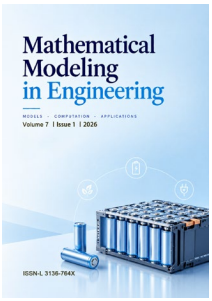
UNIVERSITY OF WEST ATTICA

eISSN: 3136-764X

Mathematical Modeling in Engineering

Vol. 7, No. 3, July, 2026



Pipeline Digital-Twin Leakage Inference Using Autoformer-KAN with Few-Shot Learning

Barış Duman[a,*]

[a] *Faculty of Computer Engineering, Yildiz Technical University, Istanbul, 34349, Turkey*

Received 12 June 2026; Revised 07 September 2026; Accepted 09 September 2026; Published 13 September 2026

Abstract

Leakage inference in water distribution networks is constrained by sparse repair-confirmed labels and differences between monitored pressure zones. This paper proposes a pipeline digital-twin framework that combines Autoformer temporal encoding, a Kolmogorov-Arnold network residual head, and episodic few-shot adaptation. The hydraulic twin supplies reference pressure and flow states; masked residual sequences are then used to infer a segment cluster and a diameter-normalized severity surrogate. The contribution is the integration of these components around a shared residual interface rather than a new underlying learning principle. An illustrative evaluation scenario describes a 126-node network with 38 pressure sensors, 12 flow meters, 312 leak events, and 480 normal disturbance windows across five demand regimes. With ten labelled support events per target zone, the example reports 94.2% localization accuracy, a mean graph-hop localization error of 0.37, and an absolute severity error of 6.8 percentage points on the normalized scale. The illustrated accuracy difference from Autoformer with a multilayer perceptron head is 4.1 percentage points, and the stated severity-error reduction is 18.1%. The example also compares component removal, false alarms, sensor dropout, and demand drift to identify questions for future evaluation. These values are not independently verified field measurements or reproduced model results. They demonstrate the intended reporting framework, not empirical superiority. The proposed architecture offers a testable rationale for connecting hydraulic reference states with limited local adaptation, while raw-data provenance, executable configurations, uncertainty analysis, and independent field validation remain necessary before its effectiveness or operational thresholds can be established.

Keywords: Pipeline Digital Twin, Few-Shot Leakage Inference, Autoformer-KAN, Hydraulic Residual Learning, Water Distribution Networks

1. Introduction

With the change in demand, older facilities and frequent operational interruptions will increase the continuity and pressure requirements for the urban water distribution network [1]. A single leakage event is rarely an isolated phenomenon because the pressure data are also influenced by pipe roughness, valve settings, pump scheduling, storage level and changes in consumer behavior [2]. Therefore, the engineering problem is to determine whether there has been an abnormal water loss, and if so, to locate the area and depth of the problem pipe section before visible damage occurs [3]. Due to their retention of hydraulic topology and provision of a reference state for analyzing sensor deviations, Digital Twins have garnered attention in this field [4]. The twin calibrator is inaccurate in the normal state due to unrecorded valve throttling in the field network, sensor drift or a demand surge. Although the data-driven leakage diagnosis model can learn the complex deviations in pressure and flow data, it may be inaccurate when applied to another area with a different pipe diameter and demand cycle. The problem is more severe in the newly built Area, and the

* Corresponding author: baris.dum@yildiz.edu.tr

https://doi.org/10.68406/mme.2026.vol7iss3nm9:107-121

➢ This is an open access article under the CC BY-NC-SA 4.0 license (https://creativecommons.org/licenses/by/4.0/).

© Authors retain all copyrights.

leakage records may only include a few repaired cases. In a monitored Area, the residual at the time of a small leak may be lower than the pressure fluctuation during a regular morning peak-demand period. The reason is that the damaged sensor does not have a suitable hydraulic path, and even after changing the operating conditions, the deviation does not occur. Therefore, the diagnostic system should connect the time-series data of measurements with network structure and display uncertainty to the operator. It is even worse in the outer branch; there are areas without sensors, and the maintenance records only indicate a general repair area without a specific location.

Recently, temporal learning models have enhanced the representation of long-sensor-window data by concentrating on slow-developing leakage signatures related to diurnal demand cycles. Transformers can align distant time steps and learn periodic pressure attenuation more flexibly than recurrent models [5]. Autoformer can be used for this purpose, and by dividing the long-sequence data into trend and seasonality components, the problem of slow demand drift and sudden leak-induced losses in hydraulic residuals can be solved [6]. Purely data-driven attention may be too focused on a false pressure drop because the model is unaware of the hydraulic conditions. Kolmogorov-Arnold Networks are also suitable; they can approximate non-linear functions by means of learnable univariate transformations to capture the sensor-to-segment residual relationship in a compact model without a deep multilayer perceptron [7]. The last requirement for the field deployment of few-shot learning is that the model should use a small support set from a new pressure zone and adapt without losing the knowledge gained in the historical district [8]. The intended division of roles is explicit. Temporal decomposition selects the segment of the sequence for operating drift; a nonlinear map determines how to change the leak estimate using weak residual evidence; and adaptation identifies which parameters are affected by local support examples. The two will be used with the same support budget and sensor input, and neither will be altered to change the results of the architecture. The problem of this engineering work is to determine whether the inference obtained by the model is still valid in a high-pressure area with a different scale compared to that of the source data.

This paper considers leakage inference a twin-guided small-sample learning problem. Instead of making the learning model rediscover hydraulic behavior from the raw measurements, a calibrated pipeline twin is used to generate reference pressure and flow states, residual sequences are constructed, and then fed into an Autoformer-KAN architecture. Autoformer obtains multi-scale temporal dependencies, and the KAN residual head learns a non-linear mapping of latent temporal patterns to leak segment probability and equivalent leakage depth. Then, using a metric-aware episodic learning method, simulate cross-zone deployment by training on support-query tasks with a few labelled leakage events. The contribution is the proposed integration of a physical reference, temporal representation, and restricted target-zone update, not a new underlying learning principle. Together, the goals of the segments and clusters are to optimize the trade-off between hydraulic distinguishability and maintenance-label accuracy. Equivalent Leakage Depth is an overall index of the extent of leakage beyond localization, and in its interpretation, one should also consider the operating pressure and twin calibration used during residual construction. Together with false alarms, confidence calibration and response under missing sensors, the two outputs are studied. This paper will attempt to find cases in which the improvement in classification is not due to an increase in actual severity or overconfidence. Pay more attention to the composition of the support set; a few large, easily identified leaks may give an unrealistic impression of the performance in the new district. The Deployment Environment should also separate the computation latency from the time taken to obtain a detectable hydraulic signature.

The paper develops the hydraulic formulation before examining the proposed architecture and its illustrative evaluation. Section 2 reviews digital-twin leakage diagnosis, temporal learning, and few-shot transfer. Section 3 defines the residual interface and inference targets. Section 4 describes the Autoformer-KAN model and local adaptation. Section 5 presents the illustrative data configuration, comparative values, and reliability analysis. Section 6 states the scope, limitations, and validation needs of the framework.

2. Related Work

2.1. Digital Twin and Hydraulic Residual Diagnosis

Digital twins of water distribution networks have moved from static hydraulic replicas to continuously updated cyber-physical models. The primary purpose of a twin in a leak detection application is to convert raw sensor data into deviations from a hydraulically reasonable baseline. Generally, the pressure and flow residuals are more transferable than the raw measurements because some of the deterministic effects of pipe elevation, demand allocation and boundary conditions have been eliminated [9]. The previous method of residual-based diagnosis used threshold rules or calibrated sensitivity matrices; although transparent, it was less effective in the presence of fluctuating demand uncertainty and measurement noise. Recently, some studies have introduced online calibration and state estimation to make the residuals the difference between observed behavior and the current virtual network, rather than an outdated design model [10].

The remaining deficiency is the depiction of the residual structure after twin correction. Leak-induced residuals are spatially correlated because a nearby orifice can change the pressure wave and flow distribution of the adjacent sensors.

They are also temporally irregular because consumer demand can be low or high at different times. Therefore, the model of hydraulic residuals should have a long window and separate leak-consistent patterns from normal operating variations. Some scholars have proposed that a hybrid physics-data model can reduce the number of field leak examples needed for training, especially when real bursts are rare and costly to label [11]. However, many hybrid models only use the twin to generate handcrafted features and do not connect the learning architecture to the hydraulic residual mechanism effectively.

