ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Efficient and Resilient Machine Learning for Industrial Applications

Philipp Wissmann[1,2], Philip Naumann[3], Daniel Hein[1], Steffen Udluft[1], Marc Weber[1], Simon Leszek[3], and Thomas Runkler[1,2] *

[1]Siemens AG, Munich, Germany
[2]TU Munich (TUM), Munich, Germany
[3]Technische Universität Berlin, Machine Learning Group, Berlin, Germany

Abstract. Machine learning is rapidly transforming industrial landscapes, yet it faces significant hurdles related to efficiency and resilience. This paper discusses industrial challenges and provides a structured overview of current approaches, encompassing data-centric methodologies, efficient training for reliable solutions, hardware-optimized deployment, and the emerging role of foundation models.

1 Introduction to industrial challenges

Machine learning (ML) and artificial intelligence (AI) are transforming industries by automating complex processes, augmenting human decision-making, and unlocking unprecedented efficiency and precision. These technologies enable businesses to tap into new opportunities: In manufacturing, technologies such as predictive maintenance and process optimization have enhanced quality control and reliability in workflows. In logistics, AI-driven route optimization, demand forecasting, and fleet maintenance revolutionize supply chain operations. Meanwhile, retail has leveraged ML for inventory management and personalized customer experiences, where forecasting models balance supply and demand with remarkable accuracy. ML algorithms often outperform traditional, rule-based methods by learning complex patterns directly from observed data. In practice, however, they face significant challenges, including issues of efficiency, reliability, and transparency. Hence, ensuring reliability in highly dynamic conditions while maintaining and improving scalability and adaptability remains a persistent focus of research – a mission that continues to drive innovation for efficiency and resilience.

Amidst this, foundation models have emerged, hinting at a new era where scalability and flexibility are integral to meeting real-world demands. They promise to improve both downstream efficiency and scalability. Nonetheless, they still face the same stringent requirements, which are often not met in practice.

In this work, we aim to provide a comprehensive overview of the challenges of ML systems in industrial applications and how they can be addressed.

*This work was funded by the German Federal Ministry for Research, Technology and Space through the project RIESIQ under project numbers 16IS24087A and 16IS24087C.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

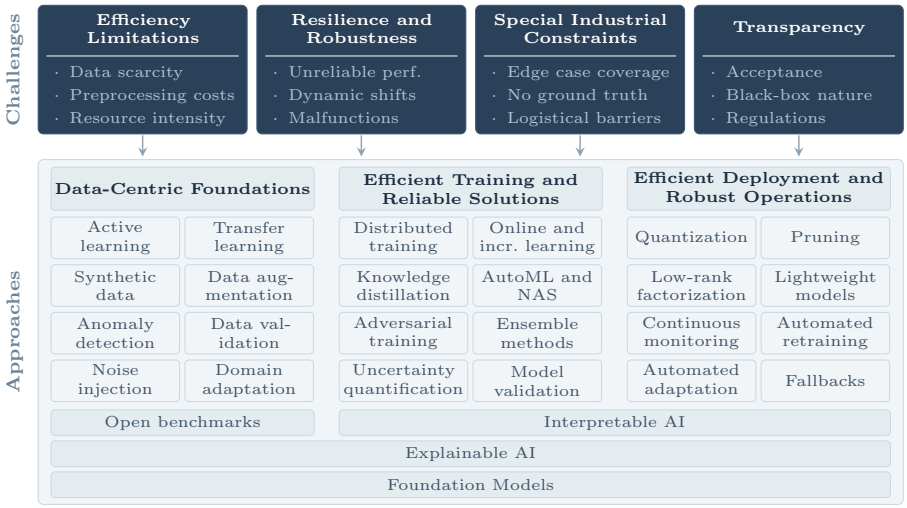

Fig. 1: High-level view of industrial challenges and practical solution ways.

