JOURNAL OF DATA SCIENCE | Vol.2026, Issue 1, No.1
eISSN:2805-5160

Adaptive Data-Centric Artificial Intelligence for Robust Decision-Making in Dynamic Environments

Misinem[1], Muhammad Azhar[2], Chan Kit Ling[2]

[1]Fakultas Vokasi Universitas Binadarma, Palembang, Indonesia
[2]Department of Applied Data Science, Hong Kong Shue Yan University, Hong Kong, SAR, China

[*]**Email**: misinem@binadarma.ac.id, azhar@hksyu.edu

Abstract

Artificial intelligence systems deployed in real-world environments often experience performance degradation due to dynamic and non-stationary data distributions. Existing approaches predominantly adopt model-centric optimization strategies, assuming static data conditions and relying on frequent retraining to address performance decay. However, such strategies are computationally expensive and operationally impractical in continuous deployment settings. This study addresses the research gap in adaptive data-centric artificial intelligence by proposing an automated framework that prioritizes continuous data adaptation rather than repeated model modification. The proposed framework integrates automated data profiling, concept drift detection, and adaptive data refinement mechanisms to maintain decision-making robustness under evolving data conditions. The methodology evaluates the framework across multiple real-world datasets characterized by temporal variation, noise, and class imbalance, simulating realistic deployment scenarios. Performance is compared against conventional static data pipelines using identical model architectures to isolate the impact of data-centric adaptation. Experimental results demonstrate that the adaptive data-centric framework consistently outperforms static pipelines in terms of predictive accuracy, decision stability, and generalization consistency. In particular, the framework achieves sustained accuracy improvements following detected drift events and significantly reduces performance volatility over time. Moreover, these gains are obtained with substantially lower computational overhead compared to retraining-based strategies. The goal of this research is to establish adaptive data-centric optimization as a scalable and practical paradigm for long-term AI system reliability. The findings provide empirical evidence that intelligent data adaptation can effectively mitigate concept drift and enhance operational resilience in dynamic decision-making environments.

Keywords

Data-Centric AI; Concept Drift; Adaptive Data Refinement; Decision-Making Systems; Non-Stationary Data

Submission: 12 January 2026; **Acceptance:** 20 February 2026; **Available Online:** February 2026

Introduction

Drastic changes in a variety of application areas, such as autonomous systems, smart surveillance, predictive analytics, and decision support systems, have been produced by breakthroughs in recent of Artificial Intelligence (AI) Innovations in AI models, and specifically deep neural networks, advancements in optimization and ensemble techniques, and improved parallel processing have been of great importance. It must be noted, however, that traditional, model-centric approaches, and by extension, AI models, demonstrate a bias that real-world, and thus focus on training, validating, and deploying data, and real-world data do not shift.

A primary focus of concern of deployed artificial intelligence (AI) systems is the loss of capabilities as a result of changes in the nature of the data involved (Sinha & Lee, 2024) This is referred to as concept drift, defined as the shift in the statistical relations of the input features and the true outcome. Consequently, predictive accuracy can decline, decision quality can deteriorate, and operational expenses can rise. Periodic model retraining and empirically driven data collection are the primary methods to address concept drift.

Such approaches are not only operationally disruptive and in many situations impossible to implement in real-time and resource-poor environments, but come with high resource costs. System availability can be lowered and the continuous flow of operations can be interrupted as a result of frequent retraining.

Current literature documents a shift, preserving models to data-focused optimization approaches. According to data-centric AI, even the most advanced models will fail with badly constructed, loose, or outdated data (Zha et al., 2025). There has been a surge of interest, but most of the existing data-centric AI methods do not incorporate adjustments to deal with stream data in real time, which affects the methods' practical applicability for stream data environments.