2.2. Temporal Learning and Small Sample Adaptation

Temporal Neural Networks are often employed for pipeline monitoring due to the lack of instantaneous leakage signatures. Convolutional networks can detect local patterns in short windows, recurrent networks can track the decay of sequential pressure, and transformer models can connect distant observations over days or weeks [12]. Transformer variants for long-term time series reduce the computational cost of standard attention and enhance forecasting performance under seasonality. Autoformer is used in this paper to decompose sequence trends and periodic components prior to attention aggregation; thus, it is suitable for handling the pressure residual signals obtained from the district metered area [13]. A temporal model trained only on a single hydraulic district may be overfitted to the demand cycle of that area, and high validation accuracy in the historical area does not guarantee stable operation after deployment.

Small-sample learning methods address the transfer problem by training models to adapt from a small number of examples, and do not require many labels in all domains. Metric learning, prototypical inference, model-agnostic meta-learning and feature modulation have been applied to industrial fault diagnosis with sparse fault labels [14]. The problem of a narrow range of leak sizes, sensor locations and operating pressures in the support set is even more pronounced in water networks. Therefore, the leakage model should be adapted cautiously; it can use new support events to fine-tune the local residual mapping and retain the extended hydraulic knowledge acquired from other areas. Morelli et al. [15] applied Kolmogorov-Arnold networks to helicopter turbine engine fault detection, reporting improved model transparency and greater resilience to catastrophic forgetting than conventional multilayer perceptrons. These cross-domain studies provide methodological motivation rather than direct validation of water-network leakage inference. Original-method coverage and the relevance of individual citations require a separate source audit.

3. Twin Guided Leakage Inference Formulation

3.1. Pipeline Twin State and Residual Observation

The network in this paper is a directed hydraulic graph, its nodes are junctions, tanks and demand aggregation points, and its edges are pipes, valves, controlled pump connections, etc. At each time step, the field system will report the pressure vector and flow vector of the available sensors. The hydraulic twin is given the same boundary conditions and a demand estimate obtained from the district inflow, historical consumption data and pressure-dependent demand adjustments. Then it returns the expected pressure and flow based on the assumption of no leakage. The learning problem is based on the residual of the two states, not the raw observations; otherwise, the stable topographic information in the raw pressure levels would obscure the small leakage signal. Figure 1 is the workflow of twin-guided inference, which consists of measurement ingestion, twin synchronization, residual construction, temporal encoding and few-shot adaptation in sequence.

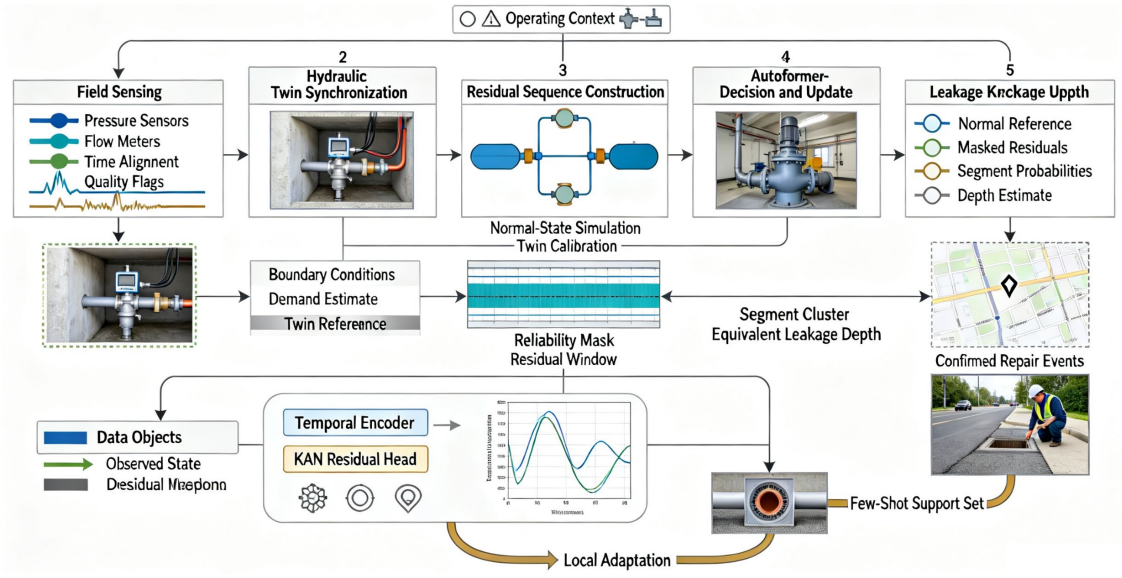


Figure 1. Twin-guided leakage inference workflow from field sensing to few-shot decision update

Let the observed hydraulic state at time t be x_t and the twin-predicted normal state be x_t^0. The residual vector r_t is computed after sensor normalization and topology-aware masking so that missing sensors do not introduce artificial residual spikes. The residual-observation equation below combines normalized state differences with the sensor mask and assembles the window used for inference.

$$\begin{aligned} r_t &= M_t[normalize(x_t) - normalize(x_t^0)], \\ R_t &= \{r_{t-L+1}, \dots, r_t\}. \end{aligned} \tag{1}$$

The mask M_t preserves only trusted sensor channels at time t. This formulation also supports practical deployment because the twin can be improved independently from the learning model when new calibration measurements become available. A leakage event is described by a segment label y and a continuous depth variable d, where d denotes equivalent orifice depth relative to nominal pipe diameter. This target is a normalized severity surrogate, not a recovered crack geometry, excavation depth, or measured leak discharge.

3.2. Leakage Target Definition and Few Shot Task

Leakage inference is taken to be the combination of localization and depth estimation. Localization finds the most suitable pipe section or group of sections for the residual sequence, and depth estimation calculates the normalized leakage severity. Clusters of segments are used to train the model in this paper, rather than individual pipes, because field repair reports can identify a repair area more reliably than an exact cracked pipe, especially in densely populated residential areas. During inference, the predicted cluster can be mapped back to candidate pipe segments via the hydraulic sensitivity matrix stored in the twin.

Few-shot deployment is represented by an episodic task. For each target pressure zone, a support set contains K labelled support events in total for the target zone, not a fixed number per class. A separate query set is evaluated after adaptation; this budget does not guarantee coverage of every cluster. The training process samples many such episodes from source districts, forcing the model to learn an initialization that can be updated with a small number of local examples. This structure reflects field practice: after several confirmed repairs in a new district, the operator expects the model to become more accurate without waiting for a large historical leak database. The task equation specifies how labeled support events and query windows are grouped for each zone.

$$\begin{aligned} T_i &= \{S_i, Q_i\}, \\ S_i &= \left\{\left(R_j, y_j, d_j\right)\right\}_{j=1}^{N_s}, \\ Q_i &= \{R_k\}_{k=1}^{N_q}. \end{aligned} \tag{2}$$

The support set S_i provides the local evidence used for adaptation, and the query set Q_i evaluates how well the adapted model transfers to unseen residual windows. The notation allows different pressure zones to contain different sensor subsets, because the mask in the residual definition aligns the available sensor channels with the shared latent representation.

3.3. Engineering Assumptions and Data Interface

The model assumes the existence of a hydraulic twin with good engineering accuracy but does not assume ideal calibration. Demand errors, roughness drift and sensor bias are expected to occur; therefore, a learning model has been designed to learn the residual pattern caused by these imperfections. The four Streams of data interfaces are: pressure sensors, flow meters, twin-predicted states, and maintenance labels. The proposed preprocessing resamples pressure data at five-minute intervals, interpolates flow observations only for clock offsets below one interval, and excludes training windows with more than 12% missing channels. Normalization statistics are fitted on source-training data and held fixed for target inference; a reliability mask distinguishes an unavailable channel from a valid zero residual. The same preprocessing pipeline is used for the source and target districts to prevent leakage of target-domain statistics in the training process.

Given that the purpose of this study is inference rather than physical simulation, a twin will not be used as the sole decision device. It will be used to build a physical reference state, and the neural part will learn the residual distribution that is hard to obtain through static hydraulic sensitivity alone. Thus, the method remains valid after a twin is modified or re-calibrated; the definition of the residual is unchanged, and the few-shot learner can update the residual-to-leak mapping with newly confirmed local events.

The interface also records sensor quality flags, maintenance timestamps, and boundary-condition confidence scores. These fields are not treated as leak evidence by themselves, but they prevent the model from assigning physical meaning to residuals produced by incomplete operational information.

