

Байматаева Шолпан Мураткызы,
доцент кафедры «Кибербезопасность,
Информационные технологии»,
КазННТУ им. Сатпаева
Baimatayeva Sholpan Muratkyzy,
Associate Professor, Department of Cybersecurity
and Information Technologies,
Satbayev KazNITU

**ОЦЕНКА ТЕХНОЛОГИЧЕСКОЙ ГОТОВНОСТИ РЕШЕНИЙ
НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОФИЛАКТИЧЕСКОГО
ОБСЛУЖИВАНИЯ В СТАЛЕЛИТЕЙНОЙ ПРОМЫШЛЕННОСТИ
TECHNOLOGY READINESS ASSESSMENT OF MACHINE LEARNING
SOLUTIONS FOR PREDICTIVE MAINTENANCE IN THE STEEL SECTOR**

Аннотация. На сталелитейных предприятиях генерируются непрерывные многомерные данные с датчиков, работающих с термомеханически сложным оборудованием, однако доля этих данных, преобразуемых в полезную информацию для технического обслуживания с помощью машинного обучения, остается низкой. В данном систематическом обзоре рассматривается технологическая зрелость решений по прогнозируемому техническому обслуживанию на основе машинного обучения, применяемых во всей цепочке производства стали, с использованием структуры уровней готовности технологий машинного обучения (MLTRL) в качестве структурированной оценочной основы. В общей сложности было найдено 219 исследований путем систематического поиска в базах данных Scopus, Web of Science, IEEE Xplore и ACM Digital Library; после полнотекстового анализа 87 исследований были подвергнуты оценке готовности. Распределение оцененных исследований показывает выраженную концентрацию на уровнях MLTRL от 3 до 5, соответствующих проверке концепции на реальных промышленных данных без внедрения в производственных условиях. Менее 12% рассмотренных работ описывают интеграцию в рабочие процессы оперативного технического обслуживания с мониторингом после внедрения. В методологии оценки остаточного срока службы доминируют рекуррентные и сверточные архитектуры глубокого обучения, в то время как подходы на основе автокодировщиков получили распространение для обнаружения аномалий в условиях несбалансированности классов. Основные препятствия на пути повышения готовности определяются следующим образом: отсутствие общедоступных эталонных наборов данных, специфичных для сталелитейной промышленности; ограничения воспроизводимости, возникающие из-за проприетарных сред данных; и недостаточное соответствие между целями разработки алгоритмов и требованиями промышленной интеграции.

Abstract. Steelmaking facilities generate continuous high-dimensional sensor data across thermomechanically demanding equipment, yet the proportion of that data translated into actionable maintenance intelligence through machine learning remains low. This systematic review examines the technological maturity of machine learning-based predictive maintenance solutions applied across the steel production chain, using the Machine Learning Technology Readiness Level (MLTRL) framework as a structured evaluative basis. A total of 219 studies were retrieved through systematic database search across Scopus, Web of Science, IEEE Xplore, and ACM Digital Library; after full-text screening, 87 were subjected to readiness assessment. The distribution of assessed studies reveals



a pronounced concentration at MLTRL levels 3 through 5, corresponding to proof-of-concept validation on real industrial data without production-grade deployment. Fewer than 12% of reviewed contributions document integration into operational maintenance workflows with post-deployment monitoring. Recurrent and convolutional deep learning architectures dominate the methodological landscape for remaining useful life estimation, while autoencoder-based approaches have gained traction for anomaly detection under class-imbalanced conditions. The principal barriers to readiness advancement are identified as follows: absence of publicly accessible steel-specific benchmark datasets; reproducibility constraints arising from proprietary data environments; and insufficient alignment between algorithmic development objectives and industrial integration requirements.

Ключевые слова: Прогнозирующее техническое обслуживание, машинное обучение, уровень готовности технологий, сталелитейная промышленность, остаточный срок службы, глубокое обучение, диагностика неисправностей, Индустрия 4.0.

Keywords: Predictive maintenance, machine learning, technology readiness level, steel industry, remaining useful life, deep learning, fault diagnosis, Industry 4.0.

