Systems & Control Transactions

Original Research Article
Peer Reviewed Conference Proceeding
ESCAPE 36 - European Symposium on Computer Aided Process Engineering
Sheffield, UK. 21-24 June 2026

PSE PRESS

Exploiting Input-Space Separation in Kolmogorov–Arnold Networks to Prevent Catastrophic Forgetting in Industrial NIR Systems

Imam M. Iqbal*, Isabell Viedt, and Leon Urbas

Technische Universität Dresden, Process-to-Order Group, Dresden, Germany
* Corresponding Author: imam_mujahidin.iqbal@tu-dresden.de.

ABSTRACT

Near-infrared (NIR) sorting systems in waste sorting plants operate under multiple settings, creating distinct input–output relationships that challenge predictive modeling. Conventional neural networks, such as multilayer perceptron (MLP), often suffer from catastrophic forgetting under continual training, limiting reliability across settings. This study evaluates Kolmogorov–Arnold Networks (KAN) for continual regression modeling of multi-setting NIR systems. KAN assign nonlinear transformations to network edges using localized spline grids, enabling structural isolation between input regions. We introduce controlled input-space manipulations (shifting successive settings to adjacent or non-overlapping grid regions) and compare KAN performance with MLPs of comparable parameter count. We also examine single-input versus multi-input configurations to assess dimensionality effects. Results show that KANs with sufficient input-space separation maintain previously learned knowledge with perfect resistance to forgetting, whereas overlapping inputs induce interference. MLP forgetting depends on learning rate and training duration and cannot be fully avoided without additional methods. Single-input KANs achieve comparable accuracy to multi-input models in this system, suggesting limited benefit from additional inputs. These findings demonstrate that KAN's structural locality, combined with controlled input-space alignment, provides a practical and robust approach for continual learning in industrial NIR systems.

Keywords: Artificial Intelligence, Machine Learning, Industry 4.0, Modelling, Process Monitoring

INTRODUCTION

Near-infrared (NIR) sorting systems in waste sorting plants operate under multiple parameter settings to accommodate variations in material composition, throughput, and product quality. These setting changes induce distinct input–output relationships, making reliable output prediction across operating conditions essential for disturbance detection, quality monitoring, and advanced control applications.

Classical models for waste sorting equipment typically rely on linear or split-factor formulations with a small number of dominant inputs [1, 2]. While interpretable, their limited expressiveness restricts their ability to capture nonlinear and setting-dependent behavior. Data-driven models, including multilayer perceptron (MLP), have demonstrated improved predictive accuracy by learning nonlinear relationships directly from data [3].

In industrial practice, however, data from different operating settings often become available sequentially, while historical data may be inaccessible or impractical to reuse. Under such continual learning conditions, neural networks such as MLP are prone to catastrophic forgetting [4], where training on new settings degrades performance on previously learned ones [5]. This behavior limits their reliability for long-term deployment in adaptive industrial systems.

Kolmogorov–Arnold Networks (KAN) have recently been proposed as an alternative neural architecture that assigns nonlinear transformations to network edges rather than nodes. Owing to their localized, grid-based parameterization, KAN may reduce interference between learned representations when trained sequentially. Initial evidence from synthetic benchmarks suggests that KAN can better preserve previously learned functional components under certain continual learning scenarios [6]. Whether these properties extend to real-world industrial regression tasks remains an open question.

Motivated by these observations, this work investigates continual regression modeling for multi-setting NIR sorting systems using KAN. To explicitly examine the role of input-space structure in catastrophic forgetting. In addition, the study evaluates whether KAN can maintain predictive performance under reduced-input configurations, reflecting practical constraints in industrial modeling.

The contributions of this paper are summarized by the following research questions:

RQ1: How does input-space separation affect catastrophic forgetting in Kolmogorov–Arnold Networks under continual training?

RQ2: To what extent does KAN retain predictive performance on previously learned settings compared to MLP under identical sequential training protocols?