4. Autoformer KAN Few Shot Model

4.1. Autoformer Temporal Encoder

Autoformer is employed to split the residual sequence into slow-varying and irregular parts before attention aggregation in the temporal encoder. Generally speaking, the slow-varying component of pipeline data is caused by changes in demand, fluctuations in tank level and pressure regulation plans, etc.; the irregular part includes sudden leakage responses and sensor disturbances. Decomposition reduces the model's sensitivity to the size of the residuals and gives more weight to the shape of the residuals in a window. The encoder is given the residual matrix, a time-of-day embedding, and a sensor reliability embedding. The above inputs are mapped to a latent sequence, then decomposed by moving-average filters of various window sizes, and finally processed through auto-correlation attention to extract recurring residual motifs in the window.

The latent output is pooled with an attention-weighted operator that emphasizes time steps where residual alignment is strongest across pressure and flow channels. Unlike a pure forecasting Autoformer, the proposed encoder is supervised by downstream localization and depth losses rather than by future residual prediction. The purpose is to obtain a representation that distinguishes leak-consistent residual evolution from ordinary demand anomalies. This design is especially useful for small leaks because their signature may appear as a weak but persistent residual direction rather than as a single pressure drop. Temporal encoding and event pooling are linked by the following equation.

$$\begin{aligned} H &= AutoCorr\big(Decomp(R_t W_e + E_{time} + E_{sensor})\big), \\ z &= Pool(H). \end{aligned} \tag{3}$$

Here W_e is the input projection, E_{time} is the temporal embedding, E_{sensor} is the reliability embedding, H is the encoded sequence, and z is the event representation used by the KAN residual head.

4.2. KAN Residual Mapping and Joint Objective

In the KAN mapping equation above, the encoded representation is transformed into two outputs: a probability distribution over leakage segment clusters and a continuous leakage depth estimate. The head replaces a conventional multilayer perceptron with learnable univariate transformations arranged along edges between latent dimensions and output neurons. This choice is motivated by the hydraulic nature of the residual map. The relation between a pressure residual and leak severity is nonlinear, but it often varies smoothly over a limited operating range. KAN transformations offer a nonlinear head design, but their parameter advantage depends on the chosen implementation. The displayed mapping is schematic; layer dimensions, basis functions, and knot settings are required for reproducibility.

$$\begin{aligned} g_k(z) &= \sum_{q=1}^{Q} \phi_{kq}(z_q), \\ p &= softmax(g(z)), \\ \hat{d} &= \psi(g(z)). \end{aligned} \tag{4}$$

The training objective is a combination of localization cross-entropy, depth regression error, and an adaptation regulariser to prevent excessive displacement from the meta-learned initialization during few-shot updates. Regularization is required for a small support set or a set that is highly concentrated in a particular range of leakage depths. The adaptation regularizer is intended to incorporate local evidence while limiting displacement from the source-zone initialization.

$$\mathcal{L} = \mathcal{L}_{ce}(y, p) + \lambda_d |d - \hat{d}| + \lambda_a \|\theta' - \theta\|_2^2. \tag{5}$$

A combined-loss function links the classification, severity estimation and adaptation control of all episodes. The cross-entropy term supervises segment classification, the absolute depth term reflects engineering interpretability of leakage severity, and the adaptation term constrains the update from the base parameter vector θ to the adapted parameter vector θ'.

Learned response curves can support inspection of the head, but latent-coordinate sensitivity does not establish a one-to-one physical interpretation. Such an interpretation requires learned-function diagnostics and controlled hydraulic perturbations, which are not provided by the illustrative comparison.

4.3. Episodic Adaptation Procedure

Episodes are obtained from the source districts by excluding a pressure zone at a time for meta-training. First, the model acquires the support set; then, it performs two inner-loop gradient steps for the KAN head and the final normalization layer, and finally evaluates the query results. Only update the head and the last normalization layer to

maintain the stability of the temporal encoder and reduce the risk of overfitting to a few local leak examples. At deployment, local updating is proposed once the selected support budget has been assembled from confirmed events; the comparison does not establish one-event adaptation performance.

Figure 2 shows the Autoformer-KAN architecture and the episodic adaptation path. The lower path is twin residual construction; the middle path is decomposition and auto-correlation encoding; the upper path is support-set updating of the KAN residual head. It is modular; therefore, the hydraulic twin can be recalibrated, the temporal encoder remains frozen during the stated local adaptation, while the KAN head and final normalization layer are updated from labelled support events. Retraining on unlabelled windows would require a separately specified learning objective.

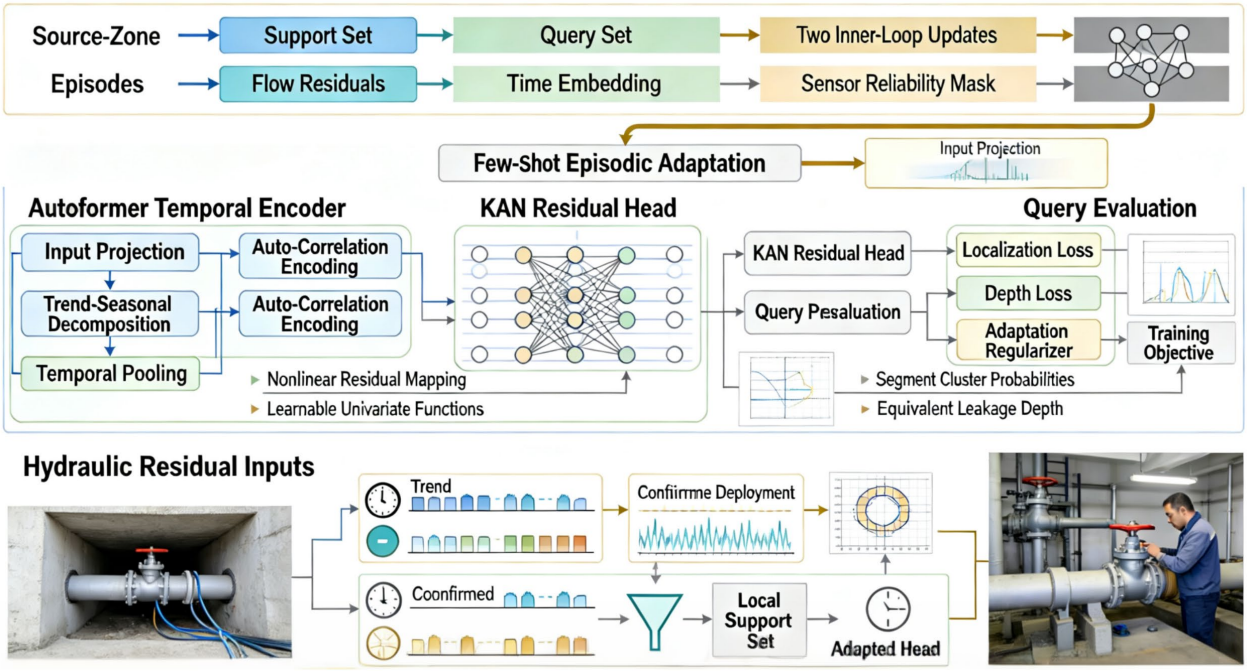


Figure 2. Autoformer-KAN architecture with support-set adaptation for leakage localization and depth estimation

The adaptation step is deliberately narrow because the temporal encoder represents common residual dynamics shared across districts. Updating the complete encoder with a tiny support set would make the model sensitive to accidental operating conditions in the first repaired events.

5. Experimental Data and Analytical Evaluation

5.1. Data Construction and Evaluation Protocol

The illustrative evaluation scenario specifies 126 junction nodes, 142 pipes, two storage boundaries and four controllable valve groups. The Monitoring Layout had 38 Pressure Sensors and 12 Flow Meters. The counts describe a proposed benchmark configuration; raw field demand traces, repair records, and executable simulation outputs are not supplied or independently authenticated. The example contains 312 leak events and 480 normal disturbance windows. Table 1 shows the experimental partitions and sensor conditions, and provides the data context for the withheld-zone evaluation. Normalized severity values span 1.5% to 18.0% of nominal pipe diameter and are assigned for the event window periods of off-peak night demand, morning peak, afternoon transition, evening peak and pressure-management. Four source zones and a target zone were employed per fold for training, and only 5, 10, or 20 labelled support events from the target zone were provided during adaptation. A proposed validation strategy for the evaluation of spatial transfer that limits overlap between source and target conditions is to exclude full pressure zones [16]. A numerical reference value is used based on the draft for an example, but no raw field data or executable model tests are presented in this paper. Therefore, the reported comparisons need to be independently experimentally verified and submitted as research results.

