Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Incremental Fuzzy-Rough Three-Way Decision with Adaptive Boundary Learning for Concept-Drifting Data Streams

**Mo Ateeb Ansari**

ateebansari2526@gmail.com

Uttaranchal University

**Aneesa Nawaz**

Fast National University of Computer and Emerging science

**Article**

**Keywords:** Fuzzy rough sets, data stream mining, incremental learning, vaguely quantified rough sets, selective classification

**Posted Date:** November 10th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11253142/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# Incremental Fuzzy-Rough Three-Way Decision with Adaptive Boundary Learning for Concept-Drifting Data Streams

Mo Ateeb Ansari[1*], Aneesa Nawaz[2]

[1]Department of Computer Science & Engineering, Uttaranchal University, 248007, India

[2]Department of Science and Humanities, Fast National University of Computer and Emerging science, Karachi, 75300, Pakistan

***Address corresponds to:** Mr. Mo Ateeb Ansari, Assistant Professor, Department of Computer Science & Engineering, Uttaranchal University, 248007, India

**E-mail:** ateebansari2526@gmail.com

## Abstract

Learning from data streams is challenging because the data distribution may change during model deployment, making previously reliable decision rules less dependable. At the same time, observations may be vague or fall within overlapping class regions, where forcing binary decisions can obscure decision reliability. Recomputing fuzzy rough models from the complete history is also computationally expensive. This paper develops IFR-3WD-CD, an incremental fuzzy rough three-way decision framework for concept-drifting data streams that addresses these challenges jointly. A vaguely quantified fuzzy rough membership is maintained over a bounded sliding window with online standardization to provide a noise-tolerant decision score. A decision-theoretic three-way rule assigns instances to accept, defer, or reject regions, while a drift adaptation layer separates detection from adaptation. A change detector triggers selective forgetting, and a bounded drift-intensity signal adaptively widens the deferral region during unstable periods. The resulting per-instance computational cost is independent of the elapsed stream length, avoiding repeated reconstruction from historical data. The framework is evaluated using a strictly chronological prequential protocol on four synthetic benchmarks representing abrupt, gradual, recurring, and incremental drift, and two real-world streams, against eight adaptive baselines. Results show competitive predictive accuracy while deferring a modest fraction of difficult instances, with committed decisions substantially more accurate than forced decisions. Ablation experiments confirm the contribution of individual components, while boundary-region analysis

demonstrates expansion after drift and contraction during recovery. The study demonstrates the utility of uncertainty-aware three-way decisions for adaptive learning under evolving data distributions.

**Keywords:** Fuzzy rough sets; data stream mining; incremental learning; vaguely quantified rough sets; selective classification

## 1. Introduction

Most supervised learning methods are designed for a fixed population: the finished model is applied to further data assumed to come from the same distribution. A growing number of applications violate this assumption because their data arrive as an unbounded, time-ordered stream and the process generating them evolves. Writing the stream as a sequence of labelled observations $\mathcal{S} = \{(x_t, y_t)\}_{t\geq 1}$ with $x_t \in \mathbb{R}^d$ and $y_t \in \mathcal{Y}$, the defining feature of this setting is that the governing distribution is time dependent, $(x_t, y_t) \sim P_t(X, Y)$, and that $P_t$ need not equal $P_{t+\Delta}$. This phenomenon, concept drift, is now recognised as the central obstacle to reliable stream learning, and recent surveys document both its many forms and the breadth of methods proposed to cope with it (Agrahari & Singh, 2022; Suárez-Cetrulo et al., 2023; Yang et al., 2026). A model that is accurate at time $t$ can degrade sharply at $t + \Delta t$ once the concept has moved, so any stream learner must contain some mechanism for keeping its internal reference current.

The transition from static to streaming data changes the problem in two ways that are usually treated separately but are, in practice, intertwined. The first concerns the arithmetic of updating. If the model summarizes the data through quantities that depend on the entire history, then each new observation in principle invalidates those quantities, and recomputing them from scratch becomes more expensive as the stream grows. A stream learner must therefore update its state at a cost that does not grow with the elapsed time, which rules out naive periodic reconstruction and motivates truly incremental maintenance (Zhang et al., 2025; Shu et al., 2024). The second concerns the quality of the evidence itself. Streaming observations are frequently noisy, and near a shifting decision boundary the classes overlap, so for many instances the available evidence simply does not justify a confident label. A learner that is nonetheless obliged to return a positive or negative verdict for every instance conflates two very different situations, the confidently classified and the truly ambiguous, and offers the downstream user no signal to distinguish them.

The second difficulty is exactly the concern of three-way decision. Rather than trisecting the universe into only acceptance and rejection, three-way decision admits a third option, a boundary or deferral region reserved for cases in which committing to either side would be premature; the accepted and rejected cases are acted on immediately, while the deferred cases are set aside for further information or review (Yao, 2023). The boundary region is not an error category but a deliberate decision outcome, and its cost is weighed against the costs of the two possible mistakes through a loss structure, which is the viewpoint of decision-theoretic rough sets and its cost-sensitive extensions (Qian et al., 2022; Liang et al., 2024). In a non-stationary stream this third option is especially natural: immediately after a change, when the model's reference still reflects the previous concept, deferring the instances it can no longer place confidently is a more defensible response than forcing them into a brittle binary split.

To defer intelligently a learner needs a graded notion of how well an instance fits each class, which is where fuzzy rough sets contribute. Fuzzy rough approximation combines a fuzzy relation that measures the similarity of instances with lower and upper approximations that quantify, respectively, the degree to which an instance certainly and possibly belongs to a concept; the gap between them is a direct measure of local uncertainty (An et al., 2021; Ji et al., 2025). Because real streams are noisy, the classical approximations based on the strict infimum and supremum are too brittle to use directly, and robustified formulations based on ordered weighted aggregation and fuzzy quantifiers have been developed precisely to tolerate mislabelled or outlying instances (Palangetić et al., 2022; Theerens & Cornelis, 2022; Yang et al., 2022). Fuzzy rough representation thus offers a principled, noise-aware measure of fit that a three-way rule can threshold, and it does so without assuming a parametric form for the class-conditional densities.

Each of these ingredients is individually mature, yet their intersection remains thinly populated. Incremental maintenance has been studied mostly for fuzzy rough attribute reduction under changing objects or attributes rather than for online decision making (Zhang et al., 2025), and three-way decision has been extended to dynamic and incomplete information systems with varying objects (Hu et al., 2026; Huang et al., 2025) but typically outside an explicit concept-drift protocol with chronological evaluation. Work that does bring three-way decision to drifting streams has so far favoured generative or ensemble realizations: an incremental Gaussian-mixture model equipped with a fuzzy three-way rule (Shen et al., 2026) and an ensemble whose members

are gated by three-way decisions (Deng et al., 2025). What is missing is a granular, instance-based treatment in which a fuzzy rough membership is maintained incrementally over a drifting stream, in which the three-way thresholds are given a decision-theoretic basis and are allowed to adapt to the current level of instability, and in which detection and adaptation are kept explicitly separate so that the contribution of the decision mechanism is not confounded with that of the change detector.

This paper develops such a framework, IFR-3WD-CD, and evaluates it as a methodological contribution rather than an application. The decision score is a vaguely quantified fuzzy rough membership computed over a bounded sliding window of recently seen, online-standardized instances, so that both the arithmetic and the evidence-quality difficulties are addressed by the same bounded state. Over this score a decision-theoretic three-way rule defines the accept, defer and reject regions, and the thresholds that delimit them are widened by a bounded drift-intensity signal while the stream is unstable, so that the model defers more readily exactly when its reference is least trustworthy. A standard change detector monitors the committed error and triggers selective forgetting, but it only supplies the trigger; the decision behaviour itself is governed by the fuzzy rough membership and the adaptive thresholds. The framework is studied under a strictly chronological prequential protocol that excludes any use of future information, on controlled synthetic benchmarks that isolate the recognised drift types and on two real streams, and it is compared with eight baselines spanning incremental, three-way and strongly drift-aware methods, with an ablation that removes each component in turn. The contributions are, concretely, an incremental vaguely quantified fuzzy rough update whose per-instance cost is independent of the elapsed stream length; a decision-theoretic three-way rule whose boundary adapts to a defined measure of drift intensity; a clean separation of drift detection from drift adaptation; an analysis of how the boundary region behaves through a drift episode; and a systematic empirical comparison, with statistical testing, against relevant streaming baselines. Only components that are implemented and tested are claimed.