RQ3: Can KAN achieve comparable or superior prediction accuracy using a reduced set of input features compared to multi-input configurations?

METHOD

Problem Definition and Data Description

A NIR sorting unit is an optical-based separation system that exploits differences in infrared spectral absorption to distinguish between material types. During operation, the NIR unit periodically measures the spectral absorption, position, and surface area of materials transported on a conveyor belt. Based on the configured operating setting, air jets located at the end of the belt are actuated to divert materials into different output bins. As a result, the NIR sorting unit also functions as a material surface flow meter. Measurements from two subsequent units provides inflow and outflow of upstream unit.

The modeling task considered in this work is to predict the outflow of the NIR unit for a given inflow under different operating settings. Each flow stream consists of 11 material types. To limit scope and ensure clarity, this study focuses on predicting the outflow of a single target material, noting that the proposed methodology is directly extensible to other material fractions.

Inflow and outflow data were collected from the NIR unit and downstream processing equipment in an European waste sorting plant over approximately ten days equivalent of continuous operation. During this period, the system operated under two distinct settings. Setting A was active for the majority of the operating time, while Setting B was active for approximately three days. Measurements were recorded at a sampling interval of one minute, resulting in datasets of size 10, 187 and 3, 771 for Settings A (Dataset A) and B (Dataset B), respectively. The box plot of each dataset is presented in Figure 1.

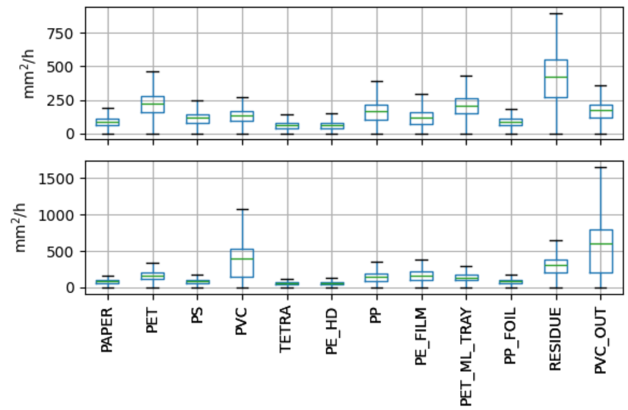

Figure 1. Flow Distribution in Dataset A (Top) and Dataset B (Bottom).

Continual Learning

To emulate a continual learning scenario, the model is trained sequentially on data from different operating settings. In the first phase (P_1), training is performed using Dataset A. In the second phase (P_2), the model is further trained on Dataset B without access to Dataset A. This setup reflects practical industrial constraints, where historical data may be unavailable when new operating conditions are introduced.

For each setting, the available data are randomly split into 70% for training, 15% for validation, and 15% for hold-out testing. The validation set is used exclusively for early stopping to mitigate overfitting with Patience of 20 steps and Max Steps of 500. Each training is repeated over five independent runs with different random seeds to account for stochastic variability in initialization and training. Model performance is evaluated on the hold-out test sets of all settings after each phase.

Prediction accuracy is quantified using coefficient of determination (R^2) for straightforward interpretation of predictive performance and forgetting behavior.

Kolmogorov Arnold Network

Kolmogorov Arnold Networks are a class of neural architectures inspired by the Kolmogorov-Arnold representation theorem, which states that any continuous multivariate function can be expressed as a finite sum of univariate functions composed with affine transformations. In contrast to MLP, where nonlinearity is applied at the nodes and edges are linear, KAN assign nonlinear transformations to the edges while nodes perform simple summation. See [6] for architectural illustrations.

Each edge in a KAN is parameterized by a learnable univariate function, typically represented using spline-based basis functions defined over a finite grid. These spline functions are locally supported, meaning that parameter updates primarily affect specific regions of the input space determined by the grid structure and spline order k. As a result, changes induced by new training data may remain confined to localized subsets of parameters.

