Guest Editorial
Special Issue on Learning From Imperfect Data for Industrial Automation

WITH the rapid development of advanced sensing, communication, and the industrial Internet of Things, it has become much easier to obtain, transmit, and, store a massive amount of real-world data. However, imperfect data is inevitable in real-world systems, such as the existence of outliers, contaminated, incomplete, inaccurate, and even missing information in the data. This phenomenon is called data imperfection, which usually makes traditional data-driven modeling and automation methods either unfeasible or ending at undesired inaccuracies. This has been a well-known challenge to data-driven methods when applied to real-world systems, such as process industry, manufacturing, energy networks, and transportation systems.

The central theme of this Special Issue is on challenges and responses in automation science and engineering for learning from imperfect data, with the purpose of providing a forum to share the state of art, new theories, and methods in dealing with imperfect data for industrial automation, as well as the achievements and lessons learnt. We received 47 submissions (excluding several invalid submissions) that cover a broad range of topics relevant to automation science and engineering for learning from imperfect data. Through a rigorous peer-review process, 18 articles have been accepted, which assemble the latest cross-industry and multidisciplinary research in academic and industrial communities on automation science, data science, artificial intelligence, and control engineering. The contributions in this Special Issue can therefore be divided into the following three categories.

1) Learning from imperfect data for modeling and optimization.
2) Learning from imperfect data for control.
3) Learning from imperfect data for process monitoring and fault diagnosis.

Specifically, the main contributions of the accepted articles, listed in the appendix, are highlighted in the following.

A. Learning From Imperfect Data for Modeling and Optimization

In [A1], a modeling and operation optimization strategy based on feedstock property and production features is presented to enhance the production performance of a distillation unit. One of the challenges is how features can be uncovered from high-dimensional and imperfect data, where imperfect data refers to product quality data that is unavailable online. In the proposed strategy, the authors inject the inherent characteristic of the process into the data-driven method to extract the feedstock property in a data-based and knowledge-oriented manner. Similarly, the work in [A2] develops a multiline rescheduling framework for data-driven surrogate modelling and networkwise optimal rescheduling of multiple lines by overcoming the difficulties of data incompleteness, imbalance, and lack of comprehensiveness.

To mitigate the impact of prediction errors on downstream production decisions, Gong et al. [A3] propose a prediction model that minimizes the decision error. Meanwhile, to enhance the adaptability to imperfect data, the prediction model is extended to a distributionally robust version, which takes into account of the worst-case formulation in the feature space. Likewise, the work in [A4] examines the worst-case scenario. The authors investigate the recovery of blackout missing data, which is essentially the interpolation of industrial time series data. A hierarchical imputation framework is developed to recover missing values under different operating conditions for incomplete datasets.

In order to realize accurate modeling of industrial process with nonstationary characteristics, two different algorithms are presented in [A5] and [A6], respectively. The work in [A5] proposes an interval type-2 fuzzy neural network based on active semi-supervised learning, which can actively identify the occurrence moments of concept drift and learn from samples with partial labels. In [A6], Wen et al. propose a novel online sequential sparse robust neural networks with random weights for imperfect industrial streaming data to achieve highly reliable online modeling of time-variant dynamic systems. This work considers the widespread outlier problem in the input and output data and thus enhances the robustness of the model by Schweppe generalized M-estimation.

B. Learning From Imperfect Data for Control

For the sintering process with various imperfect data, An et al. [A7] propose an intelligent control strategy of ignition temperature based on working-condition recognition to stabilize the ignition temperature. This work uses clustering algorithm to identify the different working conditions caused by the large variation and fluctuation of gas calorific value, which avoids the influence of the fluctuation and instability of the imperfect data in the system. The challenge of

Digital Object Identifier 10.1109/TASE.2024.3374132

obstacle avoidance of prostheses is that the recognition of environments by the human–robot interaction between amputees and prostheses is not accurate enough because of the presence of noise and individual differences. Hence, the work in [A8] develops a vision-locomotion coordination control method with imperfect data learning to help the powered lower limb prosthesis fulfill several obstacle avoidance tasks.

In addition, there are two works on data-driven robust control. Ding et al. [A9] develop an event-triggered online learning fuzzy-neural robust controller which is applied to the furnace temperature control of municipal solid waste incineration process with outliers and noise. The outliers are eliminated by the box-plot method, and the data are denoised with a Gaussian filter. The robust optimal control problem is investigated in [A10] based on data-driven robust policy iteration. This work allows for dynamic uncertainties resulted from the interconnected dynamic system, as well as unknown bounded disturbances throughout the learning process.