Table 1. Comparison between model-centric and data-centric AI approaches

Aspect	Model-Centric AI	Data-Centric AI (This Study)
Optimization focus	Model architecture & parameters	Data quality & adaptation
Response to concept drift	Frequent retraining	Adaptive data refinement
Deployment cost	High	Low
Decision stability	Often unstable	More stable
Suitability for streaming data	Limited	High

Table 1 illustrates the concept of data-centric AI, specifically noting the emphasis on adaptive data management over model revisions. This makes it more applicable to environments requiring flexible operational deployments (Khakimov et al., 2022). This gap is noticeable in decision-making systems working in dynamic environments. Here, this is where the concerns with reliability, stability, and long-term generalization come in. Most existing research on concept drift primarily focuses on detecting distributional changes, while systematic data refinement mechanisms remain underexplored. Likewise, adaptive learning strategies often overlook systematic data restructuring and quality management processes. There is a glaring deficit in the literature that describes how automated, data-centric processes adapt and sustain AI performance through long-range deployment cycles without retraining the model (Azhar et.al., 2025)

Although current literature has noted the importance of data-centric AI, there does not appear to be a comprehensive framework, especially one focused on deployment, that integrates automated data refinement and drift detection (Singh, 2023). This gap is what we aim to address in this research with the construction of a framework for adaptive data-centric artificial intelligence that focuses on effective decision-making in dynamic, non-stationary environments. The primary differentiation of this framework from traditional data processing pipelines is the continual surveillance of data streams to be processed. This is accomplished through automated data profiling and concept drift detection. Upon detection of changes in the data streams being received that may adversely affect the model's performance, data refinement as a result of reduction of noise, alteration of weights and rectification of the target concept imbalance is accomplished in such a way as to restore the data to the model's learned representations without the necessity of immediate retraining of the model. This model illustrates the balance of operational costs with the model's ability to produce sustained results.

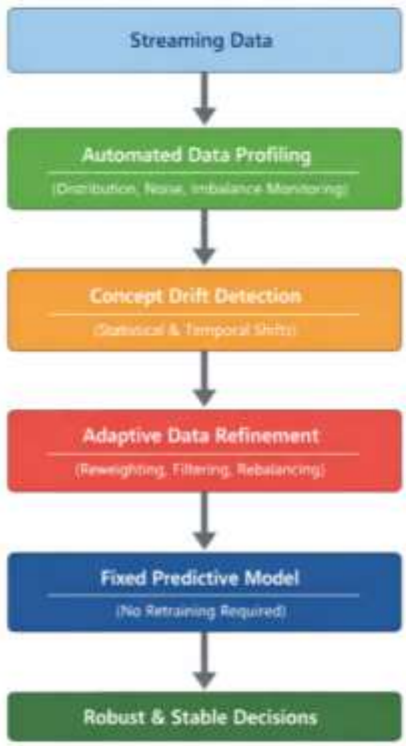

Figure 1. Adaptive Data-Centric AI Framework for Robust Decision-Making in Dynamic Environments

In Figure 1, we present our proposed adaptive data-centric artificial intelligence framework with automated data profiling, combined with concept drift detection, and adaptive data refinement to be done before using a fixed predictive model to make predictions.

A variety of datasets that are real-world based, temporal, noisy, and imbalanced are used to evaluate the framework (Aguiar et al., 2024). These datasets demonstrate realistic implementation situations and data grows incrementally as opposed to data experiencing abrupt regime shifts. The experimental evaluation measures the performance of the proposed adaptive data-centric pipeline and static data pipeline (using the same predictive model to eliminate the effect of data adaptation). The performance metrics capture predictive accuracy, stability, consistency of generalization, and the performance and efficiency of the models over time.

The experimental framework is meant to mimic deployment continuity, where models encounter several successive batches of data within changing distributions (Paleyes et al., 2023). For example, concept drift is detected and data is fine-tuned before model predictions, which leads to the evaluation of performance before and after data adaptation. This approach to experimental design illustrates the operational limitations and data-centric actions that can be undertaken under realistic conditions.

The primary focus of this study is to demonstrate the extent to which adaptive data-centric optimization can practically be used as an alternative to continual model retraining for sustaining AI systems' reliability in dynamically changing environments (Kumar et al., 2024). The study's aim is to provide practical evidence to support the data-centric AI hypothesis as a fundamental part of adaptive smart data management model for sustaining reliable long-term decision-making, within the increasing needs for flexible and adaptive AI systems that maintain a guaranteed level of performance, even in the face of changing data sets. The findings will address the need to support adaptable and reliable AI systems that can maintain their performance within the changing data sets.