This parameter locality contrasts with the globally shared weights in MLP and provides a structural mechanism by which KAN may reduce interference between sequentially learned tasks. In continual learning settings, such locality suggests that learning new operating conditions could have a limited impact on previously learned input-output mappings, motivating the investigation of KAN for multi-setting industrial regression problems.

Input-Space Manipulation Strategies

To leverage KAN's grid-based parameter locality in resisting catastrophic forgetting, this study introduces deliberate input-space manipulation strategies. The objective is to control how data from different operating settings occupy the input domain and, consequently, how they align with the spline grids of the KAN model. These manipulations are introduced solely as an experimental mechanism inspired by synthetic continual learning benchmarks and are not assumed to arise naturally in industrial NIR data.

Specifically, multiple input-shifting configurations are examined. In all shifted cases, the input features of Dataset B are translated by a fixed offset relative to Dataset A. Shifts of 2000, 4000, and 6000 mm^2/h are evaluated for spline orders k=1 and k=2, corresponding to different degrees of separation between the input regions associated with successive training phases. A baseline configuration with no input shift is also included for reference.

For all configurations, the grid width is chosen such that each dataset predominantly occupies a single grid region (i.e., 2000 mm^2 per region). Grid knot positions are fixed, while the grid range and number of knots vary according to the applied input shift. This ensures that changes in forgetting behavior can be attributed to controlled input-space separation rather than adaptive grid refinement.

By comparing these configurations, the study systematically evaluates how increasing degrees of input-space separation affect catastrophic forgetting behavior in KAN under continual training.

Experimental Design

Effect of Input-Space Separation in KAN

The combinations of input shift magnitude and spline order k, together with their corresponding grid configurations, are summarized in Table 1. Across all cases, a consistent KAN architecture and training configuration is employed to isolate the effect of input-space separation. Specifically, all models use a single-layer KAN with width of [11, 1], a zero base-function, and fixed spline scaling and bias terms (i.e., spline scale and bias are not trainable).

Table 1: Input Space Manipulation Cases and The Corresponding KAN architecture.

Shift (mm^2/h)	0	2000	4000	6000
Case k=1	K1S0	K1S1	K1S2	K1S3
Case k=2	-	K2S1	K2S2	K2S3
Grid	1	2	3	4
Max Grid	2000	4000	6000	8000

KAN vs MLP under Continual Training

To address RQ2, MLPs are trained under the same continual learning protocol as KANs and evaluated for resistance to catastrophic forgetting. Because MLPs do not exhibit inherent parameter locality, multiple MLP models are trained using different learning rates during the second training phase (L_{P_2}) to examine the effect of learning rate on forgetting behavior and to identify the best achievable performance under this constraint.

To ensure a fair comparison of model capacity, MLP architectures with approximately matched numbers of trainable parameters (n_{PAR}) are used. In particular, an MLP with structure of [11, 4, 1] is evaluated and compared against the KAN configuration exhibiting the strongest forgetting resilience identified in RQ1 (denoted KAN-K1S2). All MLPs are trained using the original input space, as input shifting is not expected to induce parameter modularity in standard MLP architectures.

Reduced-Input vs Multi-Input Configurations

To address RQ3, the predictive performance of KANs under reduced input dimensionality is examined. Two reduced-input KAN models, using a single dominant input and spline orders k=1 and k=2 (denoted KAN-K1RI and KAN-K2RI), are trained under the same continual learning protocol. Their performance is compared against the best-performing multi-input KAN models for the corresponding spline orders to assess whether comparable accuracy and forgetting resilience can be achieved with fewer input features.

Forgetting Metrics

Forgetting is quantified by measuring the change in predictive performance on the initial operating setting after training on a new setting. Specifically, the coefficient of determination R^2 on the test set of Dataset A is evaluated before and after Phase 2 training. The difference between these values is used as the primary metric.

Models exhibiting a strongly negative R^2 on Dataset A after Phase 2 are classified as having undergone catastrophic forgetting and are labeled as "Forgetful". Models maintaining near-identical R^2 values before and after

Table 2: Input Space Manipulation Results.