C. Learning From Imperfect Data for Process Monitoring and Fault Diagnosis

Aiming at the problem of less labeled and multisource heterogeneous data of plate shape quality, a self-supervised learning framework based on multisource heterogeneous contrast learning is developed in [A11]. The self-supervised feature learning phase is responsible for training the encoder with massive unlabeled data. A classifier followed by the encoder is built with less labeled data in the supervised fine-tuning phase. In [A12], Quan et al. use deep dilated causal convolutional neural network combined with logistic regression to detect and predict the stall inception using the imperfect time-series data of axial compressors with the rotating stall. The method can detect the irregular and imperfect stall inception and capture the small fault features when the characteristics of stall inception signals or data are not unobvious, irregular, and imperfect. Similarly, a fuzzy C-means-based algorithm is presented in [A13] for outlier detection of the real-world industrial data collected from a wire arc additive manufacturing pilot line.

Wind turbines (WTs) usually work in harsh environments and complex operating conditions, which easily leads to nonstationary, randomness, and multiple outliers in operating data. Aiming at solving this problem, Zhang et al. [A14] propose a fault diagnosis method for WT generators based on multitask learning. In this work, the imperfect operating data are divided into different conditions and each condition is constructed as a task. Through parameter sharing between different tasks, it overcomes the deficiency of data and realizes fault diagnosis under complex working conditions. Likewise, Lu et al. [A15] address the fault diagnosis of an offshore WT. As the collection and transmission of offshore WT data is severely restricted, the work in [A15] adopts the event-triggered federated learning to maintain high diagnostic performance while reducing communication costs.

In [A16], Wang et al. develop a well-behaved superheat degree identification model. To deal with labeled data scarcity and high annotation costs, a contrastive anchors-based-label propagation algorithm is used to predict the pseudo-labels and adopt a variational information domain adaption module and mini-batch incremental learning strategy to enhance the accuracy of model identification. In [A17], Vantilborgh et al. correlate the expert labelling with measurement data to construct a generic prognostic tool for probabilistic condition monitoring of mechatronic systems. Likewise, Xiao et al. [A18] report a work on the power consumption monitoring with imperfect data. In this work, a prediction-based approach using an interpretable data-driven model is developed and achieves satisfactory monitoring accuracy.

As guest editors, we hope that the collection of the papers in this Special Issue can reflect the current works in the area and thus promote the research information exchanges of the relevant research communities.

PING ZHOU, *Lead Guest Editor*
State Key Laboratory of Synthetical Automation for Process Industries
Northeastern University
Shenyang 110819, China
e-mail: zhouping@mail.neu.edu.cn

XUEWU DAI, *Guest Editor*
Department of Mathematics, Physics and Electrical Engineering
Northumbria University
NE1 8ST Newcastle upon Tyne, U.K.
e-mail: xuewu.dai@northumbria.ac.uk

KYRIAKOS G. VAMVOUDAKIS, *Guest Editor*
Daniel Guggenheim School of Aerospace Engineering
Georgia Institute of Technology
Atlanta, GA 30332 USA
e-mail: kyriakos@gatech.edu

JAN FAIGL, *Guest Editor*
Faculty of Electrical Engineering
Czech Technical University
166 36 Prague, Czech Republic
e-mail: faiglj@fel.cvut.cz

HONG WANG, *Guest Editor*
Energy and Transportation Science Division
Oak Ridge National Laboratory
Oak Ridge, TN 37831 USA
Department of Electrical and Electronics Engineering
The University of Manchester
M13 9PL Manchester, U.K.
e-mail: hong.wang@manchester.ac.uk

APPENDIX: RELATED ARTICLES

[A1] S. Li, Y. Zheng, S. Li, and M. Huang, "Data-driven modeling and operation optimization with inherent feature extraction for complex industrial processes," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1092–1106, Mar. 2024.

[A2] R. Liu, D. Cui, X. Dai, P. Yue, and Z. Yuan, "A data-driven surrogate modeling for train rescheduling in high-speed railway networks under wind-caused speed restrictions," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1107–1121, Mar. 2024.

[A3] H. Gong, Y. Zhang, and Z.-H. Zhang, "Training demand prediction models by decision error for two-stage lot-sizing problems," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1122–1137, Mar. 2024.