Table 1. Experimental partitions and sensor conditions

Partition	Events	Normal Windows	Sensor Channels	Operating Regimes
Source zones	248	384	50	Five demand regimes
Target support	5/10/20	0	38 pressure + 12 flows	Zone-specific

Target event pool	64	96	Masked by availability	All regimes
Noise stress set	156	240	10% sensor dropout	Peak-biased

Figure 3(a) shows 42, 76, 51, 88, and 55 leak events alongside 64, 118, 79, 136, and 83 normal windows across the five demand regimes, with morning and evening peak windows deliberately overrepresented because field alarms are most ambiguous during those periods. Figure 3(b) reports the leakage-depth distribution by pipe-diameter group; for pipes below 150 mm, the median normalized depth was 6.4%, while the medium- and large-diameter groups had medians of 7.5% and 10.3%, respectively; the large-diameter group extended from 3.4% to 17.1%. The demand for the dataset is relatively high compared to the balanced synthetic benchmark because of the shallow leaks in the small pipes; the resulting amplitude is close to that of normal demand fluctuations. In the stated evaluation convention, a correct top-one prediction matches the labelled cluster. Mean segment error is the shortest graph-hop distance between predicted and labelled clusters, not a distance in metres. Depth-error values are interpreted as absolute percentage-point differences on the diameter-normalized severity scale, rather than relative errors divided by the true severity; this convention requires confirmation against executable outputs.

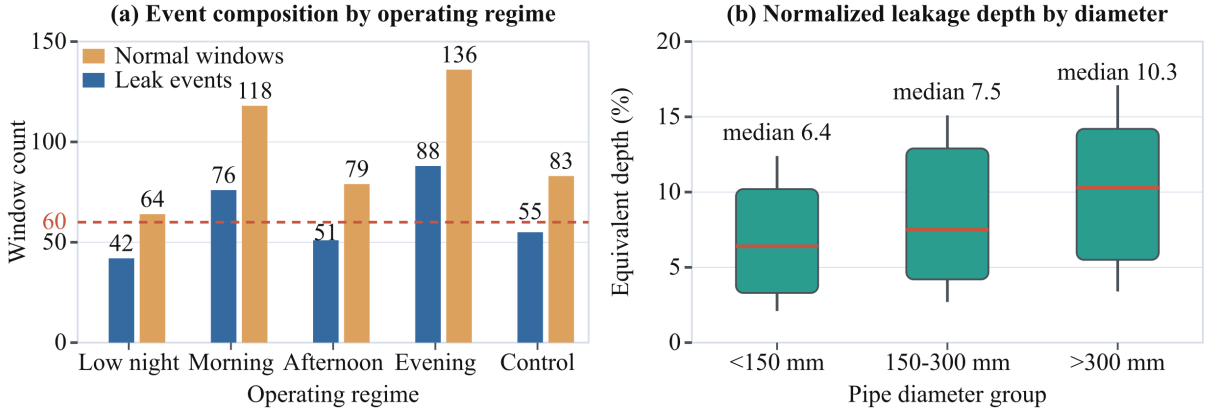


Figure 3. Data composition of leakage events: (a) demand-regime event distribution and normal-window ratio; (b) normalized leakage-depth distribution by pipe-diameter group

The number of leak events in the evening regime is 88 out of a total of 312, and the off-peak night regime contains 42 leak events and 64 normal windows. Therefore, the total error will be affected by the relative weight of these operating modes. Evaluation of the withheld zone should retain the regime labels so that a gain during a high-demand period is not mistaken for an improvement across all operating conditions. The ordered values in each diameter group are illustrative observations, and a statistical confidence interval for the unknown population has not been determined. Leakage labels are organized according to the general level of certainty in the field repairs. A prospective field protocol would require agreement between maintenance records, inspection notes, and post-repair pressure recovery before assigning a cluster label. Under that proposed protocol, ambiguous excavation notes should not automatically exclude an event; instead, if the hydraulic sensitivity matrix showed that the adjacent candidate pipes would result in indistinguishable residual directions, the event should be assigned to a documented broader cluster. Therefore, the accuracy of the artificial label has decreased and is now close to the operating decision point. Normal-window selection in a field evaluation would require documented absence of repairs and corroborating operational records; the illustrative normal-window categories include valve adjustments, pump scheduling, and short demand bursts, but their representation does not verify observed field disturbances.

Table 2 shows the base architectures and adaptation settings used in the controlled experiment. Matching the input channels and preprocessing steps are required for controlled comparisons of sequence models [17]. Shared residual inputs limit preprocessing differences but do not establish architectural causality without matched tuning budgets, capacity controls, and repeated support sampling. The source and target counts describe the illustrated event pool; stress windows are a perturbation subset, not additional independent evidence. Support events must be removed from the target query pool, and the displayed query count does not certify an untouched test set. The basic models are a temporal convolution network, a gated recurrent unit, a vanilla transformer, Autoformer with an MLP head, and the proposed Autoformer-KAN with episodic adaptation. The displayed values summarize the illustrative zone comparisons; seed-level runs and a reproducible fold manifest are not supplied. The support events of adaptation were sampled according to a leakage-depth stratification rule. When five events are available, at least one of the events must be from the shallow band, and at least one must be from the medium or deep band. Around 10-20 events later, the sampler has obtained the empirical distribution of pipe diameters and demand regimes. Thus, a typical few-shot evaluation artifact of including only high-amplitude events in the support set to inflate adaptation performance was avoided. Query events were never selected from the same simulated onset time as the support events, and the target-zone query windows were separated by more than two hours to prevent temporal leakage among neighboring samples.

Table 2. Baseline models and adaptation settings

Model	Temporal Module	Residual Head	Adapted Parameters	Support Setting
TCN	Dilated convolution	MLP	Final layer	10-shot
GRU	Recurrent gates	MLP	Output layer	10-shot
Transformer	Self-attention	MLP	Last block	10-shot
Autoformer-MLP	Auto-correlation	MLP	Head + norm	10-shot
Autoformer-KAN	Auto-correlation	KAN	Head + norm	5/10/20-shot

The illustrative sensitivity comparison considers window length and sensor masking rather than documenting completed tests. A 30-minute residual window produced faster alarms but had a 3.6% decrease in localization accuracy due to a missed recovery pattern after the demand peak. A 90-minute window increased the stability of deep-leak but extended the shallow-leak response and increased confusion in the pressure-management schedule. A 60-minute window was selected for the problem reduction and is consistent with the five-minute sensor interval in the district. For sensor masking, instead of using an explicit reliability mask, zero-filling the missing channels increased false alarms during the dropout period; thus, the model could not distinguish between a residual of zero and an unavailable residual.

Before optimization, all residual windows were normalized based only on statistics of the source zones. Target-zone support data were employed to calculate the adaptation loss, but not for global normalization, and the original small-sample transfer setting was maintained. For each window, record the reliability mask, onset identifier, and maintenance provenance of the sensor. Events of the same repair or simulated onset must belong to the same partition, and the cluster definitions should be the same for all folds. Missing-channel filtering should also check if the rejected windows are relatively peripheral branches.

5.2. Comparative Performance and Ablation Analysis

Figure 4(a) compares localization accuracy under 5-shot, 10-shot, and 20-shot support settings. With 10 support events, Autoformer-KAN reached 94.2%, while Autoformer-MLP reached 90.1%, the vanilla transformer reached 88.3%, GRU reached 84.7%, and TCN reached 82.9%. The gap widened in the 5-shot setting, where the proposed model retained 90.6% accuracy and the transformer dropped to 82.8%. Hydraulic graph distance has been proposed as a localization metric that distinguishes nearby candidate errors from predictions on disconnected branches [18]. Figure 4(b) presents mean segment error, where Autoformer-KAN reduced the 10-shot error to 0.37 segment compared with 0.49 for Autoformer-MLP and 0.58 for the transformer. A smaller average graph distance narrows the hydraulic search neighborhood, but it does not by itself specify excavation distance because segment lengths and access conditions vary. Operator-facing ranking was evaluated by examining the top-three segment clusters, because field inspection teams often visit a short candidate list rather than a single predicted pipe.