Case	K1S0	K1S1	K1S2	K1S3	K2S1	K2S2	K2S3
S_{P_1}	29±4	28±3	27±4	27±4	87±18	77±23	110±50
S_{P_2}	56±15	45±19	60±22	64±26	68±8	66±20	50±18
R^2_{A,P_2}	0.76 ±0.00	-2.99 ±0.10	0.93 ±0.00	0.93 ±0.00	-456.05 ±14.49	0.72 ±0.06	0.93 ±0.00
R^2_{B,P_2}	0.96 ±0.00	0.96 ±0.00	0.96 ±0.00	0.96 ±0.00	0.97 ±0.00	0.97 ±0.00	0.97 ±0.00
ΔR^2_A	-0.17 ±0.00	Forgetful	Resilient	Resilient	Forgetful	-0.21 ±0.06	Resilient

Phase 2 are classified as “Resilient”.

RESULTS AND DISCUSSIONS

Effect of Input-Space Separation

Table 2 summarizes the results of the input-space separation experiments. All configurations triggered early stopping in both Phase 1 and Phase 2, indicating stable convergence under the continual training protocol. Across all cases, predictive performance on the newly learned setting (Dataset B) remained high, with second-phase coefficients of determination (R^2_{B,P_2}) exceeding 96%. In contrast, performance retention on the previously learned setting (Dataset A) varied substantially across configurations.

Complete resilience to catastrophic forgetting is observed in cases K1S2, K1S3, and K2S3, where the coefficient of determination on Dataset A remains unchanged after Phase 2 training. In contrast, cases K1S1 and K2S1 exhibit complete forgetting, characterized by a severe degradation in R^2_{A,P_2}. The baseline cases K1S0 and K2S2 demonstrate partial resilience, retaining some predictive capability on Dataset A but with a noticeable reduction in performance.

These results indicate that catastrophic forgetting in KANs is strongly governed by the degree of overlap between input regions learned in successive training phases. When the input space of the new setting overlaps with, or is insufficiently separated from, the previously learned region, parameter updates propagate to shared spline basis functions, leading to interference and forgetting. Conversely, shifting the input space by at least k grid lengths from the boundary of the previously occupied region consistently prevents forgetting.

This behavior can be explained by the local support of spline basis functions in KAN. For a spline of order k, each spline basis function spans k+1 adjacent grid intervals. Consequently, basis functions located near grid boundaries influence neighboring regions outside the immediate input domain. If new training data falls within k grid intervals of a previously learned region, shared spline coefficients are updated, inducing forgetting. Ensuring a separation of at least k grid lengths guarantees that new data activates a disjoint set of spline basis functions, thereby preserving previously learned mappings.

These findings are consistent with observations reported by [7] and [8], who showed that architectures with localized or region-specific activations exhibit task-specific feature detectors. When the activation regions associated with different tasks do not overlap, earlier knowledge remains intact; when overlap occurs, forgetting increases proportionally to the extent of shared activation. In this context, KAN’s spline-based edge functions act as localized feature detectors whose spatial separation directly governs forgetting behavior.

KAN vs MLP under Continual Training

Table 3 compares the forgetting behavior of MLPs under varying second-phase learning rates with that of a KAN model of comparable parameter count. For MLPs, reducing the second-phase learning rate resulted in smaller ΔR^2_A values, indicating slower forgetting; however, this was accompanied by a decline in second-phase test performance on data B (R^2_{B,P_2}). Early stopping was not triggered for any MLP configuration, implying continued parameter updates throughout the second training phase.

Despite learning-rate tuning, none of the MLP configurations fully avoided catastrophic forgetting. Instead, forgetting was mitigated only by slowing the optimization process, confirming that forgetting in MLPs is governed by optimization duration over shared parameters.

KAN achieved the highest R^2_{B,P_2} while maintaining a comparable R^2_{A,P_2} to the best-performing MLP, with substantially lower result variability. This indicates that KAN’s improved forgetting resilience arises from its localized spline-based representations rather than from optimization hyperparameters.