The remainder of the paper is organized as follows. Section 2 reviews fuzzy rough sets and three-way decision, streaming and concept-drift learning, and the narrow intersection the paper targets, and states the gap through a structured comparison of prior work. Section 3 formalizes the streaming problem and the fuzzy rough representation, develops the incremental update and the

three-way rule, describes the drift-detection and adaptive-decision mechanism, and sets out the experimental protocol, benchmarks and metrics. Section 4 reports predictive performance and drift adaptation, the behavior of the boundary region and decision stability, and computational efficiency together with the ablation and sensitivity analyses. Section 5 discusses what the results do and do not establish, and Section 6 concludes.

## 2. Literature Review

### 2.1 Fuzzy Rough Sets and Three-Way Decision

Rough set theory models a concept through a pair of approximations induced by an indiscernibility relation, and fuzzy rough set theory generalizes this construction so that both the relation between instances and the membership in the concept may be graded. Given a fuzzy relation R on the instance space and a fuzzy concept A, the lower approximation records the degree to which instances similar to a given one all belong to A, and the upper approximation the degree to which at least one similar instance does; their difference over an instance is a local measure of the uncertainty with which that instance can be assigned to the concept. This apparatus has proved especially effective for selecting informative attributes and for measuring uncertainty in decision tables, and recent work has refined both the uncertainty measures used for feature evaluation (An et al., 2021) and the granular units over which approximations are formed (Ji et al., 2025). A recurring practical concern is robustness: because the lower approximation is an infimum over similar instances, a single mislabeled neighbor can collapse it, which makes the classical operators fragile on noisy data. Two lines of work address this directly. Ordered weighted averaging replaces the hard infimum and supremum by soft aggregations that discount extreme values, yielding approximations that remain informative under noise while retaining a granular interpretation (Palangetić et al., 2022), and vaguely quantified rough sets reformulate the approximations in terms of fuzzy quantifiers such as “most”, so that membership in the lower approximation requires most rather than all similar instances to belong to the concept (Theerens & Cornelis, 2022). Noise-aware variants that explicitly down-weight suspected mislabelled instances have also been proposed for feature selection (Yang et al., 2022). These robustified formulations are the ones suited to streaming data, where noise is the norm rather than the exception, and the decision score developed in this paper is built on the vaguely quantified construction.

Three-way decision proceeds from a complementary premise: that for many problems the appropriate response to an instance is not a binary verdict but one of three actions, namely acceptance, rejection, or a deliberate deferral when neither verdict is warranted (Yao, 2023). The three regions, positive, negative and boundary, partition the universe, and the boundary region collects the instances that current evidence cannot place confidently. The theory supplies more than a label for ambiguity: in its decision-theoretic form the two thresholds that separate the regions are derived from a loss matrix that prices the four possible outcomes of a committed decision and the cost of deferral, so that the width of the boundary region reflects an explicit trade-off rather than an arbitrary choice (Qian et al., 2022). The framework has been extended in many directions, including models that incorporate interaction between decision criteria (Liang et al., 2021), evaluation models that take a dynamic view of decision outcomes (Guo et al., 2022), and sequential and cost-sensitive formulations for classification in which granularity is refined only when deferral persists (Liang et al., 2024). The decision-theoretic derivation of the thresholds and the interpretation of the boundary region as an intentional, cost-weighted category are the aspects this paper carries into the streaming setting.

### 2.2 Data Streams, Incremental Learning, and Concept Drift

A data stream is processed under constraints that static learning does not impose: instances arrive one at a time or in small batches, they must be processed in order and essentially once, and the memory available to the learner does not grow with the stream. Two families of mechanisms keep a stream learner's reference current under these constraints. Windowing retains only the most recent instances, so that a sliding window of fixed size $w$ discards observations older than $w$ steps and thereby bounds both memory and the influence of stale data; fading instead retains older observations but down-weights them geometrically, assigning an instance seen $t - i$ steps ago a weight $\lambda^{t-i}$ with $0 < \lambda \leq 1$. Both are standard, and the choice between them trades responsiveness against stability. Incremental learning more broadly is the discipline of updating a model from each new instance without rebuilding it, and in the rough set tradition it has been developed in detail for maintaining attribute reductions as objects or attributes change over time (Zhang et al., 2025) and for online streaming feature selection, including formulations based on neighbourhood and granular-ball rough sets (Shu et al., 2024; Sun et al., 2024).

Concept drift is the change in $P_t(X, Y)$ over time, and the literature distinguishes several forms that differ in what changes and how quickly. Virtual drift alters the feature distribution $P(X)$ while leaving the posterior $P(Y \mid X)$ essentially fixed, whereas real drift changes the posterior itself and so changes the optimal decision boundary; abrupt drift replaces one concept by another within a few steps, gradual drift interleaves the old and new concepts over a transition window, recurring drift sees a previously observed concept return, and incremental drift moves the concept continuously through a sequence of small changes (Suárez-Cetrulo et al., 2023; Yang et al., 2025). Coping with drift involves two logically distinct tasks that are easily conflated: detecting that a change has occurred and adapting the model in response. Detection is the province of change detectors that monitor a statistic of the stream, typically the model's error, and signal when it departs from its recent behaviour; reviews of this area catalogue detectors based on sequential analysis, windowing and statistical tests, and discuss their sensitivity and delay (Agrahari & Singh, 2022). Adaptation is a separate decision about what to do once a change is flagged, ranging from retraining and window resetting to the management of an ensemble; strong adaptive ensembles for evolving streams, in particular adaptive tree forests, are among the most effective general-purpose stream learners (Paim & Enembreck, 2025), and a careful study must not attribute to an adaptive decision rule an improvement that is in fact due to the detector that triggers it.

### 2.3 Drift-Aware Fuzzy Rough Decision Systems and Research Gap

The combination that this paper targets, in which a fuzzy rough representation, a three-way decision rule, incremental maintenance and concept-drift adaptation operate together on a chronological stream, sits at the junction of the two literatures just reviewed and remains only partially explored. Three-way decision has been taken into dynamic environments, with incremental algorithms for updating three-way approximations as neighborhood decision systems vary and dynamic dominance-based models for data whose attribute values change (Hu et al., 2026; Huang et al., 2025); these works establish how three-way regions can be maintained under change, but they are framed as updates to an information system rather than as online decision making under an explicit drift protocol with prequential evaluation. From the streaming side, the studies closest to the present one bring three-way decision to drifting streams through generative and ensemble realizations: an incremental Gaussian-mixture model in which a fuzzy three-way rule governs how components absorb new data under drift (Shen et al., 2026), and an ensemble

whose base learners are combined through three-way decisions for drift adaptation (Deng et al., 2025). These confirm the value of a third, deferral option under drift, but they rely on a parametric mixture or on an ensemble rather than on a granular fuzzy rough membership, and the behaviour of the boundary region itself through a drift episode is not their focus.

Table 1 positions representative prior work against the components that define the target combination. The pattern it makes visible is that the individual components are well covered and several pairings are mature, but the specific intersection, a granular fuzzy rough membership maintained incrementally, read by a decision-theoretic three-way rule whose boundary adapts to a measured level of drift intensity, with detection kept separate from adaptation and with the dynamics of the boundary region studied explicitly, is not occupied by any single prior study. This is the gap the paper addresses. The claim is deliberately narrow: it is not that fuzzy rough sets, three-way decision and concept drift have never appeared together, since they have (Shen et al., 2026; Deng et al., 2025), but that an incremental, drift-aware, instance-based fuzzy rough three-way mechanism with an adaptive boundary and a clean detection–adaptation separation has not, and that such a mechanism is worth developing because it couples a noise-tolerant uncertainty representation to an explicit, cost-weighted deferral option at a per-instance cost that does not grow with the stream.