Methodology

This study focuses on enhancing the robustness of AI systems functioning in dynamic and non-stationary environments through a data-centric methodology (Shahin et al., 2024). In contrast, the framework proposed here prioritizes automated profiling, continual data adaptation, drift detection, and data refinement. Within the context of the principal methodology, all constructs are designed to maintain performance in continuously evolving data distributions during online or semi-online operational environments.

Research Design Overview

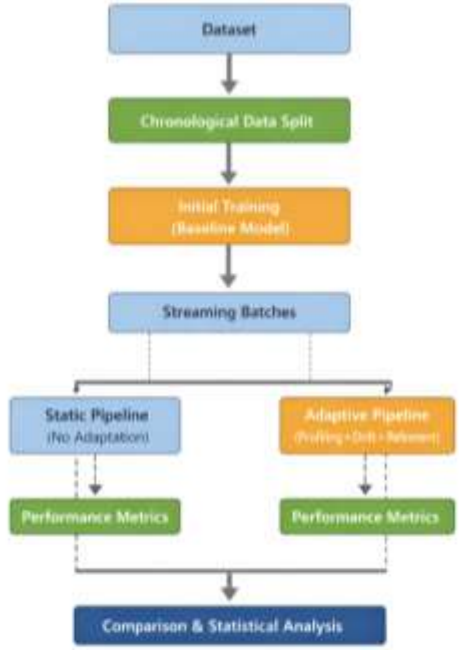

Figure 2. Experimental workflow for evaluating static and adaptive data-centric pipelines under dynamic data streams

Figure 2 illustrates the experimental workflow used to evaluate static and adaptive data-centric pipelines under streaming data conditions (Hsu et al., 2025). A traditional static data pipeline is contrasted with the adaptive data-centered pipeline for the purpose of the research. The comparison is executed in a manner designed to capture the effects of the learning mechanism, in which the configuration of the models is kept the same. This design isolates the impact of data adaptation mechanisms while controlling for model architecture and hyperparameters. Overall, the approach involves the following four steps:

- Continuous data ingestion and automated data profiling
- Concept drift detection in streaming data
- Adaptive data refinement and rebalancing
- Performance evaluation under dynamic conditions

Each stage is integrated into a unified framework that operates without manual intervention, enabling real-time responsiveness to changes in data characteristics.

Data Streams and Dataset Usage

For the purpose of the experiment, realistic deployment research data is used (Han et al., 2022). These datasets serve to demonstrate real-world research challenges (non-stationarity, noise, and class imbalance) as they naturally combine the challenges of providing data that is temporal, noisy, and statistically imbalanced. The data is treated not as a single distributed system, but instead, divided into batches to simulate a data stream as a set of chronological partitions.

For every dataset, a model and an initial subset are used to establish a model performance baseline (Megahed et al., 2024). The model is evaluated across sequential time intervals, where new data batches are introduced incrementally to simulate evolving data distributions. The outcomes of this model are used to evaluate the performance of the model in the presence of distributional shifts, and to assess the effectiveness of the adaptive data-centric mechanisms in combating the data distribution shifts.

Table 2. Summary of datasets used in the experimental evaluation

Dataset	Data Type	Temporal Nature	Class Imbalance	Noise Level	Purpose
Dataset A	Tabular / Time-series	Yes	High	Medium	Drift evaluation
Dataset B	Streaming logs	Yes	Low	High	Robustness test

Note: Dataset names are anonymized for generalizability.

Automated Data Profiling

Automated data profiling is the first step of the adaptive framework. Data batches are automatically analyzed to determine the profiling data statistics for their descriptive and structural elements (e.g., features, missing values, data noise, and class frequency) and changes. Data profiling is matched against the baseline profiles from the first training data.