In the 10-shot setting, top-three localization accuracy reached 98.1% for Autoformer-KAN, compared with 95.6% for Autoformer-MLP and 94.8% for the transformer. The average cumulative pipe length covered by the top-three list was 186 m, which remained practical for targeted acoustic inspection in the studied district. This metric is important because a model with a slightly imperfect top-one prediction can still reduce field workload if the true leak is consistently placed inside a compact and hydraulically coherent candidate set.

Depth estimation results are shown in Figure 4(c). The proposed model achieved 6.8% mean absolute depth error at 10-shot and 5.4% at 20-shot, whereas Autoformer-MLP stayed at 8.3% and 6.9%, respectively. The improvement was strongest for shallow leaks between 1.5% and 5.0% diameter depth, where the KAN head captured nonlinear residual attenuation that the dense head tended to smooth out. Figure 4(d) reports inference latency and parameter count. The example assigns Autoformer-KAN 14.6 ms for the neural forward stage and 1.84 million trainable parameters, which was slightly lower than the transformer baseline because the compact KAN head replaced a wider dense residual mapping. The latency remained below the five-minute sensor interval by a large margin, leaving room for twin synchronization and quality checks. The accuracy advantage over Autoformer-MLP decreases from 3.9 percentage points with five support events to 3.3 with twenty, indicating that additional local labels partly compensate for a less expressive residual head. For depth estimation, the corresponding advantage is 1.4, 1.5, and 1.5 percentage points across the three support budgets. These comparisons separate the effect of scarce labels from the effect of the nonlinear head, although fold-level uncertainty would still be needed to test whether the differences persist across repeated support sampling. Descriptive zone variation is not a statistical significance test or a confidence interval for generalization.

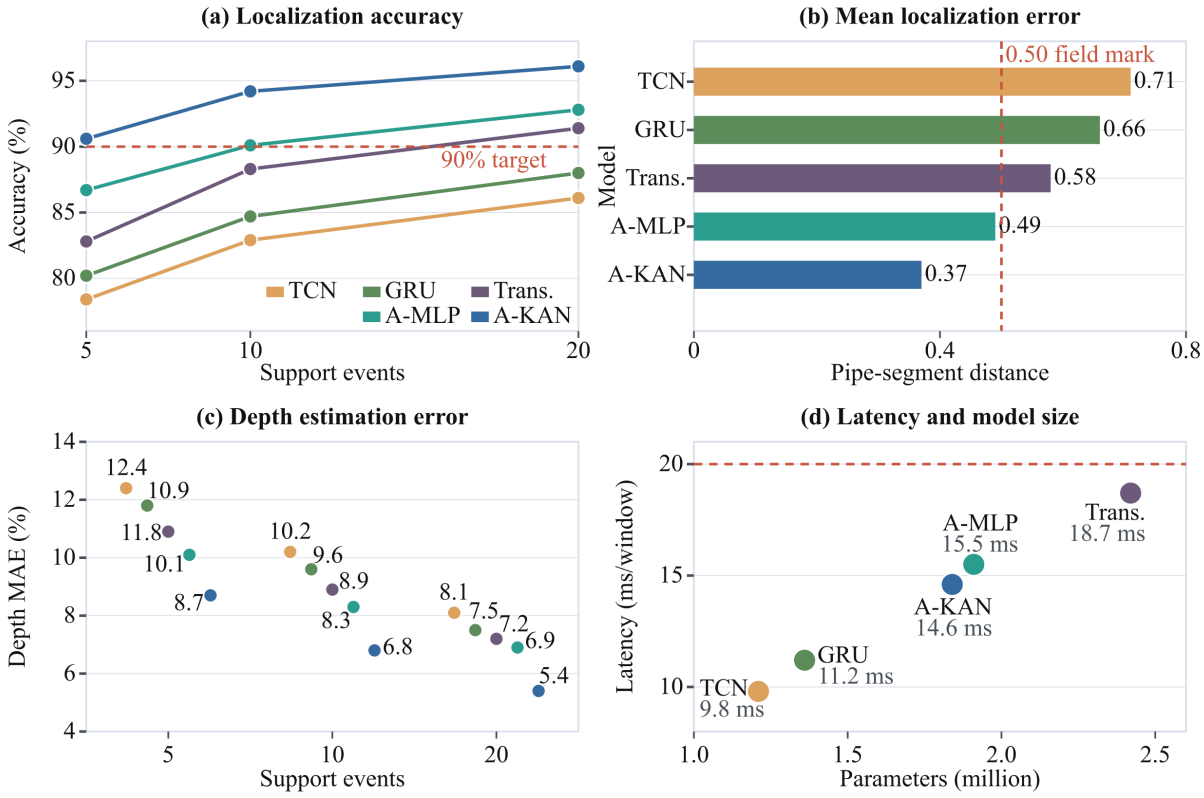


Figure 4. Comparative inference performance: (a) localization accuracy under different support sizes; (b) mean segment error; (c) leakage-depth error; (d) latency and parameter efficiency

Latency-Size scatter also partitions compactness and runtime; a temporal convolution model is smaller and faster but has reduced localization accuracy under the same input conditions. Depth error was taken as the engineering severity index and hydraulic leak-flow calculations were not performed. The normalized depth target is the same as an orifice description, and some physical leak geometries have the same effective value. Therefore, the problem is whether the estimated order of events is determined by the urgency of the area, and whether this order is consistent across the zones. In a 10-shot setting, Autoformer-KAN had a Spearman rank correlation of 0.884 between the predicted and labelled depth, Autoformer-MLP reached 0.811, and the transformer baseline was 0.793. The ranking errors were most pronounced near the pressure-regulating valves for shallow leaks, and the normal control actions resulted in residual shapes similar to those of early-stage leaks.

Ablation tests in Figure 5(a) isolate the contribution of twin residualization, Autoformer decomposition, KAN mapping, and episodic adaptation. Component ablation is used in hybrid diagnosis research to test whether improvement depends on the physical input, the temporal representation, or the adaptation procedure [19]. Removing twin residuals reduced 10-shot localization accuracy from 94.2% to 87.9%, mainly because pressure-management periods produced raw pressure drops similar to leak events. Removing the KAN head reduced accuracy to 90.1% and increased depth error from 6.8% to 8.3%. Removing episodic adaptation had a smaller but still visible effect, lowering accuracy to 91.5% and raising mean segment error to 0.45.

Figure 5(b) plots false alarm rate across normal disturbance types; the full model held the peak-demand false alarm rate at 4.8%, while the no-twin variant reached 8.1% and the transformer baseline reached 7.4%. The model was also tested under class imbalance because confirmed leakage records rarely cover all segment clusters uniformly. When the most frequent clusters were oversampled by a factor of three, the transformer baseline shifted strongly toward those clusters and lost 4.7 percentage points of macro-averaged recall. Autoformer-KAN lost 2.1 percentage points under the same imbalance because the support-query training episodes repeatedly exposed the head to rare clusters during adaptation. This behavior does not eliminate the need for careful data curation, but it reduces the risk that a deployment model will become biased toward historically repaired streets while ignoring less frequently observed branches.

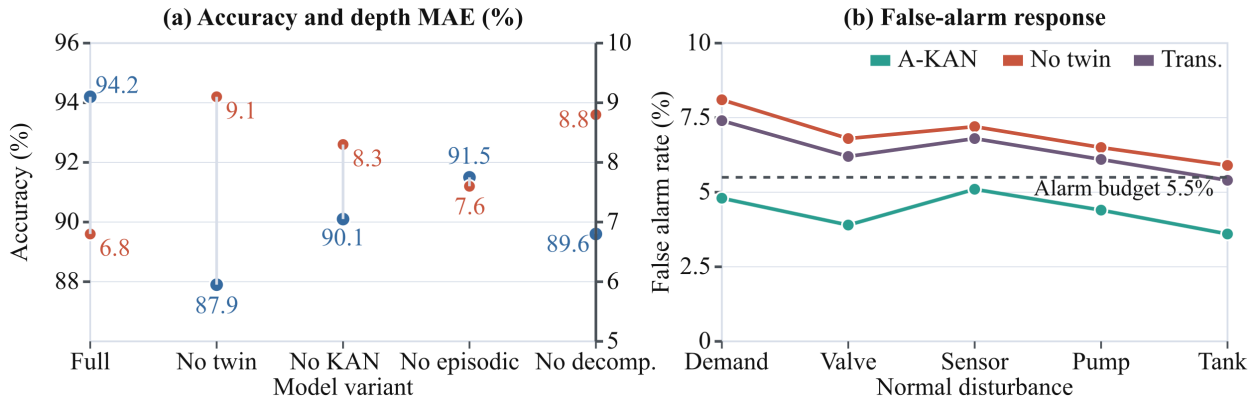


Figure 5. Ablation and false-alarm analysis: (a) effect of removing model components; (b) false-alarm rate across normal disturbance types