1.1 Efficiency

While efficiency in industrial applications often concerns the processes themselves, we focus on the computational cost of creating, running, and maintaining reliable ML systems in industrial environments. Here, we identify challenges in three key areas. Firstly, **data efficiency** addresses the availability of high-quality data and labels, which are often scarce or expensive to obtain in industrial environments. Furthermore, data regulations can limit accessibility. Efficiency also extends to data preprocessing, including data quality assurance and feature engineering. **Training efficiency** refers to the computational costs of model training, which can be significant or even prohibitive, depending on the amount of data, the complexity of the learning problem, and the model being trained. Lastly, **deployment and operational efficiency** are critical in industrial contexts, where models often must run reliably on limited hardware with stringent latency requirements.

1.2 Resilience and robustness

Robustness refers to the ability of a system to resist or avoid disruptions, while resilience emphasizes the ability to adapt to and recover from those. In ML, examples of such disruptions include noisy or altered data, adversarial attacks, or generalization to novel data, whereas in engineering, component failures, environmental changes, or unexpected loads are common. These terms are often used interchangeably in industrial settings and are crucial for ensuring reliable performance under any kind of uncertainty, as their absence can lead

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

to production downtime, increased operational costs, or safety risks. Effectively addressing them, therefore, requires managing challenges across three key areas: **data robustness**, ensuring reliability against changes or inconsistencies in the data pipeline; **training robustness**, maintaining reliable model performance; and **deployment and operational robustness**, ensuring models operate consistently on varying hardware, changing environments, and input conditions.

1.3 Special constraints in industrial applications

Developing an ML solution is only the first step; ensuring its quality and suitability for complex, real-world environments requires addressing additional challenges in data collection, evaluation, and deployment. Often, there is an **incomplete coverage of edge cases**. Complex, rare, yet critical edge cases are poorly represented in the training and test datasets. Models optimized for standard scenarios may fail when encountering these unanticipated cases, leading to unreliable predictions. Furthermore, test **datasets inherently reflect past performance**, and cannot fully anticipate future data distributions caused by process changes, equipment upgrades, or environmental variations. Known as concept drift or data drift, these discrepancies complicate model evaluation and selection, which particularly impacting control optimization tasks. For instance, policies trained with reinforcement learning (RL) may navigate unexplored regions of the state space, increasing operational risk (e.g., offline RL [1]). Even seemingly well-performing algorithms can become unsuitable, either by rendering tasks ill-posed or by their excessive sensitivity to hyperparameters and implementation details [2, 3]. **Gathering comprehensive test data** across all relevant operating conditions (e.g., varying load conditions, ambient temperatures, or wear states) is highly challenging. The logistical costs, operational disruptions, or impracticalities associated with replicating such conditions can severely limit datasets, reducing their exhaustiveness and quality for robust model evaluation. Additionally, for tasks like anomaly detection and predictive maintenance, there is often a **lack of ground truth for rare but critical events**, such as equipment failures. This complicates testing and model validation. To address these constraints, adaptive and scalable pipelines are essential, enabling ML systems to evolve over time while accommodating industry-specific challenges.

1.4 Transparency and acceptance

A major contributing factor to the employment of ML methods in industrial applications is building trust in the model's learned decision strategy. **Transparency**, i.e. the ability to comprehend how models make decisions, is a critical factor influencing their acceptance. While some algorithms are inherently more interpretable (e.g., linear models or decision trees), others learn complex functions that are essentially "black-boxes" (e.g., deep neural networks), making it difficult for users and domain experts to trust their reliability or understand potential

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

failure modes. A model may deceptively perform correctly, but possibly for the wrong reason, known as the "Clever Hans" effect [4, 5]. Unfortunately, the complexity of the learned decision function often correlates with its downstream performance [6]. Solving difficult problems, such as learning from heterogeneous data sources, is essential for building robust models, yet it often requires utilization of less transparent models. This lack of interpretability is particularly pronounced in industrial contexts, where relying on black-box mechanisms for critical, potentially safety-related tasks can introduce risks. Complex real-world tasks are rarely fully testable – enumerating all possible input-output scenarios or identifying all undesirable outcomes is practically infeasible. Additionally, legal frameworks, such as the European Union's GDPR, mandate interpretability by requiring transparency in automated decision-making systems.