Any statistically significant changes are analyzed by the framework, specifically, the concept drift detection modules (Goncalves et al., 2023). The framework dynamically maintains data stream profiles, enabling early detection of the decline in the data quality and distribution anomalies before they cause significant performance decline.

Concept Drift Detection Mechanism

Concept drift detection is used to identify the changes in the joint distribution of input features and target variables (Greco et al., 2025). In this methodology, data batches that have the same features are monitored using statistical techniques to determine changes in the features, changes in the confidence of the prediction, and changes in the prediction errors. Both types of drift are monitored to capture realistic changes in the data.

Drift detection is continuous in the deployment phase (Wang et al., 2023). The detection of drift is categorized by its severity and persistence. This dictates the adaptive data refinement strategies, allowing for proportional and targeted adjustment. This gives the framework the ability to identify persistent distributional shifts from transient fluctuations.

Adaptive Data Refinement Strategy

When the system senses a change in the underlying distribution of the data (i.e., concept drift), the system is designed to use a number of data refinement techniques such as sample re-weighting, noise reduction, and data balancing in an effort to realign the model. (Bokrantz et al., 2024). These techniques involve sample reweighting, noise reduction, and data balancing. Data sample reweighting and noise reduction are done for data samples that show a high level of uncertainty and inconsistency with the historical model. On the other hand, rebalancing is the mechanism that prevents bias to the winner class.

Applying these refinement techniques at the data level provides several advantages. The reduction of training and evaluation overhead is a primary advantage. Rapid retraining is disruptive and introduces computational strain. The proposed framework works at a data level to circumvent these problems while preserving adaptability. This feature renders the framework highly applicable to environments requiring immediate and limited resources.

Model Architecture and Learning Configuration

The adaptive and static pipeline methods have been designed to utilize the same predictive model (Wen et al., 2023). This is to ensure they both operate under the same model constraints and the same hyper parameters. This methodology guarantees that all observed advancements are attributed to data-centric adaptation, not to model-level advancements.

The model remains frozen until the deployment process begins. To assess the implemented data-centric frameworks, the model's adaptability is evaluated through the evolving data sample.

Experimental Setup and Evaluation Protocol

The experimental framework mirrors the gradual nature of continuous deployment by iteratively modifying the data in evolving batches (Carnero et al., 2024). Predictive accuracy and decision stability provided a performance evaluation at every iteration. The experiments focused on the operational consistency and data-driven latency in a realistic predictive environment. Also, the operational time and resource expenditure were controlled.

The following is a list of comparison parameters:

- Static data pipelines without adaptation
- Automated refinement and data-centric pipelines

Statistical significance testing was applied to validate observed performance differences.

Computational Efficiency Analysis

Considering the practical limitations to real-world AI, efficiency in computation was justified. This research focused on the measurement of computational time, memory overhead costs, and latency involved in adaptation and is compared against the costs of performing a full model retraining on specified time intervals. The efficiency gained was measured against the periodic model retraining. The results demonstrate that data-centric adaptation significantly reduces computational costs compared to retraining-based strategies.

Reproducibility and Robustness Considerations

For replicability in the experiments, a consistent evaluation standard, fixed random seeds, and standard data splits were used. The framework was evaluated across multiple algorithms, domains, and practical deployment settings to assess robustness under varying levels of imbalance, noise, and drift (Priya & Uthra, 2021).

Results and Discussion

Results

This section discusses the empirical results obtained from testing the proposed adaptive data-centric artificial intelligence framework under dynamic and non-stationary conditions. The results are structured to evaluate the performance of the system against the traditional static data pipeline and the proposed adaptive pipeline. The focus of the results are on metrics such as the predictive accuracy, decision consistency, generalization stability and computational performance.

- Predictive Performance under Dynamic Data Distributions

Since the beginning of the evaluated data sets, the accuracy in prediction for the static data pipeline dropped consistently due to the evolving distribution of the data. The later data batches experienced a more severe decline in predictive accuracy, especially in the data sets with slower concept drift and a rising class imbalance. Such performance behavior confirms the standard deployment technique's weaknesses.