**Table 1.** Coverage of defining components by prior work and proposed framework.

| Study | Fuzzy rough | Three-way | Incremental | Streaming | Concept drift | Boundary / defer | Adaptive update |
|---|---|---|---|---|---|---|---|
| Palangetić et al. (2022) | ● | ○ | ○ | ○ | ○ | ● | ○ |
| Theerens & Cornelis (2022) | ● | ○ | ○ | ○ | ○ | ● | ○ |
| An et al. (2021) | ● | ○ | ○ | ○ | ○ | ● | ○ |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Yao (2023) | ○ | ● | ○ | ○ | ○ | ● | ○ |
| Qian et al. (2022) | ● | ● | ○ | ○ | ○ | ● | ○ |
| Liang et al. (2024) | ○ | ● | ○ | ○ | ○ | ● | ● |
| Zhang et al. (2025) | ● | ○ | ● | ● | ○ | ○ | ● |
| Sun et al. (2024) | ● | ○ | ● | ● | ○ | ○ | ● |
| Hu et al. (2026) | ○ | ● | ● | ● | ○ | ● | ● |
| Huang et al. (2025) | ○ | ● | ● | ● | ○ | ● | ● |
| Shen et al. (2026) | ● | ● | ● | ● | ● | ● | ● |
| Deng et al. (2025) | ○ | ● | ● | ● | ● | ● | ● |
| Suárez-Cetrulo et al. (2023) | ○ | ○ | ● | ● | ● | ○ | ● |
| Paim & Enembreck (2025) | ○ | ○ | ● | ● | ● | ○ | ● |
| IFR-3WD-CD (proposed) | ● | ● | ● | ● | ● | ● | ● |

## 3. Materials and Methods

### 3.1 Streaming Problem Formulation and Fuzzy Rough Representation

Figure 1 gives an overview of the proposed framework: each arriving instance is standardized online, scored by an incrementally maintained fuzzy rough membership, and assigned by a three-way rule to an accept, defer or reject region, while a drift layer monitors the committed error and adapts both the reference set and the decision thresholds. The remainder of this section formalizes each stage, beginning with the streaming problem and the fuzzy rough representation.

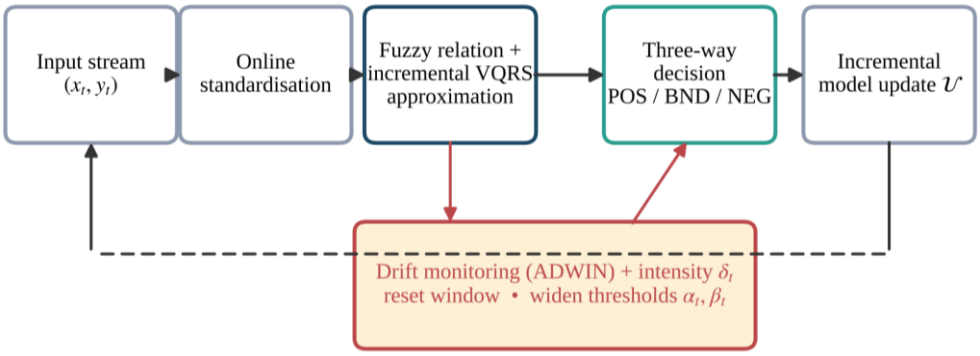



**Figure 1.** IFR-3WD-CD streaming pipeline: standardization, fuzzy rough scoring, three-way decision, drift adaptation.

The stream is a time-ordered sequence $\mathcal{S} = \{(\mathrm{x_t}, \mathrm{y_t})\}_{\mathrm{t=1}}^{\mathrm{T}}$ with feature vectors $\mathrm{x_t} \in \mathbb{R}^{\mathrm{d}}$ and labels $\mathrm{y_t} \in \mathcal{Y}$. The development and the experiments concern binary decisions, $\mathcal{Y} = \{0,1\}$, with the class of interest identified as positive; the construction extends to several classes in a one-against-the-rest fashion, which is noted where relevant but not pursued empirically. Each pair is drawn from a possibly time-varying distribution, $(\mathrm{x_t}, \mathrm{y_t}) \sim \mathrm{P_t(X, Y)}$, and concept drift is the event

$$\mathrm{P_t(X, Y)} \neq \mathrm{P_{t+\Delta}(X, Y)}, \qquad \Delta > 0. \tag{1}$$

It is useful to separate a change in the marginal $\mathrm{P(X)}$ with the posterior $\mathrm{P(Y \mid X)}$ held fixed, termed virtual drift, from a change in $\mathrm{P(Y \mid X)}$ itself, termed real drift, because only the latter moves the optimal decision boundary. Orthogonally to this distinction, a change may be abrupt, replacing one concept by another within a few steps; gradual, mixing the outgoing and incoming concepts over a transition window; recurring, when an earlier concept returns; or incremental, when the concept moves continuously through many small changes. The framework is agnostic to the kind of drift and is tested against all four controlled forms further.

Because the stream must be processed in order and without access to the future, every quantity the model uses at time $t$ is a function of $(x_1, y_1), \ldots, (x_{t-1}, y_{t-1})$ and of $x_t$ only. Features on heterogeneous scales are standardised online using running moments maintained by Welford's method. Writing $n$ for the number of instances seen so far and $\mu_k$, $M_{2,k}$ for the running mean and sum of squared deviations of feature $k$, each arrival updates

$$n \leftarrow n + 1, \quad \mu_k \leftarrow \mu_k + \frac{x_{t,k} - \mu_k}{n}, \quad M_{2,k} \leftarrow M_{2,k} + \left(x_{t,k} - \mu_k^{\text{old}}\right)\left(x_{t,k} - \mu_k\right), \tag{2}$$

so that the running standard deviation is $\sigma_k = \sqrt{M_{2,k}/(n-1)}$ and the standardised vector is $z_{t,k} = \left(x_{t,k} - \mu_k\right)/\sigma_k$. The moments are updated from feature values only and so never use a label; standardising $x_t$ with the moments that include $x_t$ uses no future information, and this is the only normalization performed. Similarity between standardized instances is measured by a fuzzy tolerance relation. Per attribute, the relation is the linear similarity

$$R_k(a, b) = \max\left(0,\ 1 - \frac{|a - b|}{\sigma_R}\right) \in [0,1], \tag{3}$$

with a bandwidth $\sigma_R > 0$ expressed in standardised units, and the attribute-wise similarities are combined by a uniform aggregation,

$$R(x, y) = \frac{1}{d}\sum_{k=1}^{d} R_k\left(z_{x,k}, z_{y,k}\right). \tag{4}$$

The uniform (arithmetic-mean) aggregation is a symmetric ordered weighted aggregation with equal weights; it is preferred here to the strict minimum t-norm because the minimum contracts too aggressively as the number of attributes grows, whereas a soft aggregation of this kind is the standard robust choice in fuzzy rough modelling (Palangetić et al., 2022). The relation is reflexive and symmetric and takes values in $[0,1]$, with $R(x, y) = 1$ when the two instances coincide. Given the relation, the fuzzy rough approximations of the positive concept $A$, identified with the indicator $A(y) = 1$ if $y$ is positive and $A(y) = 0$ otherwise, are defined through the Łukasiewicz residual implicator $I_L(a, b) = \min(1,\ 1 - a + b)$ and t-norm $T_L(a, b) = \max(0,\ a + b - 1)$:

$$\underline{R}A(x) = \inf_y I_L\big(R(x, y), A(y)\big), \qquad \overline{R}A(x) = \sup_y T_L\big(R(x, y), A(y)\big). \tag{5}$$

For a crisp concept these reduce to transparent quantities. Since $I_L(a,1) = 1$ and $I_L(a,0) = 1 - a$, the lower approximation is governed entirely by the negative instances, and since $T_L(a,1) = a$ and $T_L(a,0) = 0$, the upper approximation is governed by the positive ones:

$$\underline{R}A(x) = 1 - \sup_{y:\, y=0} R(x,y), \qquad \overline{R}A(x) = \sup_{y:\, y=1} R(x,y). \tag{6}$$

The lower approximation is therefore one minus the similarity to the most similar negative instance, and the upper approximation is the similarity to the most similar positive instance; the fuzzy boundary membership $\overline{R}A(x) - \underline{R}A(x)$ is large exactly when an instance is simultaneously close to both classes, which is the hallmark of genuine overlap.

These supremum- and infimum-based operators are, however, too sensitive to individual noisy instances to be used directly on a stream: a single mislabelled neighbor can dominate the nearest-of-class similarity. The decision score is therefore built on the vaguely quantified form of the approximations, in which membership is a function of the proportion of similar instances that belong to the concept rather than of the single most similar one (Theerens & Cornelis, 2022). Let $N_k(x)$ denote the $k$ instances in the current reference set most similar to $x$ under $R$. The positive decision score is the similarity-weighted proportion of positive instances among them,

$$p_t(x) = \frac{\sum_{y \in N_k(x)\,:\, y=1} R(x,y)}{\sum_{y \in N_k(x)} R(x,y)} \in [0,1], \tag{7}$$

and is set to $1/2$ when the reference set carries no usable similarity. This score is the argument of the vaguely quantified lower and upper approximations and inherits their robustness: averaging over the $k$ nearest instances of each class rather than taking a single extreme tolerates mislabelled neighbors, while the ratio form makes the score respond to the local balance of evidence. Its behaviour at the two kinds of uncertainty is exactly what a deferral mechanism requires. When an instance lies deep in one class, the similar instances are overwhelmingly of that class and the score approaches $0$ or $1$; when the classes overlap, high similarities to both classes drive the score toward $1/2$; and when evidence is sparse, with no strongly similar instance of either class, the numerator and denominator are both small and comparable, again yielding a score near $1/2$. Overlap and sparsity, the two situations in which a confident verdict is unwarranted, are thus mapped to the centre of the score, where the three-way rule will place them in the boundary region. Figure 2

illustrates the resulting partition on a two-dimensional snapshot, with the boundary band tracking the region of class overlap.