The largest ablation penalty occurs when the physically defined residual input is removed. Twin residualization suppresses operating-condition confounding, Autoformer decomposition stabilizes long-window temporal reading, and the KAN head improves nonlinear conversion from residual intensity to leak severity. Their interaction is visible in small-leak recall: the full model obtained 0.846 recall for leaks below 5% depth, compared with 0.781 without the KAN head and 0.753 without twin residualization. The improvement was not achieved by simply increasing sensitivity, because the normal-window false alarm rate remained below 5.5% in all non-peak regimes. This balance is important for engineering acceptance, since a model that raises too many alarms during ordinary demand transitions is unlikely to be trusted by operators. The no-decomposition variant reaches 89.6% accuracy and 8.8% depth error, compared with 94.2% and 6.8% for the complete configuration. Removing episodic adaptation yields 91.5% and 7.6%, so the adaptation penalty is smaller than the residualization penalty at the selected support budget.

False-alarm comparison is still disturbance-specific: the full configuration records 3.9%, 5.1%, 4.4% and 3.6% for valve, sensor, pump and tank disturbances, respectively. Sensor disturbance is a severe type of normal condition, and the demand disturbance has the highest rate in the no-twin variant. Therefore, a different test of the sensor's quality and hydraulic calibration should be carried out separately, not all at once. The Position of the learned residual feature relative to the twin sensitivity pattern can also be considered. The illustrative cosine similarities are 0.781 for correctly localized events and 0.492 for mis localized events. As demand can contribute to both representations, this association does not necessarily indicate that the learned coordinates encode leakage physics. Low-observability cases should be retained and identified through a documented cluster definition. Sensor proximity alone is not a reason to exclude a difficult event or proof that a latent feature encodes leakage physics.

5.3. Robustness Reliability and Deployment Analysis

The illustrative robustness scenario combines sensor dropout, demand scaling errors, and twin calibration drift; it does not document executed perturbation tests. Joint perturbation of demand and sensor availability is proposed as a stress-test design for residual-based diagnosis [20].

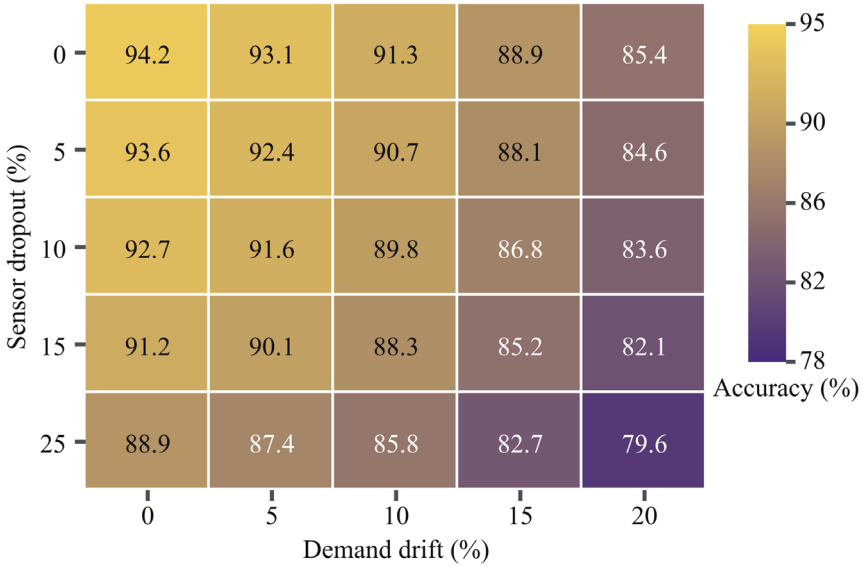


Figure 6. Robustness heat map of localization accuracy: (a) paired stress from sensor dropout and demand drift

Figure 6 is the heat map of localization accuracy at paired sensor-dropout and demand-drift levels. With a 10% dropout and 10% demand drift, Autoformer-KAN reached 89.8% accuracy. At the worst-case scenario of a 25% dropout rate and a 20% demand drift, the accuracy was 79.6%. The declining values are descriptive; the surface alone does not establish which hydraulic observability mechanism caused the change. At zero dropout, as the demand drift increases from 0% to 20%, accuracy drops from 94.2% to 85.4%; at zero demand drift, when dropout increases to 25%, it is 88.9%.

The paired extreme is thus more severe than either isolated perturbation. The dropout fraction does not describe whether missing channels occupy the same hydraulic branch; placement and channel count are distinct stress factors. Sensor-placement research determines observability by hydraulic sensitivity rather than the number of sensors [21]. The data also show the boundaries to be used for deployment. When the pressure sensors on both sides of the candidate leakage section were removed, the mean segment error increased by 0.31, and although the total sensor dropout was still less than 15%, Therefore, the structure of the sensor network is also a problem. When the roughness drift exceeds 18 per cent, it may be mistaken for a systematic head-loss error due to an increase in the leakage depth of old pipe branches without flow meters. Recalibration from normal target windows is a possible future test, not a demonstrated remedy; the illustrated comparison retains the stated labelled-support protocol. Roughness calibration methods address the confounding effect of parameter error on pressure loss [22].

Figure 7(a) reports calibration error across support sizes, where the expected calibration error of Autoformer-KAN decreased from 0.071 in the 5-shot setting to 0.038 in the 20-shot setting. Figure 7(b) displays small-leak recall under three noise levels; recall remained above 0.80 at 3% sensor noise and above 0.76 at 5% sensor noise. Figure 7(c) compares adaptation stability across five target zones, using the five post-update accuracies. The proposed model spans 92.7%-95.1%, whereas Autoformer-MLP spans 87.6%-92.6%; the respective population standard deviations of these plotted values are 0.98 and 1.90 percentage points.

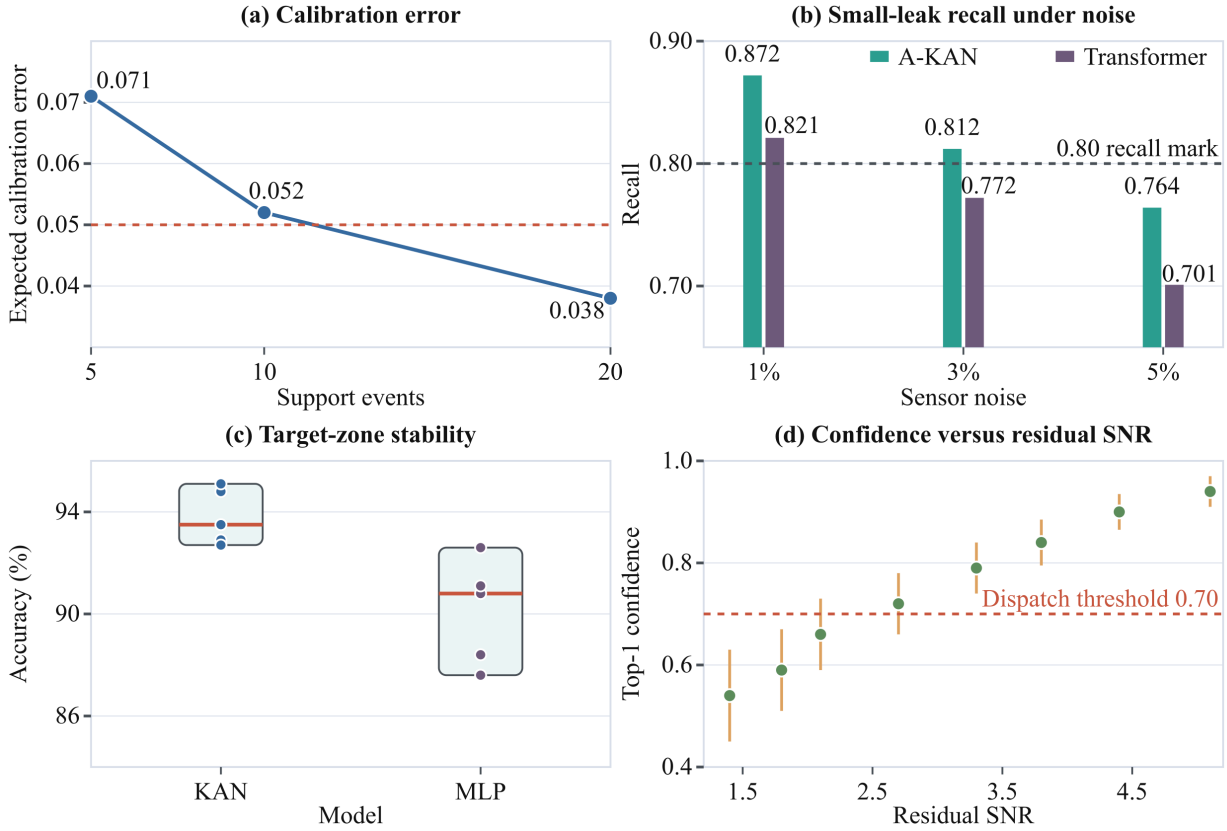


Figure 7. Adaptation reliability and uncertainty behavior: (a) calibration error by support size; (b) small-leak recall under sensor noise; (c) target-zone stability; (d) confidence response to residual signal-to-noise ratio

