Proceedings of the Seventh International Conference on
Railway Technology: Research, Development and Maintenance
Edited by: J. Pombo
Civil-Comp Conferences, Volume 15, Paper 8.2
Civil-Comp Press, Edinburgh, United Kingdom, 2026
ISSN: 2753-3239, doi: 10.4203/ccc.15.8.2

Transfer Learning in Railway Health Monitoring: Taxonomy, Applications and Future Directions

A. Sansiñena[1,2], B. Rodríguez-Arana[1,2] and S. Arrizabalaga[1,2,3]

[1] Ceit-Basque Research and Technology Alliance (BRTA), Donostia-San Sebastián, Spain
[2] Tecnun, Universidad de Navarra, Donostia – San Sebastián, Spain
[3] Institute of Data Science and Artificial Intelligence (DATAI), Universidad de Navarra, Pamplona, Spain

Abstract

The deployment of Machine Learning (ML) and Deep Learning (DL) approaches for railway health monitoring is frequently hindered by the scarcity of labelled data and the inherent domain shift between controlled laboratory conditions and diverse operational environments. In recent years, Transfer Learning (TL) has emerged as a critical paradigm to enable the application of ML and DL models in real scenarios, ensuring the strict safety and precision conditions required in the railway industry. In this paper, a comprehensive survey on the recent advancements in TL with regard to the assessment of railway assets is provided. A taxonomy is established based on the availability of data in the target domain to serve as a brief guide for new researchers interested in TL. To highlight the potential of these techniques, the latest applications on different railway assets, including faults in tracks, brake pads and bearings, are evaluated. Finally, trends and gaps for the future research are identified.

Keywords: transfer learning, railway infrastructure, faults assessment, machine learning, data scarcity, unsupervised domain adaptation

1 Introduction

Monitoring and fault detection of railway infrastructure assets has become an increasing topic of research [1]. The emergence of ML techniques has marked a turning point in the ability to predict and detect potential infrastructure failures, thereby reducing maintenance cost and enhancing safety. However, ML algorithms, particularly DL architectures, are data-driven approaches that require large amounts of labelled data for effective training. In addition, the performance of these models is highly sensitive to the conditions and distribution of the training data. Consequently, inferring from data with a different distribution typically results in a substantial decline in predictive performance. This represents the primary bottleneck for their deployment in real-world scenarios, where acquiring sufficient labelled training data is frequently unfeasible.

Suppose a model is trained on a source domain (e.g., synthetic on-board acceleration data generated via simulations of a train traversing a switch) to perform a task, such as fault detection. If we intend to deploy this model in a different domain, named the target domain (e.g., real-world operational switches), where labelled data is scarce or non-existent, a significant performance gap arises. In such scenarios, TL provides a framework for adapting knowledge from the source domain to the target domain, aiming to mitigate the challenges posed by data scarcity.

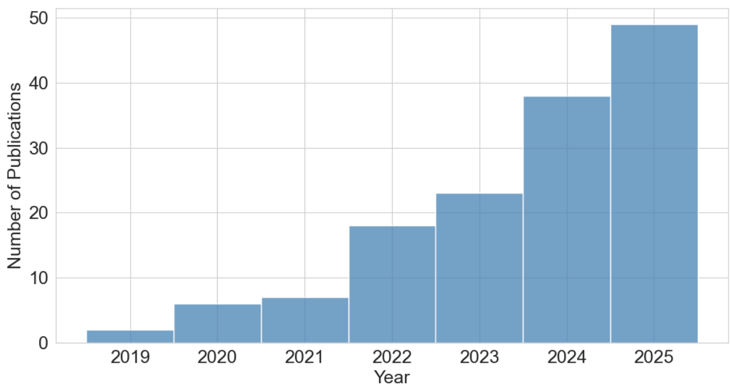

Figure 1: Publications on TL applied to railway infrastructure per year.

A bibliometric analysis conducted via the Scopus research database illustrates the novelty and growing interest of this topic within the railway sector, as evidenced by the increasing number of publications shown in Figure 1. The following query was used to search within article title, keywords and abstract: (”domain adaptation” OR ”transfer learning” OR ”domain shift” OR ”cross-domain”) AND (”fault detection” OR ”defect detection” OR ”monitoring” OR ”anomaly detection” OR ”damage detection” OR ”diagnosis”) AND (”railway” OR ”rail track” OR ”railroad” OR ”track geometry”). Driven by both the technical necessity to mitigate the aforementioned limitations and the clear academic momentum identified, this work provides a survey of TL methodologies currently employed, or with significant potential, for the health assessment of railway components. A representative selection of the articles extracted with the query will be analysed to encompass the different methodologies presented.

