combines a 1D Convolutional Neural Network (1DCNN) (Abdeljaber, Avci et al. 2017) and a Bidirectional Long Short-Term Memory network (BiLSTM) (Zhang, Zheng et al. 2015) . This hybrid model leverages CNN's spatial feature extraction and LSTM's sequential analysis, making it well-suited for time-series data. The model, depicted in Figure 2, initially covert the input time series data as multiple time sequences. Then, 1D convolutional module consisting extracts features from sequence data. Then max pooling layer is used to reduce dimensionality and computation. The extracted features are fed into the BiLSTM layer to capture temporal dependencies, and the output passes through a fully connected layer, with softmax providing probability-based class predictions.

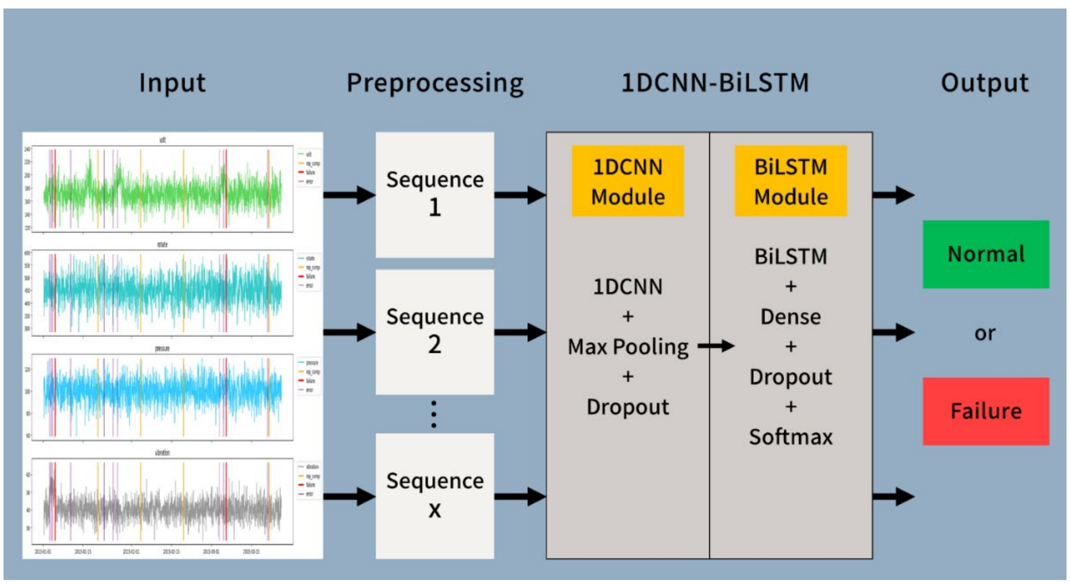

Figure 2. 1D CNN-BiLSTM model

Model parameters are summarized in Table 2. Basically, the model utilized Adam optimizer, and Binary Cross Entropy as loss function. The batch size is 128 while time step is set to be 24 based on the setting with previous study (Ahn, Lee et al. 2023). Accuracy and F1-Score were used to evaluate the model's performance.

Table 2. Model parameter settings

Layer	Parameters
Conv1D	Filter = 64, Kernel Size = 3, Stride = 1
Max Pooling	Kernel Size = 2, Stride = 1
Dropout_1	Discard Ratio = 0.2
Bidirectional_LSTM_1	Unit = 64
Bidirectional_LSTM_2	Unit = 32
Dense	Unit = 2
Dropout_2	Discard Ratio = 0.2
Dropout_3	Discard Ratio = 0.2

4. Data Collection

This research utilizes Microsoft Azure's Predictive Maintenance dataset (arnab 2020), which contains data from 100 pieces of equipment (4 different models), 876,100 total sensor readings (8,761 per device), 3,919 error logs, 3,286 maintenance records, and 761 failure records. This research conducts feature extraction based on different data sources provided by the data set, converts existing data to generate new features, and establishes a feature data set as input data for the prediction model. The types of feature data are as follows (Table 3):

Table 3. Feature Type

Feature Type	Description	Source
Sensor Data	Temperature, pressure, rotation speed	Real-time readings
Error Logs	Error frequency, type	System logs
Maintenance	Component replacements, dates	Service records
Device Info	Model type, age	Static data
Time Features	Days since last maintenance	Calculated

4.3 Data Preprocessing

Data Labeling for preprocessing are illustrated as Figure 3. The binary classification: "healthy" (0) or "faulty" (1) are labeled in the time series sequences. For the "Failure" time sequences, the fault window: current and 3 preceding timestamps before failure labeled as "faulty" based on the work in (Cardoso and Ferreira 2020).