2 Current approaches in industrial applications

We now transition to exploring practical approaches. Instead of aiming for an exhaustive survey, we provide a structured overview of key paradigms that we believe are central to addressing these challenges.

2.1 Data-centric foundations for industrial machine learning

The following methodologies aim at improving the availability, quality, usability, and reliability of industrial datasets.

Benchmarks: Open, domain-specific benchmarks are essential to foster innovation and establish a shared understanding of critical problems and perspectives in industrial ML. However, the scarcity of accessible datasets that reflect industrial complexity remains a barrier. Tasks such as distinguishing healthy from faulty manufactured components often require approaches that differ from those most open datasets represent. Therefore, industry-motivated benchmarks [7, 8] address this vital gap and provide resources tailored to real-world industrial scenarios.

Efficient data strategies refine and modify data at its core to extract maximum value from limited existing data.

Active learning: This approach selects the most informative and uncertain data points for manual annotation [9], significantly reducing the need for large labeled datasets and lowering annotation efforts.

Transfer learning and pre-trained models: Pre-trained models fine-tuned for specific industrial tasks reduce the need for domain-specific data and training time. Transfer learning and generalization techniques enhance the applicability of, e.g., RL systems with limited real-world interactions [10].

Synthetic data generation: Where real data is scarce or difficult to obtain (e.g., rare errors in production), synthetic data generation can expand datasets and improve model robustness without incurring high costs for real data collection [11].

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Data augmentation: By augmenting existing data (e.g., changing image brightness and rotation [12], or targets for an adaptive RL policy [13]), the size and diversity of datasets can be artificially increased to improve model generalization and reduce the need for data collection.

With industrial data often being noisy, incomplete, or subject to drift, **robust data strategies** are crucial to ensure consistent model performance.

Anomaly detection (AD) and multimodel data integration: AD serves as a critical tool for monitoring industrial processes and identifying deviations caused by noise, outliers, or system inconsistencies [14]. Modern approaches use deep neural networks [15, 16, 17] to identify intricate patterns and distinguish between anomalies such as noise and semantic inconsistencies. For instance, unsupervised methods can support this by clustering the noise [18]. In time-series applications, methods such as subsequence AD [19] enable precise differentiation of irregularities across temporal datasets. Multimodal data integration [20, 21] enables ML systems to fuse and process diverse data sources, thereby enhancing AD accuracy and overall system robustness.

Data validation and cleansing: Robust pipelines for continuous validation and cleansing of input data help detect and correct anomalies, outliers, and errors early on before they impact model performance [22].

Noise injection: By adding realistic noise, disturbances, or variations to the training data, models can learn to deal with such uncertainties and improve their generalization ability [23]. This is particularly useful for increasing robustness against sensor failures or measurement errors.

Domain adaptation: This approach aims to mitigate performance drops caused by shifts in data distribution between training and deployment environments [24], ensuring models remain efficient even in slightly altered scenarios.

2.2 From efficient training to reliable machine learning solutions

Creating and training ML models, especially complex ones, can be computationally intensive, time-consuming, and costly. **Efficiency in model generation** aims to optimize the training process, reducing resource usage, cutting costs, and shortening development cycles.

Efficient training strategies: Distributing training tasks across multiple machines or accelerators significantly reduces training time for large datasets and complex models [25]. Leveraging distributed computing optimizes resource utilization, making training more scalable and efficient.

Online and incremental learning: Instead of retraining models from scratch, these approaches enable continuous adaptation to new data streams [26]. This reduces the need for costly full retraining cycles and enables real-time adaptation, spreading the training effort over time.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Knowledge distillation: By training a smaller, more efficient "student" model to replicate the behavior of a larger, complex "teacher" model [27], knowledge distillation enables comparable performance with fewer parameters and much lower computational costs.

Automated machine learning (AutoML) and neural architecture search (NAS): AutoML and NAS [28] automate the time-consuming tasks of model optimization, such as searching for efficient architectures or tuning hyperparameters. These methods can drastically reduce human efforts and iteration times. However, the computational cost of these searches can be substantial, requiring efficient prioritization and resource allocation to maximize benefits.