[A4] D. Liu, Y. Wang, C. Liu, K. Wang, X. Yuan, and C. Yang, "Blackout missing data recovery in industrial time series based on masked-former hierarchical imputation framework," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1138–1150, Mar. 2024.

[A5] J. Qiao, Z. Sun, and X. Meng, "Interval type-2 fuzzy neural network based on active semi-supervised learning for non-stationary industrial processes," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1151–1162, Mar. 2024.

[A6] C. Wen, P. Zhou, W. Dai, L. Dong, and T. Chai, "Online sequential sparse robust neural networks with random weights for imperfect industrial streaming data modeling," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1163–1175, Mar. 2024.

[A7] J. An, C. Yang, M. Wu, and S. Du, "Intelligent control strategy for sintering ignition temperature based on working-condition recognition," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1176–1187, Mar. 2024.

[A8] Z. Hong, S. Bian, P. Xiong, and Z. Li, "Vision-locomotion coordination control for a powered lower-limb prosthesis using fuzzy-based dynamic movement primitives," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1188–1200, Mar. 2024.

[A9] H. Ding, J. Qiao, W. Huang, and T. Yu, "Event-triggered online learning fuzzy-neural robust control for furnace temperature in municipal solid waste incineration process," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1201–1213, Mar. 2024.

[A10] O. Qasem and W. Gao, "Robust policy iteration of uncertain interconnected systems with imperfect data," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1214–1222, Mar. 2024.

[A11] D. Li, J. Lu, T. Zhang, and J. Ding, "Self-supervised learning and multisource heterogeneous information fusion based quality anomaly detection for heavy-plate shape," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1223–1234, Mar. 2024.

[A12] F. Quan, X. Sun, H. Zhao, Y. Li, and G. Qin, "Detection of rotating stall inception of axial compressors based on deep dilated causal convolutional neural networks," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1235–1243, Mar. 2024.

[A13] J. Fang et al., "A new particle swarm optimization algorithm for outlier detection: Industrial data clustering in wire arc additive manufacturing," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1244–1257, Mar. 2024.

[A14] Y. Zhang, L. Qiao, and M. Zhao, "Fault diagnosis for wind turbine generators using normal behavior model based on multi-task learning," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1258–1270, Mar. 2024.

[A15] S. Lu, Z. Gao, P. Zhang, Q. Xu, T. Xie, and A. Zhang, "Event-triggered federated learning for fault diagnosis of offshore wind turbines with decentralized data," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1271–1283, Mar. 2024.

[A16] J. Wang, S. Xie, Y. Xie, and X. Chen, "Label propagation with contrastive anchors for deep semi-supervised superheat degree identification in aluminum electrolysis process," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1284–1296, Mar. 2024.

[A17] V. Vantilborgh, T. Lefebvre, K. Eryilmaz, and G. Crevecoeur, "Data-driven virtual sensing for probabilistic condition monitoring of solenoid valves," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1297–1311, Mar. 2024.

[A18] H. Xiao, W. Hu, H. Zhou, and G.-P. Liu, "Prediction-based power consumption monitoring of industrial equipment using interpretable data-driven models," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1312–1322, Mar. 2024.

Ping Zhou (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in control theory and engineering from Northeastern University in 2003, 2006, and 2013, respectively. He has been with Northeastern University since 2010. He is currently a Professor with the State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University. His research interests include data-driven modeling, monitoring, and control with various industrial applications. He has authored three monographs and more than 100 refereed journal publications in these areas. He won the First Prize of Zhejiang Provincial Science and Technology Award in 2023, the First Prize of the Natural Science Award of the Chinese Association of Automation in 2021, and the First Prize and Second Prize of the Natural Science Award of the Ministry of Education of China in 2022 and 2020, respectively. He was a recipient of the Top Young Talents Award of the National Ten Thousand Talents Plan of China in 2019.

Xuewu Dai (Member, IEEE) received the B.Eng. degree in electronic engineering and the M.Sc. degree in computer science from Southwest University, Chongqing, China, in 1999 and 2003, respectively, and the Ph.D. degree in electrical and electronic engineering from The University of Manchester, Manchester, U.K., in 2008. He is currently a Senior Lecturer with the Department of Mathematics, Physics and Electrical Engineering, Northumbria University, Newcastle upon Tyne, U.K. Prior to that, he was a Post-Doctoral Researcher with the Department of Engineering Science, University of Oxford, and the Department of Electronic and Electrical Engineering, University College London, U.K. His main research interests include robust adaptive signal processing, state estimation, intelligent transportation systems, automatic train control, and rescheduling. He was awarded the Early Career Research Prize by SWIG U.K.