The remainder of this article is structured to provide a comprehensive overview of the possible uses of TL within the railway sector. First, we establish a taxonomy of TL techniques, categorised primarily by their data requirements in the target domain. Following this theoretical framework, we present various applications in railway infrastructure assessment from recent literature. By synthesising these methodologies and their practical implementations, this work aims to demonstrate the versatility of TL and serve as a decision-making guide for researchers, assisting them in analysing and selecting the optimal adaptation techniques for their work.

2 Brief Overview of Transfer Learning Techniques

In this section, fundamental TL methodologies are analysed and classified based on their data requirements. The proposed taxonomy is divided into two stages. First, we distinguish between techniques that require labelled target samples and those designed for unsupervised adaptation, reflecting the pragmatic constraints of real-world applications. Then, they are further subdivided according to their underlying methodology, based on similar taxonomies found in the literature [2, 3].

2.1 Labelled data in target domain

It is a common scenario in the railway industry to have a small amount of labelled target data that is not sufficient to effectively train a ML model from scratch. It is especially challenging to find enough degraded or faulty assets. This type of TL problem is often denoted as Supervised Domain Adaptation (SDA).

Pretrained models

Fine-tuning of pretrained models is a widely used SDA approach for DL models. This strategy not only reduces the data requirements but also the computational cost of training a new model from scratch. The task of the original pretrained model can be related, for instance, a model that detects faults in a high-speed line can be adapted to fault detection in regional railways. Alternatively, general-purpose backbone models can be utilised. These are typically large-scale architectures trained on massive benchmark datasets, specialised in extracting robust, high-level features from specific data types.

Adaptation is achieved through fine-tuning: the model is retrained with a low learning rate to adapt the weights to the new data distribution while preserving useful learned characteristics. While the entire architecture can be retrained, a more common practice involves partial freezing, where the early feature-extraction layers remain fixed, and only the final task-specific layers are updated.

Meta-Learning and Few-shot Learning

In railway infrastructure, it is rare to find examples of faulty data. This constitutes a strong drawback for the development of ML models, even if techniques such as the aforementioned pretrained models are used. These scenarios in which models must be trained with just a handful of labelled target samples are known as Few-Shot Learning (FSL).

Meta-Learning, or *learn to learn*, is the most popular way to solve FSL problems. Instead of learning a specific task, Meta-Learning trains the model on a wide variety of auxiliary tasks to adjust its parameters. This enables the architecture to acquire a ”prior” knowledge that allows it to rapidly adapt its parameters to a new target domain with minimal supervision. In infrastructure monitoring, this is particularly useful for adapting a general fault detector to the unique signature of a specific, new asset [4].

2.2 Unlabeled data in target domain

When no labelled data is available in the target domain, UDA addresses this problem by searching for hidden structures and patterns shared between the source and target domains to minimise domain shift. Although researchers must consider the limitations in the discrepancy between domains that these approaches can handle, this scenario is the most common in the railway industry, both in the development and maintenance phases of an assessment system.

Distribution alignment

This strategy focuses on reducing the statistical discrepancy between the source and target feature distributions. Mathematical frameworks such as Maximum Mean Discrepancy (MMD) or Correlation Alignment (CORAL) are commonly employed to minimise the distance between the feature means or second-order statistics (covariances) of the two domains. By forcing the model to map both domains into a shared, statistically similar feature space, the classifier becomes robust to environmental variations, such as changes in track stiffness or vehicle speed.

Adversarial Methods

Adversarial methods leverage the principle of competitive training to identify a domain invariant feature representation. As illustrated in Figure 2, both labelled data from the source domain and unlabelled data from the target domain are fed into a shared feature extractor. Extracted features are then used to train the predictor and a domain discriminator. This architecture is designed to simultaneously minimise the prediction loss of the label predictor while ’fooling’ the domain discriminator by forcing the network to learn features that are both highly informative for the diagnostic task and statistically indistinguishable across different domains.