The design and training of ML models themselves play a central role in their inherent **resilience** and **robustness** to various disruptions.

Adversarial training: Training models with adversarial examples, i.e. subtle input manipulations that are almost imperceptible to humans, helps improve robustness against targeted attacks and unexpected data variations [29].

Ensemble methods: Combining multiple models (e.g., random forests, boosting, or stacking) can significantly improve the overall performance and robustness by compensating for errors or weaknesses in individual models [30].

Uncertainty quantification: Models that not only make a prediction but can also quantify the uncertainty of that prediction are valuable in industrial applications [31]. They make it possible to identify uncertain decisions and trigger human intervention or alternative strategies.

Model validation and verification: Rigorous validation and verification processes, including stress testing under extreme or unusual conditions, are essential to ensure robustness prior to deployment.

Explainable AI (XAI): XAI aims to make models more understandable to humans, e.g., by providing post-hoc explanations for opaque black-box models, thereby contributing to transparency and trustworthy AI [32]. Post-hoc XAI methods enable practitioners to audit trained models, verify that their behavior aligns with expectations, and identify issues such as relying on spurious features [4]. Beyond individual inspection, such techniques have proven effective in industry-related tasks [5, 33, 34] and can support the mitigation of unintended shortcuts or biases learned by the model without requiring full retraining [35]. At the data level, XAI methods can further be used to aggregate explanations across entire datasets and attribute distributional shifts to specific features or subgroups [36, 37], thereby providing a principled foundation for informed data curation and improved model robustness.

Interpretable AI: Interpretability refers to the extent to which a model's internal mechanisms can be directly understood by humans, and can be incorporated as an explicit design principle [38, 39, 40]. Based on this, interpretable AI approaches evaluate interpretability by selecting a quantifiable proxy, e.g., a certain model class like fuzzy rules, algebraic equations, or high-level task

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

descriptions. Algorithms are then developed to optimize within this chosen class, thereby claiming interpretability [41, 42].

2.3 Hardware-efficient deployment and robust operations

Industrial applications often require ML models to run directly on edge devices or in embedded systems, where computing power, memory, and energy consumption are severely limited. **Efficiency on target hardware** focuses on designing and optimizing models to function optimally under these constraints and make the best possible use of the available computing resources.

Quantization: Representing model parameters with lower-precision data types (e.g., 8-bit integers instead of 32-bit floats) drastically reduces memory usage and accelerates inference on hardware optimized for lower-precision arithmetic [43].

Pruning: By removing irrelevant or less influential connections and neurons in a neural network, pruning produces more compact and efficient models [44].

Low-rank factorization: Decomposing weight matrices into low-rank approximations can reduce the number of parameters and computational operations, especially in fully connected layers.

Lightweight model architectures: Architectures specifically designed for efficiency use techniques such as depthwise separable convolutions and grouped convolutions [45] to reduce computational complexity while maintaining accuracy. These are ideal for edge AI and TinyML applications [46].

Beyond the individual model, the entire ML system must be designed to **operate resiliently** under dynamic industrial conditions.

Continuous monitoring: Real-time monitoring of model performance, data quality, and system integrity is crucial [47]. Early warning systems detect performance degradation and anomalies, enabling timely interventions.

Automated retraining and adaptation: When a decline in performance or significant data changes are detected, automated retraining or fine-tuning of the model can be triggered. Online learning allows continuous updating, while transfer learning supports adaptation even in offline settings [48].

Fallbacks and fault tolerance: Fallback mechanisms, such as switching to simpler models, manual overrides, or predefined rules, ensure reliability when models fail or are uncertain. Redundant systems can increase fault tolerance.

2.4 A new paradigm: Foundation models

Foundation models are general-purpose architectures trained on massive datasets, enabling adaptability across diverse tasks with minimal fine-tuning. This paradigm shift allows for broad applicability, robustness, and scalability, making them particularly promising for industries. Among the key types, large language

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

models (LLMs) and time-series foundation models (TSFMs) currently stand out.