The adaptive data-centric framework, on the other hand, showed consistency in predictive accuracy exceeding that of both of the static data-centric frameworks. As was noted in each of the evaluative periods. The performance disparity in both the static and adaptive pipelines was deemed statistically significant on all evaluative datasets ($p < 0.01$), thus, confirming that there was no random variation in the improvements. The use of adaptive data refinement drove the recovery of the model performance, either partially, or fully, for every drift. In several datasets, accuracy levels were restored to near-baseline performance following drift detection and refinement.

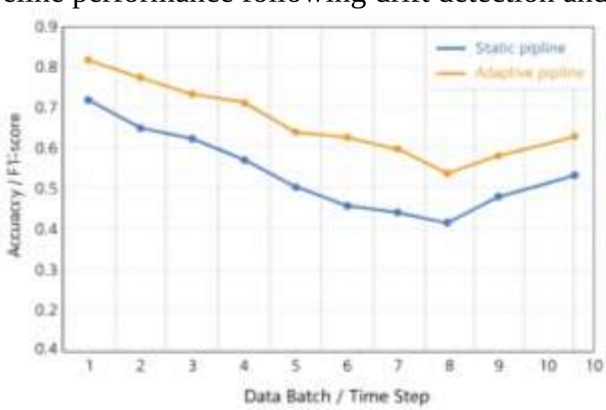

Figure 3. Predictive performance comparison between static and adaptive pipelines across streaming data batches

- <u>Decision Stability and Temporal Consistency</u>

In decision-making systems used in the real world, maintaining a steady decision-making process is just as important as overall accuracy. As drift accumulated, the static pipeline exhibited increasing volatility in prediction confidence and classification consistency.

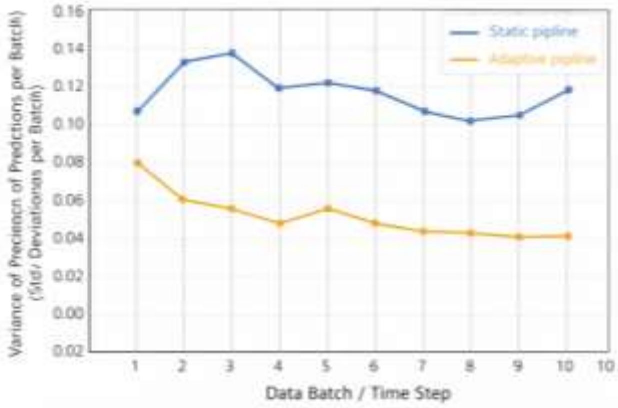

Figure 4. Temporal decision stability comparison between static and adaptive pipelines

The stability trends in Figure 4 highlight the importance of data-centric adaptation for maintaining reliable decision behavior over time. The adaptive system substantially reduced prediction volatility across successive batches. The confidence levels in predictions were comparatively more stable and there was a constant drop in the variance of predictions compared to the static system. Adaptive data refinement techniques made steps towards improving decision making reliability, a vital attribute for operational AI systems used in safety-critical and high-stakes applications. The reduction in prediction variance was also statistically significant ($p < 0.05$), further validating the stability advantage of the adaptive framework.

Table 3 outlines the improvements relative to a static pipeline that were observed from an adaptive data-centric approach relative to the average performance across all data sets.

Table 3. Average performance improvement (%) across datasets

Metric	Static	Adaptive	Improvement
Accuracy	0.52	0.63	+11%
F1-score	0.49	0.61	+12%
Stability (↓Var)	0.11	0.05	−55%

- <u>Impact of Concept Drift Detection and Data Refinement</u>

The framework successfully distinguished between gradual and sudden changes in distribution with the drift detection. Following the detection of drift, actions taken to adapt data refinement, such as sample reweighting and imbalance adjustment, resulted in substantial improvements in the performance of the downstream model.

Notably, the extent of performance recovery was associated with the level of the severity and persistence of the observed drift. Incremental data adaptation strategies show most benefits for gradual scenarios of drift, while for the more abrupt scenarios of drift more refined and aggressive approaches are required. The findings show that granular, data-centric strategies are able to effectively re-align the prevailing distributions with the learning decision boundaries of the model.