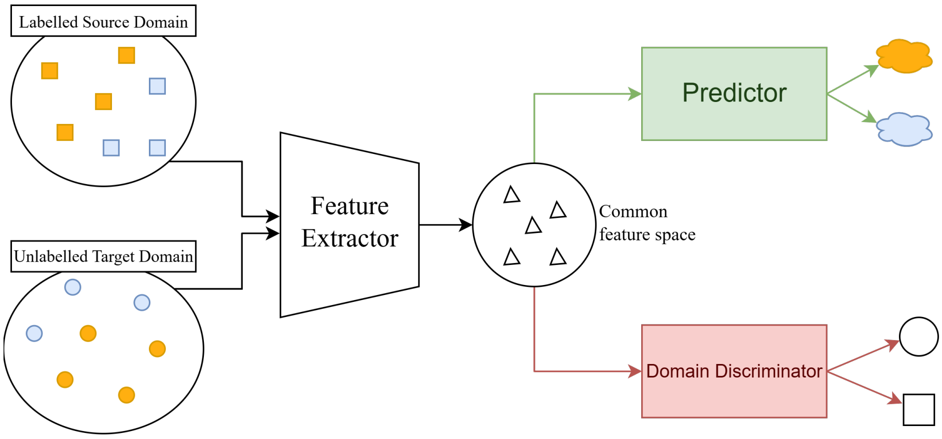

Figure 2: Schematic representation of a domain adversarial architecture featuring a shared feature extractor, label predictor, and domain discriminator.

Self-supervision

Self-supervised learning (SSL) is a paradigm in which a model is trained with unlabeled data to perform simple pretext tasks that lead it to develop a feature representation of the underlying properties. For instance, SSL can be used for image segmentation by implementing image rotation prediction as a pretext task: the model is forced to recognise objects and their orientation in an image [5]. More customised tasks can be used for focused jobs in the railway infrastructure, such as catenary support elements detection [6].

The effectiveness of self-supervised learning in domain adaptation stems from its ability to perform multi-task learning without the need for additional human-annotated labels. By concurrently optimising for a primary task (such as defect detection) in the source domain and a self-supervised pretext task across both domains, the model is forced to align its internal representations. This synergy not only enhances the stability of the training process but also ensures that the feature extractor captures universal patterns rather than domain-specific noise

Pseudo-labelling

Given a model trained on a labelled source dataset, it can be used to perform inference on the target domain. Initially, these predictions are likely to be inaccurate due to the domain shift; however, pseudo-labelling identifies the most confident predictions and treats them as ground-truth labels, termed *pseudo-labels*, to adapt the model in an iterative self-training process. Since the reliability of these labels is not guaranteed, noise accumulation can occur, leading to confirmation bias where the model reinforces its own errors and performance degrades. To mitigate this, it is imperative to implement robust filtering mechanisms and consistency constraints. By integrating dynamic thresholding and heavy stochastic perturbations during the retraining phase, the model is forced to move beyond its initial biased decision boundaries, ensuring that the self-training process converges toward a more generalised representation of the target domain rather than merely reinforcing its own predictive uncertainties [7].

3 Applications Examples in Fault Detection

The techniques outlined in the previous section depict a theoretical framework for mitigating the effects of domain shift. In this section, applications to railway elements assessment of such techniques in recent literature are exposed. The articles in this section have been selected to cover the widest possible range of TL techniques and railway assets. This includes the TL techniques disclosed in Section 2 and their application to the assessment of rails, turnout switch machines, wheelset bearings, bogies and braking pads. Table 1 organises the detailed articles according to the proposed taxonomy, categorising each study by its primary TL strategy.

	Strategy	**References**
SDA	Pretrained models	[8, 9, 10]
	Meta-Learning and Few-shot learning	[11]
UDA	Distribution alignment	[12, 13, 14, 15]
	Adversarial Methods	[16, 17]
	Self-supervision	[5, 18, 19]
	Pseudo-labelling	[7, 20]

Table 1: Classification of reviewed articles.