TSFMs (e.g., [49, 50]) excel in tasks like time-series simulation, forecasting, and AD by learning robust temporal patterns across domains. Their ability to facilitate transfer learning with minimal labeled data and operate in zero-shot settings significantly lowers barriers for new industrial applications. However, how their performance compares with traditional AI remains an area of ongoing research [51]. As LLMs demonstrate capabilities beyond natural language understanding, TSFMs may implicitly solve tasks beyond their original training objectives. Clever task framing may reveal latent abilities of such models [52].

Generative AI powered by LLMs has advanced through methods such as Low-Rank Adaptation, which enables task-specific fine-tuning without the computational demands of a full retraining [53], and by methods such as intermediate-layer distillation to make them suitable for resource-constrained environments [54].

Future foundation models can enhance adaptability to entire classes of systems rather than specific ones, including zero-shot capabilities [55, 56].

3 Conclusion

Machine learning for industrial applications demands both efficiency and resilience. This paper provides an overview of key challenges, ranging from data limitations to the need for explainability. The quest for truly adaptive, scalable, and trustworthy ML solutions remains an active and exciting area of research, accelerated by the recent transformation enabled by foundation models.

References

[1] P. Swazinna et al. Behavior constraining in weight space for offline reinforcement learning. In *ESANN*, 2021.

[2] P. Wissmann et al. Is Q-learning an ill-posed problem? In *ESANN*, 2025.

[3] P. Henderson et al. Deep reinforcement learning that matters. In *AAAI*, 2018.

[4] S. Lapuschkin et al. Unmasking Clever Hans predictors and assessing what machines really learn. *Nature Communications*, 2019.

[5] J.R. Kauffmann et al. Explainable AI reveals Clever Hans effects in unsupervised learning models. *Nat. Mac. Intell.*, 2025.

[6] D. Hein et al. Interpretable policies for reinforcement learning by genetic programming. *Engineering Applications of Artificial Intelligence*, 2018.

[7] D. Hein et al. A benchmark environment motivated by industrial control problems. In *SSCI*, 2017.

[8] R.-J. Qin et al. NeoRL: A near real-world benchmark for offline reinforcement learning. In *NeurIPS*, 2022.

[9] P. Ren et al. A survey of deep active learning. *ACM Comput. Surv.*, 2021.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

[10] Z. Zhu et al. Transfer learning in deep reinforcement learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2023.

[11] Y. Lu et al. Machine learning for synthetic data generation: A review. *arXiv:2302.04062*, 2023.

[12] C. Shorten et al. A survey on image data augmentation for deep learning. *J. Big Data*, 2019.

[13] M. Weber et al. Learning control policies for variable objectives from offline data. In *SSCI*, 2023.

[14] P. Bergmann et al. The MVTec anomaly detection dataset: A comprehensive real-world dataset for unsupervised anomaly detection. *Int. J. Comput. Vis.*, 2021.

[15] L. Ruff et al. Deep one-class classification. In *ICML*, 2018.

[16] K. Roth et al. Towards total recall in industrial anomaly detection. In *CVPR*, 2022.

[17] M. Messing et al. Label-efficient and adaptable image selection for large-scale e-commerce catalogs. In *ESANN*, 2026.

[18] M.W. Akram et al. Variational deep embedding for unsupervised clustering of industrial noise in steelmaking plants. In *ESANN*, 2026.

[19] A. Blázquez-García et al. A review on outlier/anomaly detection in time series data. *ACM Comput. Surv.*, 2022.

[20] Y. Wang et al. Multimodal industrial anomaly detection via hybrid fusion. In *CVPR*, 2023.

[21] S. Malacarne et al. Context-aware graph attention for unsupervised telco anomaly detection. In *ESANN*, 2026.

[22] D. Pyle. *Data preparation for data mining*. Morgan Kaufmann, 1999.

[23] S.-A. Rebuffi et al. Data augmentation can improve robustness. In *NeurIPS*, 2021.

[24] Y. Ganin et al. Unsupervised domain adaptation by backpropagation. In *ICML*, 2015.

[25] J. Dean et al. Large scale distributed deep networks. In *NeurIPS*, 2012.