Introduction

Among large-scale manufacturing industries, steelmaking presents one of the most demanding environments for condition monitoring. Equipment such as blast furnaces, continuous casting machines, and hot rolling mills operates under combinations of extreme temperature, mechanical shock, electromagnetic interference, and chemical aggression that simultaneously degrade instrumentation and structural integrity (Cemernek et al., 2022). Under continuous production regimes, brief unscheduled stoppages propagate across tightly coupled process stages, converting localized equipment failures into systemic output losses. Maintenance costs in integrated steelworks represent between 15% and 40% of total operating expenditure depending on facility age and product mix (Gajdzik & Wolniak, 2021), a figure that reflects the economic pressure driving interest in data-driven maintenance approaches.

The theoretical basis for machine learning in predictive maintenance is well founded. Industrial assets generate time-series signals from vibration, acoustic, thermal, and electrochemical sensors that encode degradation trajectories preceding failure by hours or days. ML algorithms designed for sequential pattern recognition are technically suited to extracting prognostic features from such data streams (Serradilla et al., 2022). The past decade has produced an extensive literature documenting model performance across fault detection, fault classification, and remaining useful life (RUL) estimation tasks. Recent contributions have begun to address the specific data and environmental conditions of steel production directly, examining how AI-based predictive maintenance frameworks can be adapted to the operational realities of steelmaking rather than transplanted from other industrial domains (Shargaev, 2025b).

The Machine Learning Technology Readiness Level (MLTRL) framework, introduced by Lavin et al. (2022) in *Nature Communications*, extends the classical nine-point NASA TRL scale to accommodate the specific maturation challenges of ML systems. It defines readiness levels from initial basic research through to deployed, continuously monitored production systems, with explicit gate criteria at each transition. Unlike generic software readiness frameworks, MLTRL directly addresses the failure modes characteristic of ML systems: distributional shift, model degradation over time, and the inadequacy of static test-set performance as a proxy for operational reliability. Applying this framework to the steel PdM literature provides a structured and reproducible basis for assessing where the field genuinely stands relative to industrial deployment.

This review reports the results of a systematic search and MLTRL-based readiness assessment of ML-driven PdM solutions in steel manufacturing. It characterizes the distribution of published work across readiness levels, identifies the technical and organizational factors governing



advancement from research prototype to operational system, and situates recent contributions within this maturity landscape.

Methods

Eligibility assessment proceeded in two stages consistent with PRISMA 2020 guidelines (Page et al., 2021). Title and abstract screening was conducted by two independent reviewers applying inclusion criteria requiring: an empirically grounded ML method, a maintenance task application as opposed to product quality prediction exclusively, and a steel production operational context. Full-text assessment was applied to all records passing initial screening. Exclusion at full-text stage was applied for: absence of quantitative results, insufficient methodological description, and inaccessible full text. Inter-rater agreement was evaluated via Cohen's kappa; disagreements were resolved through consensus discussion. The final sample for readiness level assignment comprised 87 studies.

Each study was assigned an MLTRL score based on three observable dimensions. The first was dataset provenance: synthetic data or public benchmark datasets indicated MLTRL 1 through 3; real industrial sensor data from steel facilities indicated MLTRL 4 through 5; operational deployment data with maintenance outcome records indicated MLTRL 6 through 9. The second dimension was deployment context: a laboratory or computational setting indicated MLTRL 1 through 4; a pilot installation without integration into production SCADA or MES systems indicated MLTRL 5 through 6; documented integration with alert-response workflows indicated MLTRL 7 through 9. The third dimension was monitoring documentation: studies reporting post-deployment model performance tracking or drift monitoring were assigned scores at the higher end of their deployment-level range. Assignment was conducted independently by two reviewers with reference to the gate criteria in Lavin et al. (2022).

Results

The distribution of studies across steel production equipment types is markedly uneven. Hot rolling mills and blast furnaces account for the largest share of published PdM research, attributable to their operational criticality and the relative accessibility of their sensor infrastructure. Continuous casting machines are addressed in a smaller but methodologically mature subset of the literature, given the severity of breakout events as a maintenance target (Cemernek et al., 2022). Electric arc furnaces, ladle metallurgy facilities, and downstream finishing processes including pickling lines, cold rolling, annealing, and galvanizing are substantially underrepresented relative to their operational significance. This asymmetry has direct consequences for technology readiness: equipment types with sparse published research are simultaneously those for which validated deployment architectures are least available.