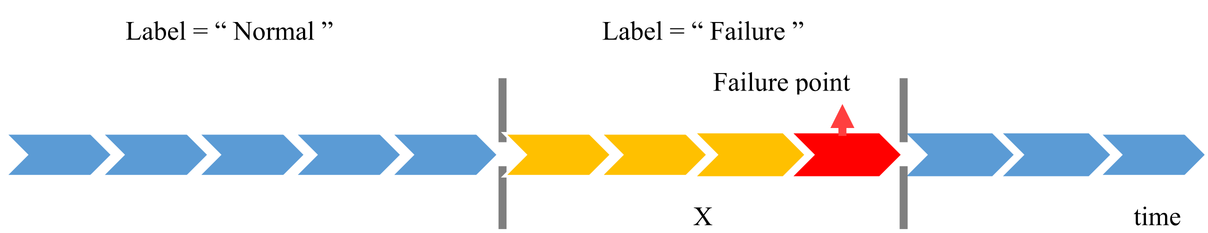

Figure 3. Label categories and construction

In the dataset, there are 876,100 samples where 761 samples are labeled as failure. To handle the class imbalance, weighted loss function, listed as Equation 4-1 by adjusting the loss function, is applied to allow the model pay more attention to categories with fewer samples during the training process.

$$class\ weight_i = \frac{N}{C \times n_i} \quad (4\text{-}1)$$

where: N = total samples; C = number of classes; n_i = samples in class i

In short, 4 preprocessing steps were conducted to prepare the input data for the model. They are 1) Timestamp alignment across data sources, 2) Feature integration and normalization, 3) Fault window labeling, and 4) Class weight calculation.

5. Results and Discussion

This research evaluates the performance of a federated learning-based predictive maintenance framework on the Microsoft Azure dataset, comparing it with individual client models that the performance was conducted by individual machine. Figure 4 illustrates the 4 corresponding research questions addressed in Introduction section. As can be seen, to answer RQ1, RQ2, and RQ3, 3 experiments were conducted to evaluate if FL model has better average performance in general, and if FL with partial training data or random partial clients can have better performance against individual, correspondingly. RQ4 aims to evaluate the performance when apply trained FL on unknown clients. In the following subsections, the results of 4 experiments are shown.