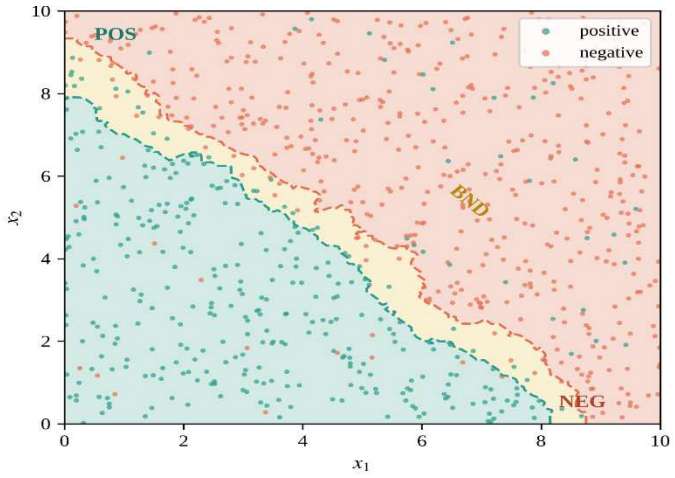



**Figure 2.** Fuzzy rough decision regions showing accept, defer, and reject bands.

### 3.2 Incremental Fuzzy Rough Updating and Three-Way Decision

The reference set over which the score is computed is a sliding window $W_t$ of the w most recent standardized instances with their labels, chosen over a fading scheme as the primary mechanism and compared against it. Maintaining the window is the incremental update operator. On the arrival of $(x_{t+1}, y_{t+1})$ the state

$$\Theta_t = \left(W_t,\ \{\mu_k, M_{2,k}\}_{k=1}^{d},\ \text{detector},\ \text{error EWMAs}\right) \tag{8}$$

is advanced by

$$\Theta_{t+1} = \mathcal{U}(\Theta_t, x_{t+1}, y_{t+1}), \tag{9}$$

where $\mathcal{U}$ updates the running moments as, appends the new instance to the window, evicts the oldest instance if the window exceeds w, and updates the drift signals. The essential point is that the score is evaluated over $W_t$, whose size is bounded by w, rather than over the whole history, so the update does not require the past to be revisited and its cost does not grow with t. This is an approximation to the full-history fuzzy rough membership, not an exact incremental reconstruction of it; the approximation is deliberate, since under drift the distant past is precisely what should be forgotten, and it is stated as such rather than presented as exact maintenance.

Over the score $p_t(x)$ the three-way rule partitions instances using two thresholds $\alpha_t$ and $\beta_t$ with $0 \le \beta_t < \alpha_t \le 1$:

$$\mathrm{POS}_t = \{x: p_t(x) \ge \alpha_t\}, \quad \mathrm{BND}_t = \{x: \beta_t < p_t(x) < \alpha_t\}, \quad \mathrm{NEG}_t = \{x: p_t(x) \le \beta_t\}. \qquad (10)$$

By construction these three regions are disjoint and exhaust the instance space, so every instance receives exactly one of the actions accept, defer or reject. Positive and negative instances are acted on, while the boundary instances are deferred as cases whose current evidence is insufficient for a confident verdict. Where a forced label is required, for comparison with two-way methods or to feed the drift detector, it is taken as 1 if $p_t(x) \ge 1/2$ and 0 otherwise.

The thresholds are given a decision-theoretic basis rather than chosen by hand. Let $\lambda_{PP}, \lambda_{BP}, \lambda_{NP}$ be the losses of accepting, deferring and rejecting an instance that is in fact positive, and $\lambda_{PN}, \lambda_{BN}, \lambda_{NN}$ the corresponding losses when it is in fact negative. Minimising the expected loss of the three actions yields the standard thresholds

$$\alpha = \frac{\lambda_{PN} - \lambda_{BN}}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})}, \qquad \beta = \frac{\lambda_{BN} - \lambda_{NN}}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})}. \qquad (11)$$

With correct decisions costless, $\lambda_{PP} = \lambda_{NN} = 0$, the two misclassifications equally costly, $\lambda_{PN} = \lambda_{NP} = 1$, and deferral priced symmetrically at $\lambda_{BP} = \lambda_{BN} = c_B$ with $0 < c_B < 1$, these expressions simplify to the symmetric base thresholds

$$\alpha_0 = 1 - c_B, \qquad \beta_0 = c_B, \qquad (12)$$

that the single cost $c_B$ controls the base width of the boundary region: a small deferral cost makes deferring attractive and widens the region, a large one narrows it. Asymmetric costs give asymmetric thresholds through the general expressions above; the symmetric case is used as the default because the benchmarks do not single out one error as more serious than the other. The thresholds are not selected by tuning on the evaluation stream; they follow from the cost, and their sensitivity is examined further.

Three structural properties follow directly from the construction. The region partition just stated, $\mathrm{POS}_t \cap \mathrm{BND}_t = \mathrm{BND}_t \cap \mathrm{NEG}_t = \varnothing$ and $\mathrm{POS}_t \cup \mathrm{BND}_t \cup \mathrm{NEG}_t = \mathbb{R}^d$, holds for any admissible thresholds. The decision is monotone in the evidence: because $p_t$ is a non-decreasing function of

the positive similarity mass and the regions are defined by thresholds on $p_t$, increasing an instance's similarity to positive instances, with all else fixed, can move it from reject toward accept but never the reverse, so a reversal of the committed label can occur only through a change in the window or the thresholds, which is to say through learning or drift adaptation rather than through evidence alone. Finally, in the absence of drift and with the window not yet forgetting, the score over $W_t$ coincides with the vaguely quantified membership computed over exactly those instances, so the incremental computation reproduces the batch computation on the same reference set; the approximation relative to the full history is only the restriction of that set to a window, which is the intended behavior.

### 3.3 Concept-Drift Detection and Adaptive Decision Mechanism

Detection and adaptation are kept deliberately separate. Detection is delegated to a standard change detector, the adaptive windowing detector ADWIN, which maintains a variable-length window of a monitored statistic and signals a change when two sub-windows exhibit significantly different means (Agrahari & Singh, 2022). The statistic monitored here is the forced $0/1$ error $e_t = \mathbb{1}[\hat{y}_t \neq y_t]$ of the committed prediction, revealed once the true label arrives. A detector alarm is used only as a trigger; it does not itself alter any decision. The adaptation it triggers is selective forgetting: on an alarm at time $t$ the window is shrunk to its most recent $r\,w$ instances, with $0 < r < 1$, so that the reference set is purged of the superseded concept and the model re-learns from recent data. This is the only role the detector plays, which ensures that the adaptive behaviour studied below is attributable to the decision mechanism and not to the detector.

Alongside the discrete detector, a continuous and bounded measure of instability drives the adaptation of the thresholds. Two exponentially weighted moving averages of the error are maintained, a fast one $\bar{e}_f$ with rate $\eta_f$ and a slow one $\bar{e}_s$ with rate $\eta_s < \eta_f$, and the drift intensity is their relative gap,

$$\delta_t = \min\left(1,\ \max\left(0,\ \frac{\bar{e}_f - \bar{e}_s}{\bar{e}_s + \varepsilon_0}\right)\right) \in [0,1], \tag{13}$$

with a small $\varepsilon_0 > 0$ preventing division by a near-zero baseline. The quantity rises when the recent error surges above its longer-term level, which is the empirical signature of a concept that the model has not yet absorbed, and returns toward zero as the fast and slow averages reconverge; it

is defined, bounded and computed from information available at time $t$ only. The intensity is not a detector and is not used to declare drift; it modulates the width of the boundary region. The thresholds in force at time $t$ are

$$\alpha_t = \alpha_0 + \kappa\,(1 - \alpha_0)\,\delta_t, \qquad \beta_t = \beta_0 - \kappa\,\beta_0\,\delta_t, \tag{14}$$

with an adaptation strength $\kappa \in [0,1]$. When the stream is stable, $\delta_t \approx 0$ and the thresholds equal their decision-theoretic base values; when instability rises, $\alpha_t$ moves toward 1 and $\beta_t$ toward 0, widening the boundary region so that the model defers more readily exactly while its reference is least trustworthy. The update is monotone in $\delta_t$ and keeps $0 \le \beta_t < \alpha_t \le 1$ throughout. A fourth property, drift adaptation, follows: after a genuine change the detector eventually triggers a window reset and the intensity transiently widens the boundary, so the model both refreshes its reference and raises its deferral rate during the recovery, after which the boundary contracts as the error subsides.