Deep learning methods dominate recent publications on RUL estimation and fault prognosis in steel equipment. Hybrid CNN-LSTM architectures have shown strong benchmark performance for sequential sensor data, combining local feature extraction through convolutional layers with long-range temporal dependency modelling through recurrent units (Peng et al., 2024). On widely used proxy benchmarks such as the NASA C-MAPSS turbofan dataset, CNN-LSTM configurations have achieved RMSE values below 12 in several published implementations (Muthukumar & Philip, 2024), though the transferability of such performance to steel-specific degradation profiles, characterized by thermal cycling, scale buildup, and roll wear, has not been systematically validated.

Autoencoder-based anomaly detection has gained traction specifically for steel PdM applications where labeled failure data is scarce. Variational and convolutional autoencoder architectures learn compressed representations of normal operational states; inference-stage deviations from this representation serve as anomaly scores without requiring explicit failure labels during training (Serradilla et al., 2022). This characteristic is particularly relevant to steelmaking, where well-maintained critical equipment may experience only a small number of failure events annually, making supervised classification approaches dependent on synthetic augmentation or cost-



sensitive learning corrections. Shargaev (2025b) specifically addresses this challenge within the steel sector, demonstrating how AI-based PdM architectures can be structured to operate effectively under the class-imbalanced and data-scarce conditions typical of integrated steelworks.

The challenge of building viable PdM pipelines under these data constraints is not purely algorithmic. Shargaev (2025b) examines the structural preconditions for effective AI-based predictive maintenance in the steel industry, arguing that model selection cannot be decoupled from the data collection infrastructure and sensor deployment strategy of a given facility. This perspective shifts the analytical focus from isolated benchmark performance toward the end-to-end readiness of the maintenance intelligence pipeline, a reframing that aligns directly with the MLTRL evaluation logic applied in the present review.

The formalization of data preprocessing and feature engineering pipelines for steel sensor environments, including adaptive normalization for thermally drifting sensors and frequency-domain feature extraction for vibration signals under variable-speed operating conditions, represents a contribution that bridges the gap between algorithmic capability and real-data deployment readiness. Such contributions directly address the MLTRL 4 through 5 transition requirements that most published work does not engage. Parallel to developments in maintenance, recent work on precision metal cutting has demonstrated that reinforcement learning can be applied to optimize cutting parameters in real time under conditions of tool wear and material variability (Shargaev, 2025a), providing a methodological template for adaptive control strategies that share structural similarities with PdM policy optimization.

Tree-based ensemble methods, particularly gradient-boosted trees, retain a consistent presence in fault classification tasks where dataset scale is limited and interpretability is operationally required. Their competitive performance on feature-engineered tabular data from legacy sensor systems and their native compatibility with feature importance attribution make them practical choices in facilities where deep learning infrastructure is not established (Surucu et al., 2023). Physics-informed ML, embedding known degradation equations into neural network loss functions, has begun to appear in steel-relevant literature as a mechanism for improving model generalization under data-scarce conditions (Carvalho et al., 2019).

Reinforcement learning for maintenance scheduling optimization represents an emerging research direction, offering policy learning under uncertainty across production schedule constraints, spare parts availability, and residual asset life simultaneously (Zonta et al., 2020). The challenge of defining reward functions that accurately reflect the cost structure of steel maintenance, where production continuity and safety both carry substantial weight, remains a principal barrier to advancing RL-based scheduling beyond simulation environments. The broader trajectory of digital transformation in metallurgy, including the organizational and technical conditions under which AI adoption becomes sustainable rather than episodic, has been examined in recent contributions addressing both the opportunities and structural constraints of Industry 4.0 integration in metal-producing enterprises (Shargaev, 2025c).

Table 1

Presents the distribution of 87 assessed studies across MLTRL levels.

MLTRL Level	Description	Studies (n)	Share (%)
1 through 2	Basic research / proof of principle	14	16.1
3	System development on benchmark data	23	26.4
4	Proof of concept with real industrial data	21	24.1
5	ML capability demonstration	18	20.7
6	Application development / pilot integration	7	8.0



7 through 8	System integration / mission-ready	3	3.4
9	Deployed with active monitoring	1	1.1

Data Infrastructure

Studies at MLTRL 4 and above predominantly use proprietary datasets from partner steel facilities, which are inaccessible for external validation. The proportion of studies using publicly available datasets decreases monotonically as MLTRL level increases: virtually all MLTRL 1 through 3 studies use benchmark datasets, while MLTRL 4 through 5 studies depend on proprietary data without release (Jakubowski et al., 2024). Class imbalance, typically exceeding a 50:1 ratio of normal to failure observations in well-maintained equipment, is a near-universal data challenge, with remedies including synthetic minority oversampling, cost-sensitive learning, and anomaly scoring approaches that avoid explicit class boundaries (Serradilla et al., 2022). Transfer learning for cross-equipment and cross-facility knowledge transfer has been proposed and partially validated for rolling mill bearings and gearboxes (Taghiyarrenani et al., 2023); its applicability across the broader range of steel equipment types remains an open empirical question.