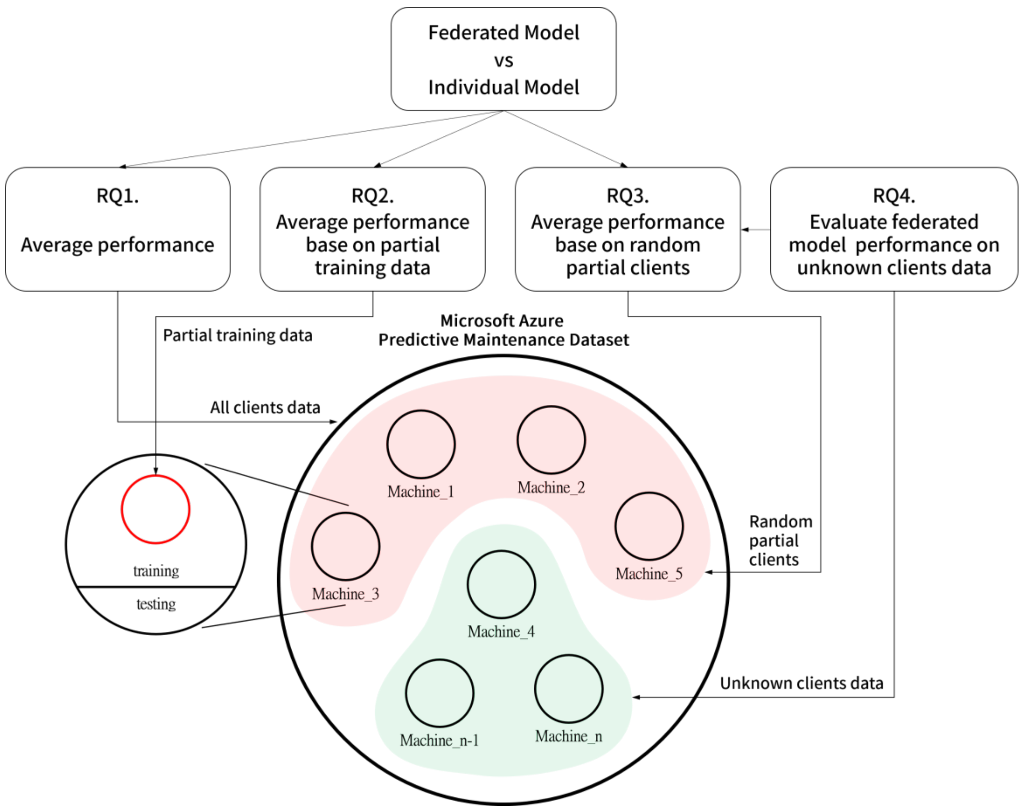

Figure 4. Experimental Overview and Research Questions

5.1 Experiment 1: Overall Predictive Performance

Table 4 show the results in terms of Accuracy, and F1-Score of FL and individual models. Obviously, the federated model, trained across 88 clients, outperformed these local individual models across all evaluation metrics. Using F1-Score as the aggregation metric, the federated model achieved 81% accuracy, a 27% improvement over individual models. Such results further confirm that FL can effectively improve the prediction performance of the model without sharing data.

Table 4. Comparison of Experiment 1 results

	Individual Model	**Federated Model**		
		aggregate base on		
		BCE	Accuracy	F1-Score
Accuracy	0.9976	0.9985	0.9986	0.9987
F1-Score	0.5456	0.7823	0.7924	0.8119

5.2 Experiment 2: Impact of Training Data Volume

Experiment 2 utilize the limited client data (25%, and 50% reduction) to evaluate if FL model can obtain the comparable performance against individual model. Obviously, shown in Table 5, when fewer client data was used for training the model, the federated model maintained high performance, with only a 10% drop in F1-Score. In contrast, local models suffered a 23% decline under the same data reduction. In short, the federated model can still maintain efficient prediction capabilities even with a small amount of training data, confirming the superiority of federated learning in situations with limited data.

Table 5. Comparison of Experiment 2 results

Partial of	**Individual Model**		**Federated Model**	
Training dataset	Accuracy	F1-Score	Accuracy	F1-Score
100%	0.9976	0.5456	0.9987	0.8119
75%	0.9957	0.4396	0.9986	0.7676
50%	0.9959	0.3108	0.9983	0.7102

5.3 Experiment 3: Effect of Client Participation

Experiment 3 evaluate the impact when the fewer clients were included in FL model training. In Table 6, the results with fewer client are listed on the right-hand side, and the left portion of results are the same as the result shown in Table as the reference. When data with fewer clients are used for training FL, reducing the number of participating clients from 88 to 22, the federated model still outperformed local individual models by 18% to 33% in accuracy with slightly losing of F1-Score. This result demonstrates the stability advantage of federated learning that is not easily affected by the number of participants.

Table 6. Comparison of Experiment 3 results

	Experiment 2				**Experiment 3**	
Partial of Training dataset	**Individual Model**		**Federated Model (88 clients)**		**Federated Model (22 clients)**	
	Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score
100%	0.9976	0.5456	0.9987	0.8119	0.9987	0.7293
75%	0.9957	0.4396	0.9986	0.7676	0.9986	0.6666
50%	0.9959	0.3108	0.9983	0.7102	0.9983	0.6435

5.4 Experiment 4: Generalization to Unseen Data

The last experiment evaluated the performance of FL when the trained FL was used to predict the failure on unknown clients. To be specific, the performance of prediction on 66 unknown clients based on the FL model trained by 22 clients is compared with the performance of that FL on individual client. In Table 7, the results with original 22 client fewer client are listed on the left-hand side, and the right portion of results show the performance of predicting failure of 66 unknown clients. Obviously, the federated model trained on 22 clients achieved strong predictive accuracy not only on the training clients but also on 66 previously unseen test clients, highlighting its excellent generalization capability.

Table 7. Comparison of Experiment 4 results

Partial of Training dataset	**Federated Model (22 clients)**		**Generalization (other 66 clients)**	
	Accuracy	F1-Score	Accuracy	F1-Score
100%	0.9978	0.7293	0.9977	0.7111
75%	0.9967	0.6666	0.9977	0.6886
50%	0.9962	0.6435	0.9977	0.6664

6. Conclusion

The application of federated learning in the smart manufacturing industry has attracted more and more attention. Predictive maintenance as an Industry 4.0 development. The core areas are easily limited by the process of accumulating data over a long period of time, making it difficult to obtain enough data for effective predictive analysis and model building in the early stages. At the same time, equipment maintenance data may contain sensitive business information, making companies reluctant to share it with other companies. The federated predictive maintenance architecture constructed in this study hopes to use the advantages of federated learning to allow multiple clients to collaboratively train models without sharing data to solve data privacy and sharing issues, and benefit from the historical experience of other clients to improve model performance.