[26] J.A. Carvajal Soto et al. An online machine learning framework for early detection of product failures in an Industry 4.0 context. *IJCIM*, 2019.

[27] G. Hinton et al. Distilling the knowledge in a neural network. *arXiv:1503.02531*, 2015.

[28] F. Hutter et al., editors. *Automated Machine Learning - Methods, Systems, Challenges*. The Springer Series on Challenges in Machine Learning. Springer, 2019.

[29] M. Zhao et al. Adversarial training: A survey. *arXiv:2410.15042*, 2024.

[30] G. Kunapuli. *Ensemble methods for machine learning*. Simon and Schuster, 2023.

[31] S. Depeweg et al. Decomposition of uncertainty in Bayesian deep learning for efficient and risk-sensitive learning. In *ICML*, 2018.

[32] L. Longo et al. Explainable artificial intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions. *Inf. Fusion*, 2024.

[33] S. Letzgus et al. An explainable AI framework for robust and transparent data-driven wind turbine power curve models. *Energy and AI*, 2024.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

[34] J. Vielhaben et al. Explainable AI for time series via virtual inspection layers. *Pattern Recognit.*, 2024.

[35] C. J. Anders et al. Finding and removing Clever Hans: Using explanation methods to debug and improve deep models. *Inf. Fusion*, 2021.

[36] P. Naumann et al. Wasserstein distances made explainable: Insights into dataset shifts and transport phenomena. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2026.

[37] C.J. Anders et al. Software for dataset-wide XAI: From local explanations to global insights with Zennit, CoRelAy, and ViRelAy. *PloS one*, 2026.

[38] F. Doshi-Velez et al. Towards a rigorous science of interpretable machine learning. *arXiv:1702.08608*, 2017.

[39] T. Wang et al. A Bayesian framework for learning rule sets for interpretable classification. *J. Mach. Learn. Res.*, 2017.

[40] A. Perzylo et al. Intuitive instruction of industrial robots: Semantic process descriptions for small lot production. In *Intelligent Robots and Systems (IROS)*, 2016.

[41] D. Hein et al. Trustworthy AI for process automation on a Chylla-Haase polymerization reactor. In *GECCO*, 2021.

[42] G. Mondal et al. Enhancing power converter efficiency with reinforcement learning-based PWM control. In *IECON*, 2025.

[43] M. Courbariaux et al. Binarized neural networks: Training deep neural networks with weights and activations constrained to +1 or -1. *arXiv:1602.02830*, 2016.

[44] S. Singh et al. Pruning and quantization for deeper artificial intelligence (AI) model optimization. In *RCAAI*, 2022.

[45] A. G. Howard et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv:1704.04861*, 2017.

[46] Y. Abadade et al. A comprehensive survey on TinyML. *IEEE Access*, 2023.

[47] M. Shafiq et al. Continuous quality control evaluation during manufacturing using supervised learning algorithm for Industry 4.0. *Int. J. Adv. Manuf. Technol.*, 2023.

[48] B.B. Ormancı et al. TEA: Trajectory encoding augmentation for robust and transferable policies in offline reinforcement learning. In *ESANN*, 2025.

[49] A.F. Ansari et al. Chronos-2: From univariate to universal forecasting. *arXiv:2510.15821*, 2025.

[50] C. Feng et al. General time transformer: An encoder-only foundation model for zero-shot multivariate time series forecasting. In *CIKM*, 2024.

[51] C. Calisir et al. A comparison of open time-series foundation models for industrial manufacturing applications. In *ESANN*, 2026.

[52] M. Tokic et al. TSFM in-context learning for time-series classification of bearing-health status. In *ESANN*, 2026.

[53] E.J. Hu et al. LoRA: Low-rank adaptation of large language models. *arXiv:2106.09685*, 2021.

[54] X. Jiao et al. TinyBERT: Distilling BERT for natural language understanding. *arXiv:1909.10351*, 2020.

[55] M. Palatucci et al. Zero-shot learning with semantic output codes. In *NeurIPS*, 2009.

[56] A. Touati et al. Does zero-shot reinforcement learning exist? In *ICLR*, 2023.