- Computational Efficiency and Operational Overhead

From a computational standpoint, the adaptive data-centric pipeline added little overhead to the static pipeline. The costs associated with data profiling, drift detection, and refinement were orders of magnitude lower than the costs of a fully retrained model. Across all trials, the memory and latency costs were consistently within an acceptable operational range.

In comparison to retraining-based baselines, the proposed method resulted in a significant reduction of computational costs while predictive performance remained the same, if not better. This clearly demonstrates the resource efficiency of data-centric adaptation approaches in scenarios where computational or time resources are limited.

Table 4. Computational efficiency comparison between static, adaptive, and retraining-based pipelines

Approach	Avg. Processing Time	Memory Overhead	Retraining Cost
Static	Low	Low	None
Adaptive	Moderate	Moderate	None
Retraining	High	High	High

Discussion

The findings confirm the core idea of the study- the potential that dynamically optimizing data in a smart way has to improve the performance of AI systems in changing situations. The adaptability that the framework provides is in contrast to traditional approaches that are model-centric, and provides evidence of the ability to achieve continuous performance with the proposed data adaptation methods, without the need to retrain models constantly.

In order to further clarify the individual contributions of components of the proposed framework, a qualitative ablation analysis was performed by turning off certain mechanisms. When prior drift detection was not applied, recovery of performance was both weaker and less stable over time. Likewise, refinement without drift detection was ineffective in mitigating distributional changes. The proposed method's justification comes from the grouping mechanism of profiling, drift detection, and refinement, not from the individual components.

One of the most important points was the noted improvement in the stability of the decision. Most previous research has focused on the metrics of accuracy. However, the framework's adaptable reduction in the volatility of predictions over time has improved a frequently critical but inadequate need in AI's operational hold. The primary objective of operational AI systems is to maintain stable and consistent decision outputs over time, particularly in decision support, monitoring, and autonomous systems where instability may undermine trust and reliability. There is a possibility that operational instability may compromise user confidence and even result in abandonment of the system.

The performance in a number of scenarios was found within the data distributions which were evolving in relation to a model aligned, rather than a limitation with the model. This finding is in conflict with the dominant belief that the only plausible solution for performance that is degraded to an unacceptable level is active retraining of the model.

In the context of data validity, the controlled experimental design offers value. With data revisions being the variable of focus, and model architecture and hyperparameters being consistent, the designer can pinpoint data revisions as the greatest contributor to performance enhancement. This speaks to a historical critique of the adaptive learning literature, wherein model and data changes occurred simultaneously, inhibiting causal attribution and the pinpointing of actual performance improvement drivers.

The computational efficiency results further support the practical viability of this approach. In the real world applications of the system, a retraining strategy may introduce unmanageable costs, lengthy latencies, operational disruptions, and complex system behaviors. For these reasons, retraining is not a viable operative solution. The framework introduces a more adaptive solution that supports various operational approaches that are computationally sustainable in long-term system deployment.

Some limitations must be recognized despite these benefits. The framework focuses on operational strength, but the effectiveness of the framework ultimately depends on the reliability of drift detection signals and the granularity of data profiling, which may vary across application domains. In some cases, environments with highly sudden or adversarial distribution changes may require other control measures or hybrid adaptive adjustments. This form of control is an important research area in the future.

The findings demonstrate that adaptive data management significantly enhances operational robustness under evolving data conditions. Collectively, these results reinforce the ongoing shift toward data-centric artificial intelligence, highlighting the need for active data governance and adaptive control rather than further algorithmic complexity. These findings demonstrate that robust, deployed AI systems require active data governance and adaptive control rather than being a problem of algorithmic optimization.

Conclusion and Future Work

Conclusion

Collectively, the findings indicate a significant step toward operationalizing adaptive data-centric artificial intelligence frameworks for dynamic environments. It demonstrates how decision-making systems can be designed to operate reliably in dynamic and non-stationary environments. The framework integrates adaptive mechanisms that respond to evolving data distributions while preserving model stability.