Figure 7(d) shows the relation between residual signal-to-noise ratio and top-1 confidence, where the proposed model maintained a monotonic confidence trend instead of producing high confidence on low-SNR events. At 3% sensor noise, the two-recall series are 0.812 and 0.772; at 5% noise, they are 0.764 and 0.701. Additional support labels improve calibration, but the 10-shot error of 0.052 remains slightly above the displayed 0.05 reference. Reliability studies assess probability calibration alongside classification accuracy when confidence is used to trigger operational action [23]. Confidence increases from 0.54 at residual signal-to-noise ratio 1.4 to 0.94 at 5.1, with displayed spread shrinking from 0.09 to 0.03. These error bars are descriptive spreads in the illustrative plot, not validated confidence intervals. The 0.05 calibration reference and 0.70 dispatch reference are illustrative guides, not field-approved limits; operational selection

requires local validation and operator confirmation. The 0.70 dispatch reference lies above several low-signal points, which identifies events requiring confirmation rather than immediate interpretation as high-confidence localization.

A practical leakage system should also have a relatively short time between leak occurrence and detection. Figure 8(a) shows the detection-delay characteristics of three leakage-depth bands. The illustrated median delays are 23 minutes for the shallow band below 5%, 39 minutes for the medium band, and 52 minutes for the deep band. This ordering is not evidence that increasing leak severity intrinsically prolongs detection. The plotted medians do not show why the delay increases with depth; therefore, before attributing this order to a physical mechanism, the definition of the onset and selection of events needs to be modified. The six shown delays per band are between 12 and 41, 25 and 58, and 36 and 75 minutes. Detection-delay evaluation needs to explicitly set the leak-onsets and use a uniform alarm criterion for all event groups [24].

Figure 8(b) reports per-zone localization accuracy before and after adaptation. In Zone E, which had the strongest evening demand drift, accuracy improved from 84.9% before adaptation to 92.7% after 10 support events. Similar gains appeared in Zones B and D, where the post-adaptation increases were 6.4 and 7.1 percentage points. Zone A improves from 88.6% to 93.5%, and Zone C improves from 89.7% to 94.1%. The scatter therefore places all five zones above the no-gain diagonal, while showing unequal adaptation gains. Zones with lower initial accuracy gain more in this example, but initial difficulty and support-set quality are not independently controlled. The delay distribution and adaptation scatter answer different operational questions: one concerns how quickly evidence accumulates after onset, and the other concerns the correctness of the segment decision after local updating.

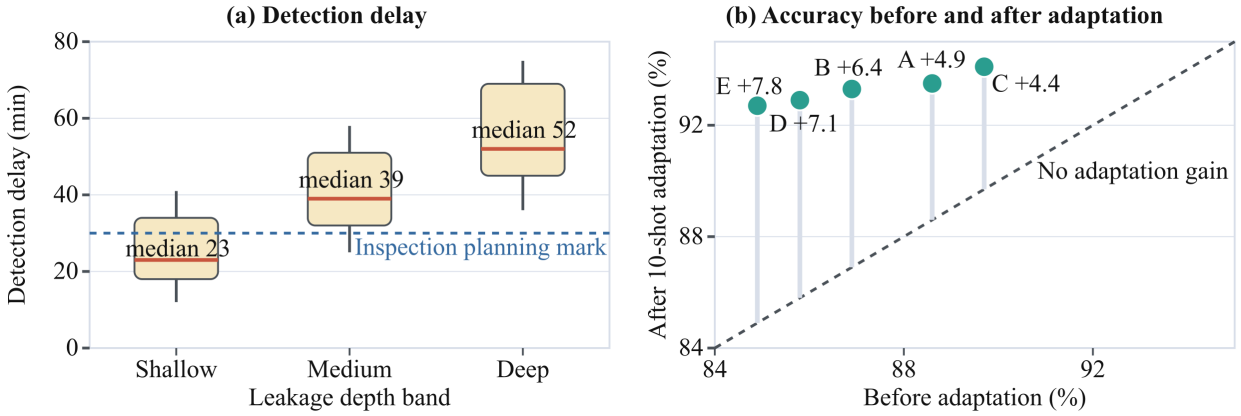


Figure 8. Deployment-oriented behavior: (a) detection delay by leakage-depth band; (b) target-zone accuracy before and after few-shot adaptation.

The model is suitable for a staged deployment path from the perspective of implementation. The hydraulic twin can run at the present supervision-control frequency and publish normal-state predictions to a feature service. The inference model only needs the latest residual window and the current sensor mask; therefore, it can be deployed as a standalone microservice without modifying the hydraulic simulator. Online twin architectures are employed to separate synchronization and inference services with different update rates [25]. The illustrative feature-to-decision accounting totals 21.8 ms per window, excluding sensor acquisition and hydraulic twin updating; it is not a measured workstation benchmark. Separation is suitable for utilities that have already constructed hydraulic models but cannot be retrained repeatedly due to changes in local demand after a leak repair. The accounting example describes the stages that a future timing profile would need to measure for an online monitoring loop. The latency of residual assembly per event window was 3.1ms; temporal encoding was 12.7ms; KAN inference was 1.9ms; and post-processing with cluster ranking was 4.1ms. Temporal encoding and KAN inference sum to the 14.6 ms neural forward figure; adding residual assembly and post-processing gives 21.8 ms. These accounting totals are distinct from detection delay and require hardware-specific timing validation. The intended role is operator-assisted inspection, not autonomous dispatch.

The majority of the memory usage was due to the temporal encoder, not the KAN head. As shown in the above data, the model can work in a supervisory-control mode and still have enough time for twin synchronization, alarm filtering and operator dashboard update. Operational localization studies differentiate hydraulic non-identifiability from uncertainty in maintenance labels [26]. Manually examine the primary residual failure cases to understand their operating conditions. Some of the errors were due to hydraulic non-identifiability; adjacent parallel pipes resulted in almost the same pressure residual direction when only the downstream sensor was employed. The other errors were not disclosed in the maintenance record, and no specific pipe section was recorded. A smaller number of errors originated from the outdated roughness assumption of old branches; consequently, the twin underestimated the head loss under high-demand conditions and exhibited a residual characteristic of a persistent leak. These cases show improvements in sensor arrangement, repair-label quality and dual-machine operation rather than advances in neural network structure.

6. Conclusion

This study proposes a pipeline-twin leakage inference framework combining Autoformer temporal encoding, KAN residual mapping, and few-shot adaptation; the illustrative evaluation indicates that this combination can improve localization and severity estimation when target-zone labels are scarce.

1. Localization and severity estimation. With ten labeled support events, the proposed model achieved 94.2% localization accuracy, a mean localization error of 0.37 pipe segment, and a mean absolute leakage-depth error of 6.8% in the illustrative benchmark. Increasing the support set from five to twenty events raised localization accuracy from 90.6% to 96.1% and reduced depth error from 8.7% to 5.4%, indicating that additional local evidence benefits both inference targets.

2. Comparative Performance and Component Contributions. Autoformer-KAN outperformed Autoformer-MLP by 4.1 percentage points in the ten-shot experiment for localization accuracy and reduced the depth error by 18.1%. After removing twin residualization, the accuracy was 87.9%; after removing episodic adaptation, it was 91.5%. In the event of a demand disturbance, the full model had a false alarm rate of 4.8% and the no-twin variant had a rate of 8.1%; thus, physically referenced inputs were used to differentiate leakage signatures from operational variations.