The results validated the efficacy of the proposed federated learning framework in addressing predictive maintenance challenges. Key findings include 1) Federated learning enhances predictive accuracy by integrating insights from multiple clients. 2) The federated model maintains high performance even with limited training data per client. 3) The federated model is resilient to reductions in client participation during training. 4) The federated model demonstrates strong generalization to new, unseen data. Based on these findings, the framework's ability to balance data privacy, efficiency, and model performance underscores its suitability for real-world smart manufacturing deployments.

In future work, federated predictive maintenance will not only require continued improvements in predictive model performance. It is necessary to consider the communication costs caused by information exchange during the aggregation process of the model, and reduce unnecessary information transmission to improve communication efficiency. In addition, the collaborative process of federated learning is vulnerable to attacks by malicious participants or tampering with information, leading to system paralysis or errors. Therefore, the implementation of strong security protection mechanisms such as encryption technology is also one of the important tasks in the future. Through these efforts, the practical application value of federated learning can be further enhanced and ensure its widespread application in the field of smart manufacturing.

References

Abdeljaber, O., et al., Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks, *Journal of sound and vibration,* vol.388, pp. 154-170, 2017.

Ahn, J., et al., Federated learning for predictive maintenance and anomaly detection using time series data distribution shifts in manufacturing processes, *Sensors,* vol.23, no.17, pp.7331, 2023.

Arnab, Microsoft Azure Predictive Maintenance, *Kaggle*, 2020

Budach, L., et al., The effects of data quality on machine learning performance, *arXiv preprint arXiv:2207.14529*, 2022

Cardoso, D. and L. Ferreira, Application of predictive maintenance concepts using artificial intelligence tools, *Applied Sciences,* vol.11, no.1, pp.18, 2020

Ge, N., et al., Failure prediction in production line based on federated learning: an empirical study, *Journal of Intelligent Manufacturing,* vol.33, no.8, pp.2277-2294, 2022

Hashemian, H. M., State-of-the-art predictive maintenance techniques, *IEEE Transactions on Instrumentation and measurement,* vol.60, no.1, pp.226-236, 2010

Liu, Y., et al., Deep anomaly detection for time-series data in industrial IoT: A communication-efficient on-device federated learning approach, *IEEE Internet of Things Journal*, vol.8, no.8, pp.6348-6358, 2020

Shubyn, B., et al., Federated learning for improved prediction of failures in autonomous guided vehicles, *Journal of Computational Science,* vol.68, no.101956, 2023

Wan, J., et al., Software-defined industrial internet of things in the context of industry 4.0, *IEEE Sensors Journal,* vol.16, no.20, pp.7373-7380, 2016

Weber, Ingo, et al., Untrusted business process monitoring and execution using blockchain, *Business Process Management: 14th International Conference, BPM 2016, Rio de Janeiro, Brazil, September 18-22, 2016. Proceedings 14*, Springer International Publishing, pp. 329-347, 2016

Zhang, Shu, et al., Bidirectional long short-term memory networks for relation classification. In: *Proceedings of the 29th Pacific Asia conference on language, information and computation*, pp. 73-78*,* 2015.

Biographies

Chao-Lung Yang received a B.S. degree in Mechanical Engineering and an M.S. degree in Automation Control from National Taiwan University of Science and Technology, Taiwan, in 1996 and 1998, respectively. He also received the M.S.I.E. and Ph.D. degree in Industrial Engineering from Purdue University, West Lafayette, in 2004 and 2009, respectively. He is currently with the Department of Industrial Management, National Taiwan University of Science and Technology, Taiwan, as a professor. His research area is data mining, machine learning, big data analytics, metaheuristic algorithm, and human action recognition. He has developed a data streaming analytics framework by applying deep learning models such as CNN and LSTM, and metaheuristic algorithms to quickly detect the process shift and classify the product detects. Recently, he works in the computer vision domain to develop machine learning models to recognize human action, particularly in manufacturing operations. He is a member of INFORMS, IEEE, and ASME.

Hwai-Chen Qiu is currently pursuing his PhD at the Graduate Institute of Intelligent Manufacturing Technology at National Taiwan University of Science and Technology (NTUST). He holds a Master’s degree from the Department of Electrical Engineering at National Cheng Kung University. He is also a manager at Hon Hai Precision Industry Co., Ltd. (Foxconn). His research interests include artificial intelligence, digital transformation, digital twins, and AI agents.

Cheng-Yeh Yang received his master’s degree in Industrial Management from National Taiwan University of Science and Technology (NTUST) in 2024. His research interests include Industry 4.0, predictive maintenance, and artificial intelligence.