The proposed adaptive data-centric approach demonstrates that data-level mechanisms can mitigate performance deterioration due to concept drift, noise, and class imbalance, after thorough experimental evaluation on several real-world datasets. The approach also demonstrates that, unlike static data pipelines, predictive accuracy is maintained, decision outputs are stabilized, and generalization performance is enhanced during extended periods of operation. Furthermore, the proposed adaptable data-centric approach results in less model retraining which consequently reduces the cost and complexity of operation.

These findings imply that the cause of performance decay in deployed AI systems is due to the misalignment of the data and not the model. AI systems can continue to perform optimally even in the face of shifting data and evolving systems. These findings once again emphasize that data-centric optimization is a principal rather than an ancillary component for the sustainable deployment of AI systems.

The real-world example of the developed framework demonstrates that adaptive data-centric mechanisms enhance the resilience of a system by reducing the effects of distributional changes and maintaining operational stability in a system's adaptive working environment. This is the first evidence of adaptive data-centric systems intelligence as an effective approach to robust AI systems in changing environments. Overall, the findings shift the adaptation paradigm from reactive model retraining to proactive data governance, offering a scalable pathway for sustainable AI deployment.

Future Work

While the framework is practical and demonstrates overall solid performance and practical viability, there are multiple pathways to pursue. First, the focus of data-level adaptation in the current implementation may yield additional performance improvements. Future work may integrate hybrid strategies that combine data-level adaptation with selective model updates to address extreme or adversarial drift scenarios.

Extending adaptive data-driven methods to streams of multi-modal and/or high-dimensional data may be beneficial. For example, the integration of adaptive data-driven methods in multi-modal data frameworks may benefit the overall integration.

Thirdly, future work may focus on the automated policy learning for adaptive data refinement. Rather than using refinement strategies that have been predefined, reinforcement learning, or meta-learning may be used for dynamically refining or optimizing the strategies that are used for the adaptation of data based on the anticipated long-term outcome and the constraints of the process.

Integrating explainability and interpretability into data-centric adaptation processes remains an important open challenge. Being able to offer insight into the reasons for particular data refinements would foster trust, enhanced accountability, and ease of compliance to regulations for data-centric adaptive systems in practice.

Lastly, large-scale longitudinal evaluations in industrial and safety-critical environments are necessary to further validate the framework's effectiveness under real-world operational constraints.

Acknowledgements

The researcher did not receive any funding for this study, and the results have not been published in any other sources.

References

Aguiar, G., Krawczyk, B., & Cano, A. (2024). A survey on learning from imbalanced data streams: taxonomy, challenges, empirical study, and reproducible experimental framework. Machine Learning, 113(7), 4165–4243. https://doi.org/10.1007/s10994-023-06353-6

Azhar, M., Amjad, A., Dewi, D. A., & Kasim, S. (2025). A Systematic Review and Experimental Evaluation of Classical and Transformer-Based Models for Urdu Abstractive Text Summarization. *Information*, *16*(9). https://doi.org/10.3390/info16090784

Bokrantz, J., Subramaniyan, M., & Skoogh, A. (2024). Realising the promises of artificial intelligence in manufacturing by enhancing CRISP-DM. Production Planning and Control, 35(16), 2234–2254. https://doi.org/10.1080/09537287.2023.2234882

Carnero, A., Martín, C., Jeon, G., & Díaz, M. (2024). Online learning and continuous model upgrading with data streams through the Kafka-ML framework. Future Generation Computer Systems, 160(452), 251–263. https://doi.org/10.1016/j.future.2024.06.001

Goncalves, V. P. M., Silva, L. P., Nunes, F. L. S., Ferreira, J. E., & Araújo, L. V. (2023). Concept drift adaptation in video surveillance: a systematic review. Multimedia Tools and Applications 2023 83:4, 83(4), 9997–10037. https://doi.org/10.1007/s11042-023-15855-3