It is acknowledged that several established techniques such as regularization-based methods, replay strategies, or parameter-isolation approaches can further improve forgetting resilience in MLPs. However, learning-rate control was deliberately selected here to isolate the effect of optimization dynamics alone, without introducing additional mechanisms that explicitly modify the network structure or training data. This ensures a fair comparison focused on intrinsic architectural differences between MLP and KAN.

Table 3: KAN vs MLP in Continual Learning Results.

Case	K1S2	MLP1	MLP2	MLP3
L_{P_2}	1	1x10^{-4}	2x10^{-5}	1x10^{-6}
n_{PAR}	44	53	53	53
S_{P_1}	27±4	323±167	323±167	323±167
S_{P_2}	60±22	500±0	500±0	500±0
R^2_{A,P_2}	0.93 ±0.00	0.54 ±0.05	0.89 ±0.02	0.93 ±0.00
R^2_{B,P_2}	0.96 ±0.00	0.95 ±0.00	0.92 ±0.01	0.90 ±0.00
ΔR^2_A	Resilient	-0.38 ±0.05	-0.04 ±0.02	Resilient

Input Dimensionality Analysis

Table 4 summarizes the input dimensionality analysis. Across both k=1 and k=2, multi-input KAN models consistently achieve slightly higher R2 values than their single-input counterparts; however, the performance gap remains marginal.

This indicates that interactions between multiple material inputs contribute only weakly to predictive performance in the studied setting. Single-input models with k=2 exhibit a small but consistent improvement over k=1, suggesting the presence of mild nonlinear correlations in the input–output relationship.

Given the negligible accuracy gain from additional inputs and higher model complexity, a single-input KAN with k=1 emerges as the most effective configuration, delivering comparable predictive accuracy with substantially fewer trainable parameters and reduced model complexity.

Table 4: Input Dimensionality Analysis Result.

Case	K1S2	K2S3	K1RI	K2RI
n_{INPUT}	11	11	1	1
n_{PAR}	44	66	4	6
S_{P_1}	27±4	110±50	20±0	22±1
S_{P_2}	60±22	50±18	26±7	22±2
R^2_{A,P_2}	0.93 ±0.00	0.93 ±0.00	0.92 ±0.00	0.92 ±0.00
R^2_{B,P_2}	0.96 ±0.00	0.97 ±0.00	0.96 ±0.00	0.97 ±0.00
ΔR^2_A	Resilient	Resilient	Resilient	Resilient

Practical Implications for Industrial Modeling

From a practical perspective, the proposed input-space shifting strategy offers a simple yet effective mechanism for maintaining predictive accuracy across multiple operating settings without requiring access to historical data.

The results of this study suggest that, when using KANs, deliberate alignment of new operating data to previously unused regions of the input space can substantially mitigate catastrophic forgetting. This enables long-term model reuse across settings using a single model instance, reducing maintenance effort and simplifying deployment. Moreover, the strategy is purely on data-level and does not require architectural modification or specialized continual learning algorithms, making it attractive for industrial adoption.

Limitations

Several limitations of the proposed approach should be acknowledged i.e.:

- The input-space manipulation strategy operates exclusively at the input level and explicitly exploits the localized grid structure of shallow KAN. As network depth increases, interactions between layers may reduce the degree of effective parameter locality, potentially diminishing the strength of input-space separation.
- The strategy relies on deliberate and controlled input transformations, which may not always align with naturally occurring process variations. While suitable for controlled modeling scenarios, its applicability to systems with complex, high-dimensional, or poorly bounded input spaces requires further investigation.
- Finally, the present study focuses on a single real-world system and a limited number of operating settings. While sufficient for mechanism validation, broader conclusions regarding generalization across domains should be drawn cautiously.