Kyriakos G. Vamvoudakis (Senior Member, IEEE) received the Diploma degree in electronic and computer engineering from the Technical University of Crete, Greece, in 2006, and the M.Sc. and Ph.D. degrees in electrical engineering from The University of Texas at Arlington in 2008 and 2011, respectively. He is currently the Dutton-Ducoffe Endowed Professor with the Daniel Guggenheim School of Aerospace Engineering, Georgia Tech. His research interests include reinforcement learning, control theory, cyber-physical security, bounded rationality, and safe/assured autonomy. He is a member of the IEEE Control Systems Society Conference Editorial Board. He was a recipient of the 2019 ARO YIP Award, the 2018 NSF CAREER Award, the 2018 DoD Minerva Research Initiative Award, and the 2021 GT Chapter Sigma Xi Young Faculty Award. His work has been recognized with best paper nominations and several international awards, including the 2016 International Neural Network Society Young Investigator Award, the Best Paper Award for Autonomous/Unmanned Vehicles at the 27th Army Science Conference in 2010, and the Best Researcher Award from the Automation and Robotics Research Institute in 2011. He is an Associate Editor of *Automatica*, IEEE Transactions on Automatic Control, IEEE Transactions on Systems, Man, and Cybernetics: Systems, IEEE Transactions on Artificial Intelligence, *Neurocomputing*, *Journal of Optimization Theory and Applications*, and IEEE Transactions on Neural Networks and Learning Systems.

Jan Faigl (Member, IEEE) received the Ing. degree in cybernetics and the Ph.D. degree in artificial intelligence and biocybernetics from Czech Technical University in Prague (CTU), Prague, Czech Republic, in 2003 and 2010, respectively. From 2013 to 2014, he was a Fulbright Visit Scholar with the University of Southern California, Los Angeles, CA, USA. Since 2013, he has been leading the Computational Robotics Laboratory, Artificial Intelligence Center. He is currently a Professor of computer science with the Faculty of Electrical Engineering, CTU. He is also the Co-Founder of the Center for Robotics and Autonomous Systems. His current research interests include combinatorial motion planning, robotic information gathering, long-term autonomous missions with online incremental learning, autonomous navigation, aerial systems, and multilegged locomotion control. He received the Antonin Svoboda Award from the Czech Society for Cybernetics and Informatics in 2011. He served as the Guest Editor of the Special Issue on Online Decision Making in Multi-Robot Coordination for *Autonomous Robots*. He serves as an Associate Editor for IEEE Transactions on Automation Science and Engineering.

Hong Wang (Fellow, IEEE) received the B.Sc. degree from Huainan Institute of Mining Engineering in 1982 and the M.Sc. and Ph.D. degrees from Huazhong University of Science and Technology, China, in 1984 and 1987, respectively. He was a Post-Doctoral Research Fellow with the University of Salford, Brunel University, and the University of Southampton, before joining the University of Manchester Institute of Science and Technology (UMIST), U.K., in 1992. He was a Chair Professor in process control from 2002 to 2016, where he was the Deputy Head of the Paper Science Department and the Director of the UMIST Control Systems Centre (which was established in 1966 and is the birthplace of Modern Control Theory) from 2004 to 2007. He was a member of the University Senate and General Assembly and has been a Professor Emeritus of The University of Manchester since 2016 (https://research.manchester.ac.uk/en/persons/hong-wang). From 2016 to 2018, he was a Laboratory Fellow with the Pacific Northwest National Laboratory and the Chief Scientist and was the Co-Leader of the Control of Complex Systems (https://controls.pnnl.gov/). He joined the Oak Ridge National Laboratory, as a Senior Distinguished Scientist, in January 2019. His research interests include stochastic distribution control, fault diagnosis and tolerant control, and intelligent controls with applications to transportation systems. He is a member of three IFAC committees. He was an Associate Editor of IEEE Transactions on Automatic Control, IEEE Transactions on Control System Technology, and IEEE Transactions on Automation Science and Engineering. He is an Associate Editor of IEEE Transactions on Neural Networks and Learning Systems.