Within the scope of SDA, the majority of studies use fine-tuning to adapt a trained model to different domains. In [8], the authors train the AlexNet architecture on the large ImageNet dataset, which contains more than a million images, enabling it to effectively identify useful patterns for general image analysis tasks. Only the last layers of AlexNet are retrained on the frequency-spectrum dataset for fault detection in wheelset-axlebox assemblies. The ImageNet dataset is also used in [9] for pretraining several models, including ResNet-101, AlexNet, GoogleNet, and DenseNet. Authors show that the integration of these pretrained models with a BiLSTM outperforms other models trained from scratch for ride comfort classification. These studies employ a generic large dataset to train the feature extraction layers of the networks, so they can be rapidly adapted to other tasks. Despite the advantages that this technique offers in some cases, it may be detrimental in other cases: for instance, some critical physics-related characteristics may not be captured. In [10], synthetic data from simulations are leveraged to pre-train models that learns to extract the relevant physics characteristics for wheel-rail force inversion. This approach exploits the reduced domain shift between physical simulations and actual operational environments, allowing for successful adaptation with a significantly lower volume of real-world labelled samples.

Although Few-Shot Learning constitutes a promising SDA paradigm in the railway industry because the data requirements are aligned with the working conditions in practice, few examples of its application have been found in the literature, perhaps due to the technical difficulty of its implementation or because there is still a lack of knowledge within the field. [11] employs a Meta-Learning framework to achieve robust turnout switch machine fault diagnosis from multi-source signals, obtaining promising results with just five and even one labelled samples.

Among the UDA approaches, distribution alignment is probably the most straightforward and the one most often used in combination with other strategies. In [12], four different methods are employed and compared to construct a mapping transformation to align extracted features from recorded accelerations from two different lines, one labelled and the other unlabeled. The transformed space is then used to train a kNN classifier to predict track status (healthy or damaged). Beyond adaptation, Hu et al. CORAL distance is applied in [13] to minimise the domain shift in the latent space of an autoencoder for fault diagnosis in bearings from vibration signals under different load and speed conditions. While the distribution alignment in these studies is made globally, in the sense that the alignment does not distinguish between internal subgroups termed subdomains, such as different labels, [14] propose a Deep Subdomain Generalisation Network that employs Local Maximum Mean Discrepancy (LMMD) to align class-specific distributions across multiple source domains, enabling the model to generalise to unseen railway brake pad wear scenarios. As illustrated in Figure 3, the primary advantage of LMMD over the standard Maximum Mean Discrepancy (MMD) lies in its granularity; while MMD performs a global alignment of the marginal distributions, potentially misaligning different fault categories that share similar statistical properties, LMMD focuses on subdomain alignment. Similarly, [15] proposes a deep subdomain transfer learning framework for wheelset bearing fault diagnosis, integrating a spatial attention ConvLSTM network to capture spatiotemporal dependencies and applying subdomain alignment to ensure diagnostic robustness under varying high-speed operating conditions

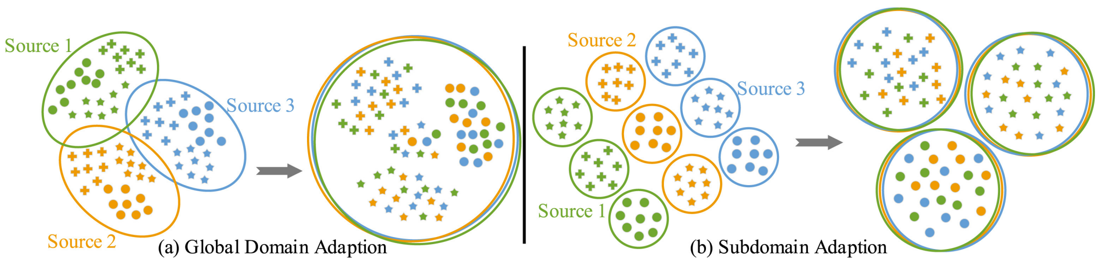

Figure 3: Comparison between global and subdomain distributions alignment for UDA [14].

While statistical alignment focuses on minimizing distance, Adversarial Training adopts a competitive learning strategy to handle more complex domain shifts. [16] deals with the common class imbalance scenario in rolling bearing faults diagnosis through a cross-domain dynamic balanced adversarial network. Unlike standard adversarial neural networks, this method dynamically weighs the contribution of different classes during the adversarial alignment, ensuring that rare bearing faults are not overwhelmed by the dominant healthy patterns. A deep convolutional generative adversarial network (DCGAN) is employed in [17] not just for alignment, but for data augmentation, generating realistic time-frequency images of brake pad damages, effectively expanding the small dataset of vibration signals.