The complete procedure is summarised in three algorithms. Algorithm 1 is the incremental fuzzy rough update; Algorithm 2 is the drift-aware three-way decision taken for each instance; Algorithm 3 is the overall prequential loop that ties them together.

---

**Algorithm 1** Incremental fuzzy-rough update $\mathcal{U}(\Theta_t, x_{t+1}, y_{t+1})$

---

**Input:** State $\Theta_t = (W_t, \{\mu_k, M_{2,k}\}_{k=1}^{d}$, detector, EWMAs); new instance $(x, y)$

**Output:** Updated state $\Theta_{t+1}$

1. **Step 1.** Running moments (Welford):

$$n \leftarrow n + 1,\ \mu_k \leftarrow \mu_k + \frac{x_k - \mu_k}{n}$$

$$M_{2,k} \leftarrow M_{2,k} + \left(x_k - \mu_k^{\text{old}}\right)(x_k - \mu_k)$$

$$\sigma_k = \sqrt{\frac{M_{2,k}}{n-1}},\ z_k = \frac{x_k - \mu_k}{\sigma_k}$$

2. **Step 2.** Forced label:

$$\hat{y}(x) = \begin{cases} 1 & \text{if } p_t(x) \geq \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$$

3. **Step 3.** Error EWMAs:

$$\hat{e}_f \leftarrow \eta_f \, e + \left(1 - \eta_f\right) \hat{e}_f, \; \hat{e}_s \leftarrow \eta_s \, e + (1 - \eta_s) \, \hat{e}_s$$

$$\delta_t = \min \left(1, \; \max \left(0, \; \frac{\hat{e}_f - \hat{e}_s}{\hat{e}_s + \epsilon_0}\right)\right)$$

4. **Step 6.** Window reset:

$$W \leftarrow \{(z_j, y_j) \in W : j \geq \mid W \mid - \lceil r \cdot w \rceil\}$$

5. **return** $\Theta_{t+1}$

---

**Algorithm 2** Drift-aware three-way decision for instance $x_t$

---

**Input:** Instance $x_t$; window $W_t$; running moments $\{\mu_k, M_{2,k}\}_{k=1}^{d}$; base thresholds $(\alpha_0, \beta_0)$; drift i ntensity $\delta_t$; adaptation strength $\kappa$; neighbourhood size $k$

**Output:** Action in {POS, BND, NEG}and forced label $\hat{y}_t$

1. **Standardize** $x_t$using past and present running moments:

$$z \leftarrow \text{standardise}(x_t)$$

2. **Compute fuzzy similarities** for each window instance $(z_j, y_j) \in W_t$:

$$R_j \leftarrow \frac{1}{d} \sum_{k=1}^{d} \max \left(0, \; 1 - \frac{\mid z_k - z_{j,k} \mid}{\sigma_R}\right)$$

3. **Select** the $k$nearest neighbours by similarity:

$$N \leftarrow \text{indices of the } k \text{ largest } R_j$$

4. **Aggregate** similarity mass per class:

$$s_{\text{pos}} \leftarrow \sum_{j \in N,\, y_j = 1} R_j \,,\; s_{\text{neg}} \leftarrow \sum_{j \in N,\, y_j = 0} R_j$$

5. **Compute decision score** (positive membership):

$$p \leftarrow \begin{cases} 0.5 & \text{if } s_{\text{pos}} + s_{\text{neg}} = 0 \\ \dfrac{s_{\text{pos}}}{s_{\text{pos}} + s_{\text{neg}}} & \text{otherwise} \end{cases}$$

6. **Adapt thresholds** to drift intensity:

$$\alpha \leftarrow \alpha_0 + \kappa\,(1 - \alpha_0)\,\delta_t,\; \beta \leftarrow \beta_0 - \kappa\,\beta_0\,\delta_t$$

7. **Assign action** by the three-way rule:

$$\text{action} \leftarrow \begin{cases} \text{POS} & \text{if } p \geq \alpha \\ \text{NEG} & \text{if } p \leq \beta \\ \text{BND} & \text{otherwise} \end{cases}$$

8. **Determine forced label:**

$$\hat{y}_t \leftarrow \begin{cases} 1 & \text{if } p \geq 0.5 \\ 0 & \text{otherwise} \end{cases}$$

9. **return** action, $\hat{y}_t$

---

**Algorithm 3** IFR-3WD-CD prequential stream processing

---

**Input:** Stream $S = \{(x_t, y_t)\}_{t=1}^{T}$; parameters $(w, \sigma_R, \kappa, c_B, \delta_{\text{ADWIN}}, r, k, \eta_f, \eta_s)$

**Output:** Per-instance actions $\{\text{action}_t\}_{t=1}^{T}$ and running performance

1. **Initialize** empty window $W_0 \leftarrow \emptyset$, zero running moments $\{\mu_k, = 0\; M_{2,k} = 0\}_{k=1}^{d}$,

   detector, error EWMAs $\hat{e}_f, \hat{e}_s$; set base thresholds:

$$\alpha_0 \leftarrow 1 - c_B,\; \beta_0 \leftarrow c_B$$

2. **for** $t = 1$ **to** $T$ **do**

3. **Update** running moments with features of $x_t$ *(no label used)*:

$$n \leftarrow n+1,\ \mu_k \leftarrow \mu_k + \frac{x_{t,k} - \mu_k}{n}$$
$$M_{2,k} \leftarrow M_{2,k} + \left(x_{t,k} - \mu_k^{\text{old}}\right)\left(x_{t,k} - \mu_k\right)$$

4. **Decide** using Algorithm 2 on $x_t$:

$$(\text{action}_t,\ \text{forced}_t) \leftarrow \text{Algorithm 2}\left(x_t, W_{t-1}, \{\mu_k, M_{2,k}\}, \alpha_0, \beta_0, \delta_{t-1}, \kappa, k\right)$$

5. **Emit** $\text{action}_t$; record performance of $\text{forced}_t$ once $y_t$ is revealed:

$$\text{Acc}_t \leftarrow \mathbb{1}[\text{forced}_t = y_t]$$

6. **Update** state via Algorithm 1 with revealed label:

$$\Theta_t \leftarrow \mathcal{U}(\Theta_{t-1}, x_t, y_t)$$

7. **end for**

---

The cost of the framework follows from these algorithms. Scoring an instance evaluates the relation against the window and selects the k nearest, which is $O(w\,d)$; the running-moment update is $O(d)$; the window maintenance is amortised $O(1)$; and the detector update is amortised near-constant in w. The per-instance cost is therefore

$$C_{\text{update}} = O(w\,d), \tag{15}$$

independent of the elapsed stream length t, and the memory footprint is

$$M_{\text{stream}} = O(w\,d), \tag{16}$$

the window plus a constant number of scalars. Processing the whole stream costs $O(T\,w\,d)$ in time and $O(w\,d)$ in memory. The natural point of comparison is a fuzzy rough model rebuilt from the accumulated history, which at step t must score against all t past instances at cost $O(t\,d)$ and must store them, giving a total time of $\sum_{t=1}^{T} O(t\,d) = O(T^2\,d)$ and memory $O(T\,d)$ that both grow with the stream. The incremental scheme thus replaces a quadratic total time and an unbounded memory by a linear total time and a constant memory, a difference derived from the update rules rather than inferred from measured runtimes; the measured runtimes are reported only as corroboration. Table 2 collects the per-instance and total costs.

**Table 2.** Per-instance and whole-stream computational cost comparison for incremental versus batch.

| Quantity | Incremental (proposed) | Batch reconstruction |
|---|---|---|
| Fuzzy relation / scoring per instance | $O(w\,d)$ | $O(t\,d)$ |
| Running-statistics update | $O(d)$ | $O(t\,d)$ |
| Drift monitoring per instance | $O(\log w)$ amortised | $O(\log t)$ |
| Three-way decision per instance | $O(1)$ | $O(1)$ |
| Per-instance total | $O(w\,d)$ | $O(t\,d)$ |
| Whole-stream time | $O(T\,w\,d)$ | $O(T^2\,d)$ |
| Memory | $O(w\,d)$ | $O(T\,d)$ |

### 3.4 Experimental Protocol, Benchmarks, and Evaluation Metrics

Every method is evaluated under the prequential, test-then-train protocol, which respects the streaming constraints. For each instance in chronological order the model first predicts or decides, its performance on that instance is then recorded, the true label is revealed, and only afterwards is the model updated. The stream is never shuffled, no instance is used before its time, and all normalization uses running statistics computed from past and present instances only, so that no future information can leak into a present decision; the five steps of receiving, deciding, scoring, revealing and updating are applied uniformly to every method. Because the benchmarks are processed in their native order and the only statistics kept are running ones, the protocol excludes the familiar leakage routes of random train–test splitting of a temporal stream, future-aware normalization, retrospective threshold optimization and tuning on the evaluation stream.