Future Work

The following are several directions for future research emerge from this work:

- Extending the experimental setup to deep KAN would allow assessment of forgetting resilience in deep KAN.
- Evaluating the proposed mechanism on other architectures that employ localized representations, such as radial basis function models, would help determine whether the observed behavior is specific to KANs or reflects a broader architectural principle.
- Embedding shallow KAN as base layer of MLP and applying input-space separation might enhance the forgetting resilience of the resulting complex

model.

- Combining input-space manipulation with complementary continual learning techniques, such as replay buffers or regularization-based methods, may further enhance robustness while maintaining feasibility.

CONCLUSION

Shifting input regions by at least k grid lengths prevent catastrophic forgetting in KAN, while overlapping or insufficiently separated regions cause interference in spline coefficients. This demonstrates that KAN's local-support spline functions can be systematically exploited to retain prior knowledge in continual training.

KAN maintains high accuracy on both old and new settings, whereas MLP forgetting depends on learning rate and training duration and cannot be fully avoided without additional techniques. This shows that KAN's structural locality is key to its superior continual learning performance.

Multi-input KAN models offer only marginal gains over single-input models, indicating minimal effect from multi-material interactions. Single-input KAN with k=1 achieves comparable accuracy with fewer parameters, making it the most efficient choice for practical deployment.

Together, the results demonstrate that Kolmogorov–Arnold Networks provide an effective and robust framework for continual learning in multi-setting NIR systems, highlighting its practical suitability for industrial deployment where both adaptability and model reliability are critical.

ACKNOWLEDGEMENTS

As part of the Federal Government's seventh energy research program 'Research for an environmentally friendly, reliable and affordable energy supply', research funding is provided by the BMWK and Project Management Jülich (PtJ) under the title of project EnSort (FKZ 03EN2083B).

REFERENCES

1. Kleinhans K, Hallemans M, Huysveld S, Thomassen G, Ragaert K, Van Geem KM, Roosen M, Mys N, Dewulf J, De Meester S. Development and application of a predictive modelling approach for household packaging waste flows in sorting facilities. Waste Management 120:290-302 (2021). https://doi.org/10.1016/j.wasman.2020.11.056
2. Tanguay-Rioux F, Legros R, Spreutels L. On the limits of empirical partition coefficients for modeling material recovery facility unit operations in municipal solid waste management. Journal of Cleaner Production 293:126016 (2021). https://doi.org/10.1016/j.jclepro.2021.126016
3. Kroell N, Maghmoumi A, Dietl T, Chen X, Küppers B, Scherling T, Feil A, Greiff K. Towards digital twins of waste sorting plants: developing data-driven process models of industrial-scale sensor-based sorting units by combining machine learning with near-infrared-based process monitoring. Resources, Conservation and Recycling 200:107257 (2024). https://doi.org/10.1016/j.resconrec.2023.107257
4. McCloskey M, Cohen NJ. Catastrophic interference in connectionist networks: the sequential learning problem. Psychology of Learning and Motivation :109-165 (1989). https://doi.org/10.1016/s0079-7421(08)60536-8
5. van de Ven GM, Soures N, & Kudithipudi D. Continual Learning and Catastrophic Forgetting (Version 1). arXiv (2024). https://doi.org/10.48550/ARXIV.2403.05175
6. Liu Z, Wang Y, Vaidya S, Ruehle F, Halverson J, Soljačić M, Hou TY, & Tegmark M. KAN: Kolmogorov-Arnold Networks (Version 5). arXiv (2024). https://doi.org/10.48550/ARXIV.2404.19756
7. Lee A, Gomes HM, Zhang Y, Kleijn WB. Kolmogorov-arnold networks still catastrophically forget but differently from MLP. AAAI 39:18053-18061 (2025). https://doi.org/10.1609/aaai.v39i17.33986
8. Rahman MM, Wang G, Zhou K, Chen M, & Yang F. Catastrophic Forgetting in Kolmogorov-Arnold Networks (Version 1). arXiv (2025). https://doi.org/10.48550/ARXIV.2511.12828

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See https://creativecommons.org/licenses/by-sa/4.0/