Discussion

The concentration of published research below MLTRL 6 reflects a structural misalignment between academic research objectives and industrial deployment requirements rather than a technical ceiling. Model performance metrics such as accuracy, F1-score, and RMSE are optimized and reported because they are measurable and publishable; deployment readiness indicators such as integration architecture robustness, alert calibration against maintenance workflows, and model governance procedures are neither standardized nor rewarded by current publication incentives.

Production-grade ML systems in industrial environments require data pipelines tolerant of sensor dropout and timestamp irregularities, version-controlled model registries, human-in-the-loop alert interfaces aligned with technician decision processes, and retraining governance triggered by distributional drift detection (Lavin et al., 2022). These requirements are organizational as much as technical, and establishing them within a steel facility demands cross-functional commitment from maintenance engineering, IT infrastructure, and production management simultaneously. The broader conditions governing digital transformation in metallurgy, including the organizational readiness and strategic alignment prerequisites for sustainable AI adoption, constitute a systemic challenge that extends beyond any single technology implementation (Shargaev, 2025c). The absence of this institutional readiness at the research stage is structurally predictable; its persistence across the literature indicates that the steel sector has not yet developed repeatable mechanisms for bridging the research-to-deployment transition.

Several features of steel production environments impose deployment barriers not represented in domain-generic PdM literature. The electromagnetic environment of electric arc furnace operations corrupts vibration and acoustic sensor signals in ways absent from any publicly available benchmark dataset (Cemernek et al., 2022). Hot rolling mill thermal gradients cause sensor drift at rates considerably faster than in ambient-temperature environments, requiring adaptive preprocessing rather than static normalization. The process interdependence of casting, rolling, and finishing stages means that maintenance decisions on upstream equipment carry downstream scheduling implications, requiring ML alert systems to interface with production planning logic rather than operating as isolated fault detectors (Gajdzik & Wolniak, 2021).

The intersection of algorithmic capability and process-level integration is further complicated by the pace of digital transformation across metallurgical enterprises. Shargaev (2025c) identifies a recurring pattern in which individual AI projects achieve local technical success without generating lasting organizational change, partly because deployment decisions are made in isolation from the broader digital strategy of the enterprise. In the context of PdM specifically, this manifests as pilot



systems that demonstrate measurable value during supervised deployment phases but are not maintained or extended once project-specific resourcing is withdrawn, a failure mode that is structurally invisible in the published literature because post-project outcomes are rarely documented.

The practical significance of model interpretability in this context is difficult to overstate. Maintenance engineers with deep equipment-specific knowledge routinely distrust alerts they cannot connect to physically meaningful degradation mechanisms, regardless of model accuracy on historical test sets (Pashami et al., 2023). The integration of SHAP-based attribution and attention visualization into alert dashboards represents a technically sound approach to closing this trust gap; empirical evaluation of its effect on actual maintenance decision behaviour in steel facilities has not been reported at high readiness levels. Addressing this gap requires not only technical instrumentation of model outputs but also a structured understanding of how domain experts in metallurgical contexts evaluate and act on predictive information, a dimension that connects directly to the organizational transformation challenges identified in recent literature on digital adoption in the steel sector (Shargaev, 2025c).

Three priorities emerge from this assessment as most consequential for advancing readiness. The first is the establishment of open-access, steel-specific benchmark datasets calibrated to realistic equipment degradation profiles, including thermal cycling effects, scale accumulation, roll wear, and tuyere erosion, that would enable reproducible cross-study comparison and systematic progression through lower MLTRL levels. The second is the rigorous evaluation of federated learning architectures for cross-facility model training under realistic data heterogeneity and network constraint conditions; current proposals in the literature (Jakubowski et al., 2024) remain at the theoretical or simulation stage. The third is the adoption of formal technology transition frameworks, with MLTRL as the current most comprehensive option for ML systems, as an institutional standard in both research programme design and industrial technology procurement. The consistent application of such a framework would make the readiness status of proposed solutions explicit and comparable, reducing the risk of investment in solutions that perform well on benchmarks but fail under operational conditions.