Self-supervision learning for transfer learning remains a relatively unexplored frontier within the railway industry. For instance, Lu et al. [18] propose a framework that utilises self-supervised pretext tasks to extract unbiased features from raw signals, thereby improving the accuracy of fault diagnosis in mechanical components like bearings. In the context of infrastructure maintenance, [19] develop a model that learns to identify rail surface defects by training on simulated anomalies within defect-free samples, effectively reducing the reliance on scarce manual labels. Finally, [5] showcases how simple auxiliary tasks, such as predicting image rotations, can be generalised to enhance object recognition and scene segmentation across varying environmental conditions.

The last category in our taxonomy is pseudo-labelling. As it was exposed in the previous section, an inherent risk of this approach is that incorrect pseudo-labels can propagate and compound errors. To mitigate this risk, a progressive distribution alignment label correction is employed in [7] for railway track monitoring, iteratively updating confidence in the pseudo-labels. Similarly, [20] combines a progressive pseudo-labelling mechanism with adversarial learning to localise bogie faults, employing a dynamic attention network to ensure robust diagnostic performance across varying operational speeds.

4 Conclusions

The main transfer learning techniques used for railway infrastructure assessment have been analysed and classified according to a taxonomy based on target domain data requirements. To illustrate these methods, several application examples have been extracted from the literature to inspect different assets, including tracks, brake pads or bearings. These articles show the potential of transfer learning techniques for the assessment of a variety of assets, and most importantly, how they can be combined with standard supervised learning frameworks to enable their implementation in real industry, overcoming the common limitations of noisy, scarce and unbalanced labelled data. However, the literature on this topic is very recent: most of the reviewed articles were published less than three years ago. In addition, there is still a small number of publications for each topic, which underscores the novelty and necessity for new research. Among the reviewed TL methodologies, fine-tuning for SDA, and distribution alignment and adversarial methods for UDA stand as the most explored ones in the literature. However, meta-learning and few-shot learning, because of the alignment of their data requirements with the actual conditions in railway scenarios, and pseudo-labelling, because of its superior capacity to adapt to new data and to be used alongside other UDA approaches, are the most promising ones for future research.

Furthermore, most TL techniques introduce an additional difficulty for model explainability, which is paramount in the railway industry to rely on and interpret the model predictions. Besides, the lack of public benchmarks in the railway industry makes it impossible for a researcher to effectively compare different techniques without having to implement them one by one or rely on the reported results on other datasets, which are likely private and may not replicate the scenario the researcher is working on. This ultimately hinders the progress of research. We believe these two aspects should be considered to further advance the applicability of ML models in the railway industry.

Acknowledgements

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Europe's Rail Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them. The projects "FP1-MOTIONAL" and "FP3-IAM4Rail" are supported by the Europe's Rail Joint Undertaking and its members

References

[1] I. Bravo, U. Alvarado, J. Nieto, P. Ciáurriz, "On-Board Accelerometers in Railway Track Condition Monitoring. A Systematic Review", Transportation Research Interdisciplinary Perspectives, 33, 1–18, 2025

[2] G. Chen, X. Kong, H. Cheng, S. Yang, X. Wang, "Deep Transfer Learning in Machinery Remaining Useful Life Prediction: A Systematic Review", Measurement Science and Technology, 36, 1–18, 2025

[3] Y. Xiao, H. Shao, S. Yan, J. Wang, Y. Peng, B. Liu, "Domain Generalization for Rotating Machinery Fault Diagnosis: A Survey", Advanced Engineering Informatics, 64, 1–20, 2025

[4] L.-C. Quiñonez-Baca, G. Ramirez-Alonso, A. Guzman-Pando, J. Camarillo-Cisneros, D. R. Lopez-Flores, "Advances in Meta-Learning and Zero-Shot Learning for Multi-Label Classification: A Review", Digital Signal Processing, 163, 1–22, 2025

[5] J. Xu, L. Xiao, A. M. Lopez, "Self-Supervised Domain Adaptation for Computer Vision Tasks", IEEE Access, 7, 156694–156706, 2019