The benchmarks are of two kinds, summarized in Table 3. Four synthetic streams isolate the recognised drift types under controlled conditions. The SEA generator produces three-dimensional instances with two relevant features and labels an instance positive when the sum of its first two features falls below a threshold, and the threshold is changed over the stream to induce drift: SEA-Abrupt changes it sharply at three equally spaced points, SEA-Gradual interpolates between successive concepts over a transition window, and SEA-Recurring alternates between two concepts so that an earlier one returns. The rotating-hyperplane generator produces instances

labelled by a linear separator whose weights drift continuously, giving incremental drift with no discrete change points. These generators are standard controlled benchmarks in the drift literature, and their known change points allow detection delay and recovery to be measured against ground truth. Two real streams assess external validity. The Electricity stream records changes in electricity price against demand and transfer indicators, with six features already scaled to the unit interval and a binary target; the Weather stream records daily meteorological measurements with eight features on heterogeneous scales, which the online standardization must accommodate, and a class-imbalanced binary target. For the real streams the true change points are not documented, so no drift ground truth is assumed and any detected changes are treated as estimates rather than as known events.

The Electricity (elec2) dataset is publicly available through the scikit-multiflow streaming-datasets repository (https://github.com/scikit-multiflow/streaming-datasets) and the River library (https://riverml.xyz), originally introduced by Harries (1999). The Weather (NOAA) dataset is available through the scikit-multiflow streaming-datasets repository and was reported by Elwell and Polikar (2011). The synthetic SEA and rotating hyperplane streams were generated following the benchmark descriptions in Street and Kim (2001; https://doi.org/10.1145/502512.502568) and Hulten et al. (2001; https://doi.org/10.1145/502512.502529), respectively.

**Table 3.** Benchmark streams: synthetic drift types and real-world data characteristics.

| Stream | Type | Drift | Instances | Features | Classes | Positive rate | Notes |
|---|---|---|---|---|---|---|---|
| SEA-Abrupt | synthetic | abrupt (3 changes) | 30,000 | 3(2relevant) | 2 | ≈0.39 | 5% label noise |
| SEA-Gradual | synthetic | Gradual (3 transitions) | 30,000 | 3(2relevant) | 2 | ≈0.39 | transition width 2,000 |
| SEA-Recurring | synthetic | recurring (A, B, A, B) | 30,000 | 3(2relevant) | 2 | ≈0.42 | 5% label noise |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| HYP-Incremental | synthetic | incremental (continuous) | 30,000 | 5 | 2 | ≈0.50 | drifting hyperplane |
| Electricity | real | undocumented | 45,312 | 6 | 2 | 0.425 | features in [0,1] |
| Weather | real | undocumented | 18,159 | 8 | 2 | 0.314 | heterogeneous scales |

The framework is compared with eight methods chosen to cover the relevant alternatives rather than only weak baselines. An incremental Gaussian naive Bayes and a Hoeffding tree represent standard streaming classifiers without explicit drift handling; a Hoeffding adaptive tree and an adaptive random forest represent strong drift-aware learners, the forest being among the most effective general-purpose stream classifiers (Paim & Enembreck, 2025); a sliding-window k-nearest-neighbour classifier represents incremental instance-based learning, the closest conventional analogue to the window used here; a static fuzzy rough classifier, fitted on an initial window and then frozen, isolates the effect of never updating; and a three-way naive Bayes, which applies the same decision-theoretic three-way rule to an incremental Bayesian posterior, represents three-way decision without a fuzzy rough representation. The proposed framework is the ninth method. To isolate the contribution of each component, seven ablations are also run: replacing the graded fuzzy similarity by a crisp tolerance (−Fuzzy); replacing the fuzzy rough neighbourhood membership by an online class-conditional Gaussian posterior (−Rough); removing the boundary region by setting a single threshold at one half ($-3WD; replacing the bounded window by reconstruction over the whole history (Incremental);disabling both the detector triggered reset and the threshold widening (-$ Drift Adapt); fixing the thresholds at their base values while retaining the reset (static thresholds); and replacing the sliding window by a fading scheme (fading memory). The full framework is the eighth configuration in this group.

Performance is reported with metrics appropriate to a streaming decision problem, so that ordinary accuracy is never the sole figure. Predictive quality is summarized by the prequential accuracy, the balanced accuracy, the macro-averaged $F_1$ score and, given the class imbalance of one real stream, the geometric mean of the class recalls, all computed from the forced predictions so that three-way and two-way methods are directly comparable. The three-way behaviour is summarised

separately by the boundary rate, the fraction of instances deferred; the coverage, its complement, the fraction committed; the selective accuracy, the accuracy on the committed instances; the precision of the positive and negative regions; and the accuracy–coverage relationship, obtained by sweeping the deferral threshold and tracing selective accuracy against coverage. Drift adaptation is quantified, on the synthetic streams where the change points are known, by the detection delay, the number of steps between an induced change and the first subsequent detector alarm,

$$D_{\text{detect}} = t_{\text{detected}} - t_{\text{drift}}, \tag{17}$$

reported only where ground truth exists; by the recovery time, the number of steps after a change until the fading prequential accuracy returns to within a fixed fraction of its pre-change level and remains there,

$$D_{\text{recover}} = t_{\text{stable}} - t_{\text{drift}}, \tag{18}$$

with the stability criterion stated explicitly as a sustained return to $97\%$ of the pre-change accuracy; and by the accuracy drop $\Delta\text{Acc} = \text{Acc}_{\text{before}} - \text{Acc}_{\text{after}}$ across the change. Computational cost is reported as the total runtime, the mean per-instance time and the working memory, the last measured as the number of stored instances for the window-based methods and as the serialized model size for the learners, with the measurement basis stated so the figures are not misread as directly comparable across method families. Decision stability is reported as a switch rate: after each update a small set of recent instances is re-decided, and the per-update fraction of their decisions that change is averaged over the stream, a quantity interpreted with care because some switching is required to track genuine drift and fewer switches are not automatically better.

Comparisons across benchmarks are supported by statistical tests. Treating each dataset–seed combination as an independent problem instance, the methods are ranked on each instance and the Friedman test is applied to the ranks, with the Nemenyi critical difference used to judge which rank gaps are significant at the $5\%$ level; the proposed framework is additionally compared against each baseline by the Wilcoxon signed-rank test over the instances, with the rank-biserial correlation reported as an effect size. Significance is not treated as the only evidence, and effect sizes and the direction and size of the differences are reported alongside the p-values. All experiments are run in Python 3.13 with NumPy 2.5, SciPy 1.18, scikit-learn 1.9 and the river 0.26

streaming library, on a two-core machine. Randomness in the synthetic generators and the ensemble learners is seed-controlled, and the synthetic results are averaged over five seeds with the standard deviation reported; the real streams are deterministic and run once. The framework's hyperparameters are fixed a priori at the values listed in Table 4 and are held constant across all benchmarks; they are not selected from the evaluation streams, and their influence is characterized separately by the sensitivity analysis, in which each parameter is varied in turn around these defaults.

**Table 4.** Default hyperparameters of IFR-3WD-CD fixed a priori across benchmarks.

| Parameter | Symbol | Value | Role |
|---|---|---|---|
| Window size | w | 800 | bounded reference set |
| Similarity bandwidth | $\sigma_R$ | 2.0 | width of the fuzzy tolerance relation |
| Neighborhood size | k | 25 | instances in the vaguely quantified membership |
| Deferral cost | $c_B$ | 0.30 | base thresholds $\alpha_0 = 0.70,\ \beta_0 = 0.30$ |
| Adaptation strength | $\kappa$ | 0.50 | threshold widening with drift intensity |
| Detector sensitivity | $\delta_{ADWIN}$ | 0.05 | change-detection confidence |
| Reset fraction | r | 0.10 | window retained after an alarm |
| EWMA rates | $\eta_f, \eta_s$ | 0.02, 0.002 | fast and slow error averages |

## 4. Results

The results are organized to follow the questions set out earlier: predictive performance and drift adaptation first, then the behaviour of the three-way boundary and the stability of the decisions, and finally computational efficiency together with the ablation and sensitivity analyses. Throughout, forced predictions are used for the accuracy, balanced-accuracy, macro-$F_1$ and geometric-mean figures so that every method is compared on the same basis, and the three-way quantities are reported separately. Synthetic results are means over five seeds; the standard deviations are small, below one accuracy point in almost every cell, and are given with the per-dataset accuracies in Table 5.