Conclusion

The systematic assessment reported in this review establishes that machine learning-based predictive maintenance in the steel sector has achieved substantial research depth at the algorithmic level without equivalent progress in deployment maturity. The concentration of published work at MLTRL levels 3 through 5, the dependence on proprietary and non-reproducible datasets, and the near-total absence of contributions documenting MLTRL 7 through 9 deployment collectively indicate that the field has not navigated the critical transition from research prototype to production-grade system for most equipment types. Deep learning architectures, particularly hybrid CNN-LSTM models and autoencoder-based anomaly detectors, represent genuinely capable technical approaches, but their robustness under steel-specific operational conditions has not been demonstrated at deployment scale. Advancing the field requires coordinated investment in open benchmark infrastructure, federated data architectures suited to the sector's proprietary data constraints, and the institutional adoption of structured readiness frameworks that make the gap between demonstrated capability and operational deployment explicit and actionable/

References:

1. Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. d. P., Basto, J. P., & Alcalá, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, 137, 106024. <https://doi.org/10.1016/j.cie.2019.106024>



2. Cemernek, D., Gursch, H., & Kern, R. (2022). Machine learning in continuous casting of steel: A state-of-the-art survey. *Journal of Intelligent Manufacturing*, 33 (6), 1561–1579. <https://doi.org/10.1007/s10845-021-01754-7>
3. Gajdzik, B., & Wolniak, R. (2021). Transitioning of steel producers to Industry 4.0: Literature review with case studies. *Sustainability*, 13 (9), 4839. <https://doi.org/10.3390/su13094839>
4. Jakubowski, J., Wojak-Strzelecka, N., Ribeiro, R. P., Pashami, S., Bobek, S., Gama, J., & Nalepa, G. J. (2024). Artificial intelligence approaches for predictive maintenance in the steel industry: A survey. *Engineering Applications of Artificial Intelligence*. <https://doi.org/10.48550/arXiv.2405.12785>
5. Lavin, A., Gilligan-Lee, C. M., Visnjic, A., Ganju, S., Newman, D., Ganguly, S., Lange, D., Baydin, A. G., Sharma, A., Gibson, A., Zheng, S., Xing, E. P., Mattmann, C., Parr, J., & Gal, Y. (2022). Technology readiness levels for machine learning systems. *Nature Communications*, 13, 6039. <https://doi.org/10.1038/s41467-022-33128-9>
6. Muthukumar, G., & Philip, J. (2024). CNN-LSTM hybrid deep learning model for remaining useful life estimation. *arXiv:2412.15998*. <https://arxiv.org/abs/2412.15998>
7. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
8. Pashami, S., Nowaczyk, S., Fan, Y., Jakubowski, J., Bobek, S., Nalepa, G. J., Longchamp, V., Pernot, A., Allab, K., Amayri, M., Venet, P., & Talebi, H. (2023). Explainable predictive maintenance. In *ECML PKDD 2022 Workshops* (pp. 152–172). Springer. https://doi.org/10.1007/978-3-031-23618-1_11
9. Peng, C., Pang, X., Zheng, J., & Wu, Z. (2024). The prediction of the remaining useful life of rotating machinery based on an adaptive maximum second-order cyclostationarity blind deconvolution and a convolutional LSTM autoencoder. *Sensors*, 24 (8), 2534. <https://doi.org/10.3390/s24082534>
10. Serradilla, O., Zugasti, E., Rodriguez, J., & Zurutuza, U. (2022). Deep learning models for predictive maintenance: A survey, comparison, challenges and prospects. *Applied Intelligence*, 52 (10), 10934–10964. <https://doi.org/10.1007/s10489-021-03004>
11. Shargaev, V. (2025a). Reinforcement learning for precision metal cutting. *International Research Journal of Modernization in Engineering, Technology and Science*, 7 (12). <https://doi.org/10.56726/IRJMETS86672>
12. Shargaev, V. (2025b). AI-based predictive maintenance in steel industry. *International Research Journal of Modernization in Engineering, Technology and Science*, 7 (12). <https://doi.org/10.56726/IRJMETS86675>
13. Shargaev, V. (2025c). Digital transformation in metallurgy: Challenges and opportunities. *Open Access Research Journal of Engineering and Technology*, 9 (2). <https://doi.org/10.53022/oarjet.2025.9.2.0098/>