Greco, S., Vacchetti, B., Apiletti, D., & Cerquitelli, T. (2025). Unsupervised Concept Drift Detection From Deep Learning Representations in Real-Time. IEEE Transactions on Knowledge and Data Engineering, 37(10), 6232–6245. https://doi.org/10.1109/TKDE.2025.3593123

Han, J., Xia, T., Spathis, D., Bondareva, E., Brown, C., Chauhan, J., Dang, T., Grammenos, A., Hasthanasombat, A., Floto, A., Cicuta, P., & Mascolo, C. (2022). Sounds of COVID-19: exploring realistic performance of audio-based digital testing. Npj Digital Medicine 2022 5:1, 5(1), 16-. https://doi.org/10.1038/s41746-021-00553-x

Hsu, C. Y., Li, W., & Wang, S. (2025). Geospatial foundation models for image analysis: evaluating and enhancing NASA-IBM Prithvi's domain adaptability. International Journal of Geographical Information Science, 39(9), 2096–2125. https://doi.org/10.1080/13658816.2024.2397441

Khakimov, A., Elgendy, I. A., Muthanna, A., Mokrov, E., Samouylov, K., Maleh, Y., & El-Latif, A. A. A. (2022). Flexible architecture for deployment of edge computing applications. Simulation Modelling Practice and Theory, 114, 102402. https://doi.org/10.1016/j.simpat.2021.102402

Kumar, S., Datta, S., Singh, V., Singh, S. K., & Sharma, R. (2024). Opportunities and Challenges in Data-Centric AI. IEEE Access, 12, 33173–33189. https://doi.org/10.1109/ACCESS.2024.3369417

Megahed, F. M., Chen, Y. J., Jones-Farmer, L. A., Rigdon, S. E., Krzywinski, M., & Altman, N. (2024). Comparing classifier performance with baselines. Nature Methods, 21(4), 546–548. https://doi.org/10.1038/s41592-024-02234-5

Paleyes, A., Urma, R. G., & Lawrence, N. D. (2023). Challenges in Deploying Machine Learning: A Survey of Case Studies. ACM Computing Surveys, 55(6). https://doi.org/10.1145/3533378

Priya, S., & Uthra, R. A. (2021). Deep learning framework for handling concept drift and class imbalanced complex decision-making on streaming data. Complex & Intelligent Systems 2021 9:4, 9(4), 3499–3515. https://doi.org/10.1007/s40747-021-00456-0

Shahin, M., Chen, F. F., Maghanaki, M., Firouzranjbar, S., & Hosseinzadeh, A. (2024). Evaluating the fidelity of statistical forecasting and predictive intelligence by utilizing a stochastic dataset. The International Journal of Advanced Manufacturing Technology 2024 138:1, 138(1), 193–223. https://doi.org/10.1007/s00170-024-14505-8

Singh, P. (2023). Systematic review of data-centric approaches in artificial intelligence and machine learning. Data Science and Management, 6(3), 144–157. https://doi.org/10.1016/j.dsm.2023.06.001

Sinha, S., & Lee, Y. M. (2024). Challenges with developing and deploying AI models and applications in industrial systems. Discover Artificial Intelligence 2024 4:1, 4(1), 55-. https://doi.org/10.1007/s44163-024-00151-2

Wang, W., Zhu, Q., & Gao, H. (2023). Drift Detection of Intelligent Sensor Networks Deployment Based on Graph Stream. IEEE Transactions on Network Science and Engineering, 10(2), 1096–1106. https://doi.org/10.1109/TNSE.2022.3227909

Wen, K., Jiao, J., Zhao, K., Yin, X., Liu, Y., Gong, J., Li, C., & Hong, B. (2023). Rapid transient operation control method of natural gas pipeline networks based on user demand prediction. Energy, 264, 126093. https://doi.org/10.1016/j.energy.2022.126093

Zha, D., Bhat, Z. P., Lai, K. H., Yang, F., Jiang, Z., Zhong, S., & Hu, X. (2025). Data-centric Artificial Intelligence: A Survey. ACM Computing Surveys, 57(5), 129. https://doi.org/10.1145/3711118