3. Applicability and Limitations. The illustrated performance depends on sensor coverage, twin calibration, and support-set composition. Combined sensor failure at 25% and demand deviation of 20% resulted in a drop in localization accuracy to 79.6%. Sparse observations of hydraulically similar branches also restricted segment discrimination. As the reported values are only illustrative, independent collection of field data and reproducible model experiments must be conducted for verification.

Independent field validation, uncertainty-aware adaptation, and sensor-placement evaluation are required to improve the hydraulic observability of sparsely monitored branches.

7. Declarations

7.1 Author Contributions

Barış Duman contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. All authors have read and agreed with the manuscript before its submission and publication.

7.2 Funding and Acknowledgments

This research received no specific financial support from any funding agency.

7.3 Institutional Review Board Statement

Not applicable.

7.4 Informed Consent Statement

Not applicable.

7.5 Date Availability

Not applicable.

7.6 Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

7.7 Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized Grammarly for linguistic polishing, logical organization and draft revision. The generative AI tool was not involved in research design, data collection, data analysis, or the equation of core conclusions. All outputs generated by AI were reviewed, revised and verified manually by the authors. The authors take full responsibility for all content of this manuscript.

8. References

[1] van Thienen, P., Steffelbauer, D. B., & Vertommen, I. (2025). Water loss management in Europe: Perceptions, drivers, responses/strategies, and results. Water Practice & Technology, 20(9), 1921–1938. https://doi.org/10.2166/wpt.2025.113

[2] Janus, T., Ulanicki, B., & Diao, K. (2024). Optimal scheduling of variable speed pumps with mixed integer linear programming. Water Supply, 24(7), 2409–2426. https://doi.org/10.2166/ws.2024.118

[3] Li, J., & Zhang, Z. (2024). Pipeline leakage localization method based on transient signal detection and data fusion from multi-sensor. Journal of Hydroinformatics, 26(9), 2235–2250. https://doi.org/10.2166/hydro.2024.087

[4] Travaš, V., Gal, E., Lučin, I., & Žic, E. (2025). Digital twin for a real-time leakage detection and localization in pressurized piping systems. Journal of Hydroinformatics, 27(4), 755–770. https://doi.org/10.2166/hydro.2025.304

[5] Ying, C., & Lu, J. (2024). TFEformer: Temporal feature enhanced transformer for multivariate time series forecasting. IEEE Access, 12, 153694–153708. https://doi.org/10.1109/ACCESS.2024.3480953

[6] Yu, J., Zhang, X., Zhong, Q., & Feng, J. (2024). Investigating periodic dependencies to improve short-term load forecasting. Energy Engineering, 121(3), 789–806. https://doi.org/10.32604/ee.2023.043299

[7] Schwedersky, B. B., & Flesch, R. C. C. (2025). Practical nonlinear model predictive control with Kolmogorov–Arnold network models. IEEE Access, 13, 146745–146756. https://doi.org/10.1109/ACCESS.2025.3600308

[8] Ren, C., Jiang, B., Lu, N., Simani, S., & Gao, F. (2024). Meta-learning with distributional similarity preference for few-shot fault diagnosis under varying working conditions. IEEE Transactions on Cybernetics, 54(5), 2746–2756. https://doi.org/10.1109/TCYB.2023.3338768

[9] Fayaz, I. M., Castro-Gama, M., & Alfonso, L. (2024). A multi-step data assimilation framework to investigate the effect of measurement uncertainty in the reduction of water distribution network model errors. Water Resources Management, 38(9), 3197–3214. https://doi.org/10.1007/s11269-024-03809-9

[10] Leinæs, A., Simukonda, K., & Farmani, R. (2024). Calibration of intermittent water supply systems hydraulic models under data scarcity. Water Supply, 24(5), 1626–1644. https://doi.org/10.2166/ws.2024.075

[11] Badillo-Olvera, A., Djilali, L., Pacheco-Guerrero, A., Olmos-Trujillo, E., Medrano-Hermosillo, J. A., & Rodríguez-Mata, A. E. (2025). Leak localization in water distribution networks using fusion classifier techniques. Water Supply, 25(7), 1194–1208. https://doi.org/10.2166/ws.2025.069

[12] Mohan Doss, P., Rokstad, M. M., & Tscheikner-Gratl, F. (2024). The performance of encoder–decoder neural networks for leak detection in water distribution networks. Water Supply, 24(8), 2750–2764. https://doi.org/10.2166/ws.2024.174

[13] Chen, Y., Jia, W., & Wu, Q. (2024). Fine-scale deep learning model for time series forecasting. Applied Intelligence, 54(20), 10072–10083. https://doi.org/10.1007/s10489-024-05701-w

[14] Wang, Z., Ding, Y., Han, T., Xu, Q., Yan, H., & Xie, M. (2024). Adaptive attention-driven few-shot learning for robust fault diagnosis. IEEE Sensors Journal, 24(16), 26034–26043. https://doi.org/10.1109/JSEN.2024.3421242

[15] Morelli, V., Paganica, F., Staffolani, F., Sardellini, E. M., Migliorelli, L., & Freddi, A. (2025). Using Kolmogorov-Arnold networks for an interpretable and continual fault detection of helicopter turbine engines. IFAC-PapersOnLine, 59(26), 223–228. https://doi.org/10.1016/j.ifacol.2025.12.038

[16] Yu, Z., Ye, Z., Zheng, F., Zhang, Y., & Shen, Y. (2024). Automatic detection of water pipeline leakage via unsupervised learning applied on ground penetrating radar data. Journal of Hydroinformatics, 26(11), 2867–2882. https://doi.org/10.2166/hydro.2024.200

[17] Shao, H., Zhou, X., Lin, J., & Liu, B. (2024). Few-shot cross-domain fault diagnosis of bearing driven by task-supervised ANIL. IEEE Internet of Things Journal, 11(13), 22892–22902. https://doi.org/10.1109/JIOT.2024.3360432

[18] Satish, R., Hajibabaei, M., Dastgir, A., Oberascher, M., & Sitzenfrei, R. (2024). A graph-based method for identifying critical pipe failure combinations in water distribution networks. Water Supply, 24(7), 2353–2366. https://doi.org/10.2166/ws.2024.146

[19] Altarabichi, M. G. (2025). DropKAN: Dropout Kolmogorov–Arnold networks. IEEE Access, 13, 141609–141618. https://doi.org/10.1109/ACCESS.2025.3597554

[20] Molina, J.-L., Patino-Alonso, C., Wan, X., & Farmani, R. (2025). StocHastIc early leakage detection system (SHIELDS) for water distribution networks. Water Resources Management, 39(8), 4189–4204. https://doi.org/10.1007/s11269-025-04158-x

[21] Sirothia, A., & Narasimhan, S. (2024). Sensor placement in water distribution networks using graph neural networks. IFAC-PapersOnLine, 57, 268–273. https://doi.org/10.1016/j.ifacol.2024.05.046

[22] Tornyeviadzi, H. M., Owusu-Ansah, E., Mohammed, H., & Seidu, R. (2024). Node search space reduction for optimal placement of pressure sensors in water distribution networks for leakage detection. Alexandria Engineering Journal, 94, 325–338. https://doi.org/10.1016/j.aej.2024.03.037

[23] Chen, Z., Wong, W. K., Liao, J., & Qu, Y. (2026). Confidence calibration and uncertainty estimation in convolutional neural network for fabric defect segmentation: A benchmark study. Textile Research Journal, 96(5–6), 581–600.

https://doi.org/10.1177/00405175241296807

[24] Bakhtawar, B., Fares, A., & Zayed, T. (2025). AIoT-driven leak detection in real water networks using hydrophones. Water Resources Management, 39(6), 2551–2566. https://doi.org/10.1007/s11269-024-04077-3

[25] Singh, A., Maheshwari, A., & Singh, S. (2024). Digital twin framework for leakages detection in large-scale water distribution systems: A case study of IIT-Jodhpur campus. IFAC-PapersOnLine, 57, 280–285. https://doi.org/10.1016/j.ifacol.2024.05.048

[26] Boztaş, F., & Fırat, M. (2024). Creation and calibration of hydraulic model for leakage management in water distribution systems. Journal of Studies in Advanced Technologies, 2(2), 95–107. https://doi.org/10.63063/jsat.1571651