### 4.1 Predictive Performance and Drift Adaptation

Table 5 reports prequential accuracy for every method on every benchmark, and Table 6 the corresponding macro-$F_1$. The pattern is consistent across the two measures. On raw forced accuracy the proposed framework is competitive with the strongest adaptive baselines rather than uniformly dominant: on the abrupt-drift stream the adaptive random forest is marginally ahead and the sliding-window k-nearest-neighbour classifier is level with the framework, while the framework is clearly ahead of the two Hoeffding trees, naive Bayes, the static fuzzy rough classifier and the three-way naive Bayes. This ordering is the expected one. The adaptive random forest is an ensemble of many trees with its own internal drift handling and is among the most capable general-purpose stream learners, so matching it to within a point with a single granular model, while also returning a calibrated deferral signal that the forest does not provide, is the intended outcome; the framework is not proposed as a way to squeeze out the last fraction of forced accuracy. The macro-$F_1$ results tell the same story and confirm that the accuracy figures are not an artefact of class balance. The framework's distinctive strength is in adaptation, which Table 7 quantifies on the synthetic streams whose change points are known. After an abrupt change the framework returns to within three per cent of its pre-change accuracy far more quickly than the tree-based learners and the instance-based baseline, because the combination of detector-triggered window resetting and threshold widening refreshes the reference set and raises the deferral rate during the unstable interval rather than committing to a stale concept. The adaptive random forest, although slightly more accurate on average, recovers substantially more slowly, since its ensemble must grow and prune trees before the new concept is reflected in the vote; the Hoeffding trees are slower still.

Figure 3 shows the prequential accuracy over the abrupt-drift stream and makes the mechanism visible: all methods dip at each change point, but the proposed framework's trajectory returns to its plateau sooner and with a shallower trough. Figure 4 isolates the first change and contrasts the framework with its own no-adaptation ablation, with the detector alarms marked; the drift-aware configuration recovers ahead of the configuration in which resetting and widening are disabled, which attributes the faster recovery to the adaptation mechanism and not to passive forgetting alone. Detection delay, reported only where ground truth exists, is modest on the abrupt and recurring streams and, as expected, not applicable under continuous incremental drift, where there is no discrete change to detect and the continuous intensity signal carries the adaptation instead.

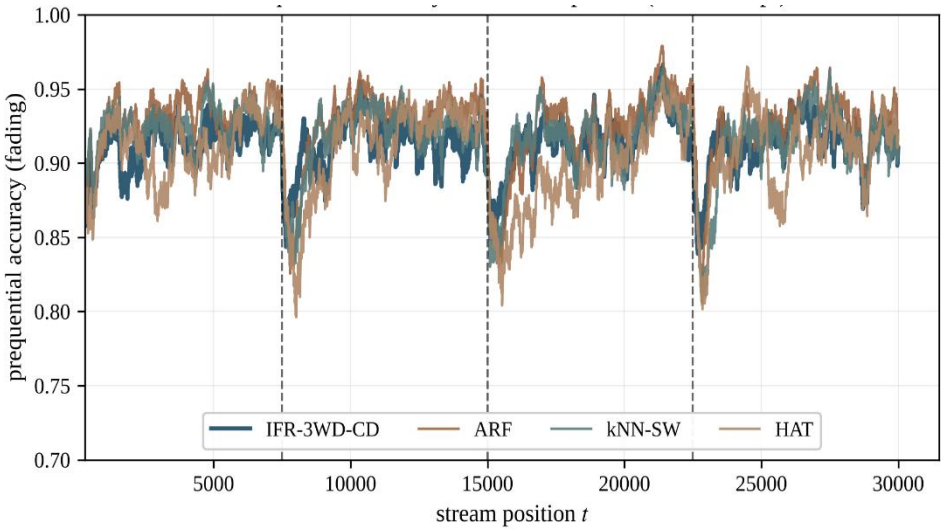



**Figure 3.** Prequential accuracy over abrupt-drift SEA stream for proposed framework and baselines.

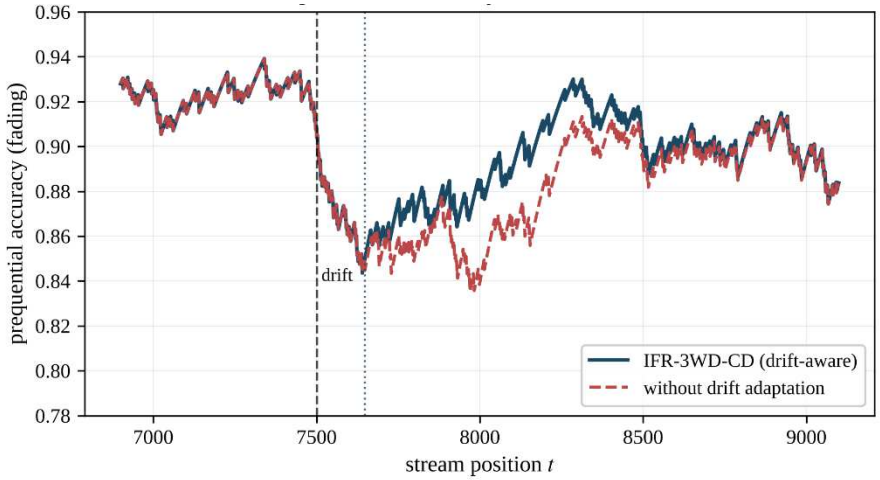



**Figure 4.** Adaptation around first drift: drift-aware versus non-adaptive configuration.

Robustness across the four drift types is summarized in Figure 5, which places the framework beside the adaptive baselines on the abrupt, gradual, recurring and incremental streams. The framework is stable across all four, with no drift type on which it collapses; it is strongest, relative to the baselines, on the abrupt and recurring streams where its resetting mechanism is most useful, and on the incremental stream it edges ahead of the sliding-window instance learner, the best baseline there. On the two real streams the forced-accuracy ranking is more mixed: the adaptive random forest is clearly ahead on electricity and the instance learner leads on Weather, so the

framework does not win on raw forced accuracy there. Its value on these streams appears instead through deferral, examined where the accuracy of its committed decisions exceeds the forced accuracy of the strongest baseline once a portion of the hardest instances is set aside. The benefit of the non-parametric representation over the parametric three-way baseline is also clearer on the real streams than on the synthetic data, consistent with the real streams departing from the linear-Gaussian structure that favors parametric models.

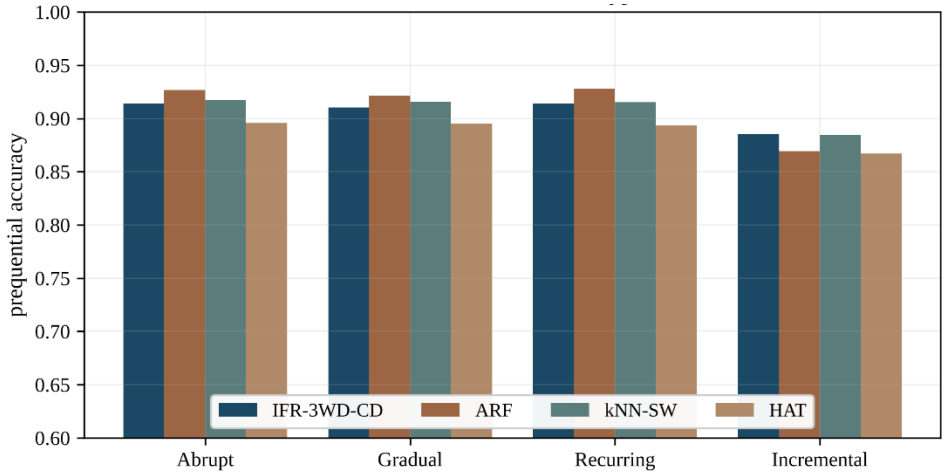



**Figure 5.** Forced prequential accuracy across abrupt, gradual, recurring, and incremental drift types.

### 4.2 Three-Way Boundary Behavior and Decision Stability

The purpose of the three-way structure is not to raise forced accuracy but to expose a reliable confidence ordering, and Table 8 shows that it does. On every benchmark the selective accuracy, computed on the committed instances only, exceeds the forced accuracy by a clear margin while the framework defers a minority of instances on the cleaner streams and more on the noisier ones; the deferred cases are disproportionately the ones the model would otherwise have got wrong, so setting them aside improves the quality of the committed decisions. The effect is largest where it matters most. On the synthetic streams the selective accuracy reaches about 0.95 at a deferral rate near 0.17, roughly three points above the forced accuracy. On the real streams the gain is wider: on electricity the committed decisions reach 0.90 at a deferral rate of about 0.32, which exceeds the 0.86 forced accuracy of the strongest baseline on that stream, and on Weather the committed accuracy of 0.88 likewise surpasses every baseline’s forced accuracy once the hardest instances are deferred. The practical reading is that a user willing to defer a portion of the hardest cases