[6] H. Yang, Z. Liu, N. Ma, X. Wang, W. Liu, H. Wang, D. Zhan, Z. Hu, "CSRM-MIM: A Self-Supervised Pretraining Method for Detecting Catenary Support Components in Electrified Railways", IEEE Transactions on Transportation Electrification, 11, 4, 10025–10037, 2025

[7] R. Ghiasi, N. Lestoille, C. Diaine, A. Malekjafarian, ”A Novel Domain Adaptation Method for Drive-by Railway Track Monitoring Using Progressive Distribution Alignment Based on Label Correction”, Structural Health Monitoring, 2025

[8] Y. Bai, J. Yang, J. Wang, Y. Zhao, Q. Li, ”Image Representation of Vibration Signals and Its Application in Intelligent Compound Fault Diagnosis in Railway Vehicle Wheelset-Axlebox Assemblies”, Mechanical Systems and Signal Processing, 152, 1–14, 2021

[9] A. Kasimu, W. Zhou, Q. Meng, Y. Wang, Z. Wang, Q. Zhang, Y. Peng, ”Performance Evaluation of Pretrained Deep Learning Architectures for Railway Passenger Ride Quality Classification”, Alexandria Engineering Journal, 118, 194–207, 2025

[10] T. Zhu, Y. Ren, H. Shi, Y. Ye, P. Feng, Z. Su, C. Yao, G. Ma, ”Wheel-Rail Force Inversion via Transfer Learning-Based Residual LSTM Neural Network with Temporal Pattern Attention Mechanism”, Mechanical Systems and Signal Processing, 224, 1–21, 2025

[11] Y. He, D. He, Z. Lao, Z. Yao, H. Sun, C. He, Z. Yuan, ”A Class-Center Fine-Tuning Prototypical Network for Few-Shot Fault Diagnosis of Turnout Switch Machine Driven by Multi-Source Signals”, Measurement, 242, 2–15, 2025

[12] R. Ghiasi, N. Lestoille, C. Diaine, A. Malekjafarian, ”Unsupervised Domain Adaptation for Drive-by Condition Monitoring of Multiple Railway Tracks”, Engineering Applications of Artificial Intelligence, 139, 1–21, 2025

[13] X. Shao, C.-S. Kim, ”Unsupervised Domain Adaptive 1D-CNN for Fault Diagnosis of Bearing”, Sensors, 22, 1–19, 2022

[14] R. Hu, M. Zhang, X. Meng, Z. Kang, ”Deep Subdomain Generalisation Network for Health Monitoring of High-Speed Train Brake Pads”, Engineering Applications of Artificial Intelligence, 113, 1–11, 2022

[15] J. Wang, S. Yang, Y. Liu, G. Wen, ”Deep Subdomain Transfer Learning with Spatial Attention ConvLSTM Network for Fault Diagnosis of Wheelset Bearing in High-Speed Trains”, Machines, 11, 1–21, 2023

[16] H. Ren, J. Wang, J. Dai, Z. Zhu, J. Liu, ”Dynamic Balanced Domain-Adversarial Networks for Cross-Domain Fault Diagnosis of Train Bearings”, IEEE Transactions on Instrumentation and Measurement, 71, 1–12, 2022

[17] M. Zhang, R. Hu, J. Mo, Z. Xiang, Z. Zhou, ”A Cross-Domain State Monitoring Method for High-Speed Train Brake Pads Based on Data Generation under Small Sample Conditions”, Measurement, 226, 1–13, 2024

[18] N. Lu, H. Xiao, Z. Ma, T. Yan, M. Han, ”Domain Adaptation With Self-Supervised Learning and Feature Clustering for Intelligent Fault Diagnosis”, IEEE Transactions on Neural Networks and Learning Systems, 35, 6, 7657–7670, 2024

[19] Y. Xu, H. Wang, Z. Liu, M. Zuo, ”Self-Supervised Defect Representation Learning for Label-Limited Rail Surface Defect Detection”, IEEE Sensors Journal, 23, 23, 29235–29246, 2023

[20] Q. Liu, B. Wu, N. Qin, D. Huang, T. Wang, ”A Pseudo-Labeling Training-Based Progressive Multi-Scale Adaptive Network for Fault Localization of High-Speed

Train Bogies", Transactions of the Institute of Measurement and Control, 1–11, 2025