obtains, from the framework, committed decisions that are more accurate than any baseline's forced decisions on the same stream. The positive- and negative-region precisions are correspondingly high, confirming that the two committed regions are trustworthy and that the boundary region is absorbing genuine ambiguity rather than arbitrary cases. The three-way naive Bayes baseline also shows a selective-accuracy gain, since the mechanism is general, but it commits from a weaker underlying model and so its committed decisions are less accurate than the frameworks at a comparable deferral rate.

This advantage is made quantitative by the accuracy-coverage relationship in Figure 6, obtained by sweeping the deferral threshold and tracing selective accuracy against the fraction committed. The curve for the proposed score lies above those of the baseline confidences across the whole range, which means that at any chosen coverage the committed decisions are more accurate; the framework's own operating point, set by its decision-theoretic thresholds rather than by tuning, sits on the favourable part of this curve. The interpretation is that the fuzzy rough membership provides a better-ordered confidence than the posterior of a parametric model or the vote share of an ensemble, and that the three-way rule turns this ordering into an actionable accept–defer–reject partition.

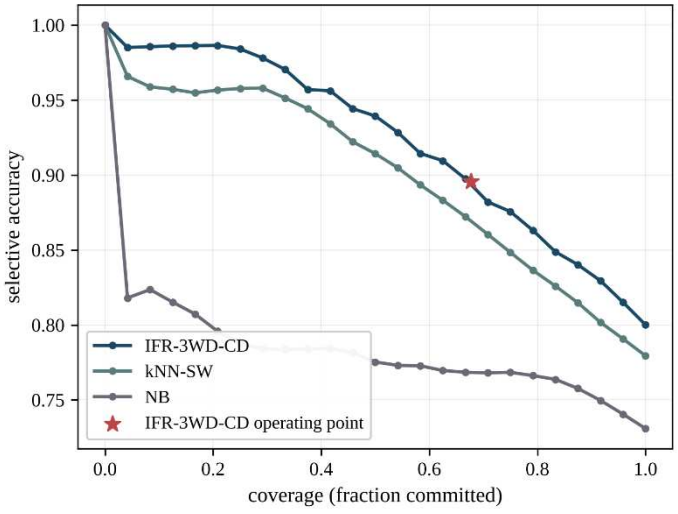



**Figure 6.** Accuracy-coverage trade-off on electricity stream with operating point marked.

How the boundary region behaves over time is central to the paper's claims, and Figures 7 and 8 examine it directly. Figure 7 shows the composition of the three regions over the abrupt-drift

stream: during stable periods the positive and negative regions dominate and the boundary is thin, and at each change the boundary swells as instances that the stale reference can no longer place confidently are deferred, before contracting again as the reference refreshes and the decisions re-sharpen. Figure 8 overlays the boundary rate with the drift-intensity signal and confirms that the two move together, the boundary widening as the intensity rises after a change and narrowing as it subsides; the induced change points coincide with the peaks of both curves. This is the behaviour the framework was designed to produce, and it is observed rather than assumed: the boundary region is an adaptive buffer that expands exactly when the model is least certain and contracts when confidence is restored. The same figures show that the boundary does not expand without limit; it stabilizes at a low level once each new concept is learned, so the deferral cost is incurred mainly during the transient and not in steady state.

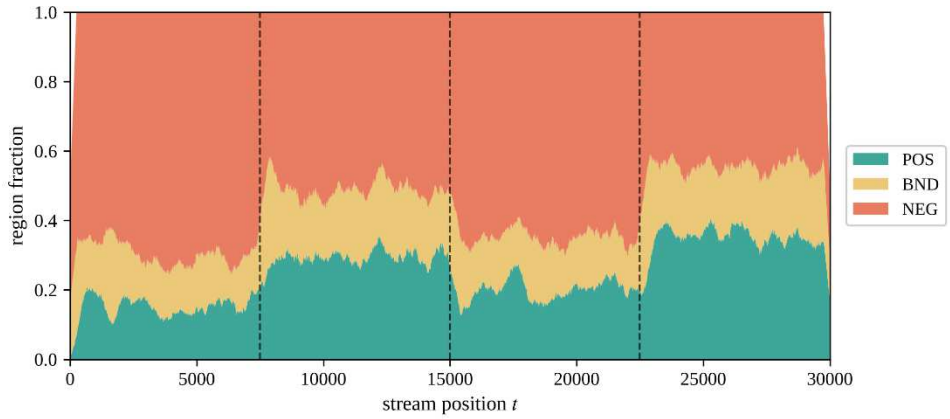



**Figure 7.** Decision-region composition over time on abrupt-drift SEA stream.

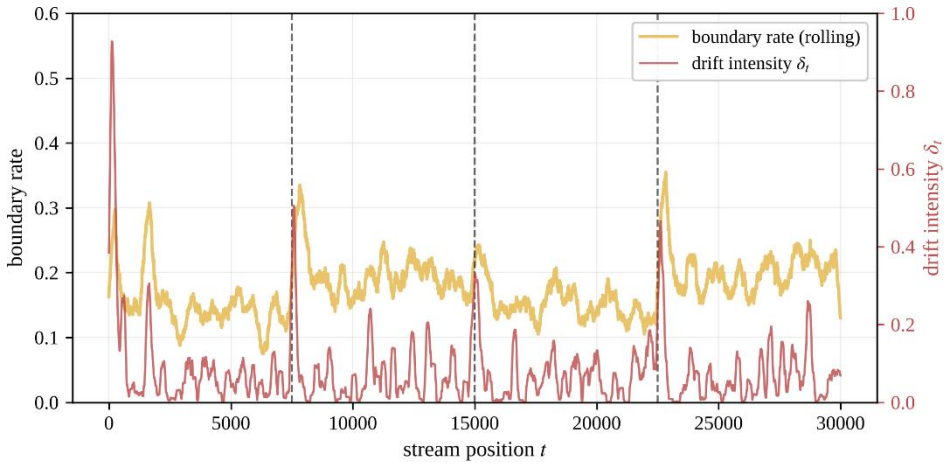


**Figure 8.** Boundary-region dynamics and drift intensity over abrupt-drift stream.

Decision stability is reported in Table 9. The switch rate, the mean per-update fraction of recently seen decisions that change after a model update, is uniformly small for every method, so none of them thrashes; the differences are nonetheless informative. The proposed framework and the adaptive Hoeffding tree have the highest switch rates, reflecting that they actively revise decisions to track the evolving concept, whereas the static fuzzy rough baseline, which never updates its reference, and the sliding-window instance learner, whose large neighborhood is insensitive to a single arrival, switch least but also recover from drift more slowly. This is the stability–plasticity tension in miniature: a method that never revises a decision cannot adapt, and the proposed framework occupies a sensible middle ground, revising when drift requires it and holding steady otherwise. A low switch rate is therefore not a goal in itself, and is interpreted here alongside the recovery results rather than on its own.

### 4.3 Computational Efficiency, Ablation, and Sensitivity

The complexity analysis predicts a per-instance cost independent of the elapsed stream length for the incremental scheme and a growing cost for batch reconstruction, and Figure 9 confirms this empirically. The incremental scheme's per-instance time is flat as the stream lengthens and its stored-memory footprint is constant at the window size, whereas the batch-reconstruction variant's per-instance time rises with the amount of accumulated history and its memory grows without bound. The cumulative runtime of the incremental scheme is therefore linear in the stream length while that of batch reconstruction is superliner, exactly as derived. Table 10 places the framework's per-instance time among the methods: it is heavier than the parametric learners, as an instance-based method evaluating a fuzzy relation against a window must be, but it remains in the same order as the sliding-window k-nearest-neighbour baseline and is well within the budget of an online setting, and its memory is bounded by the window rather than growing with the stream.

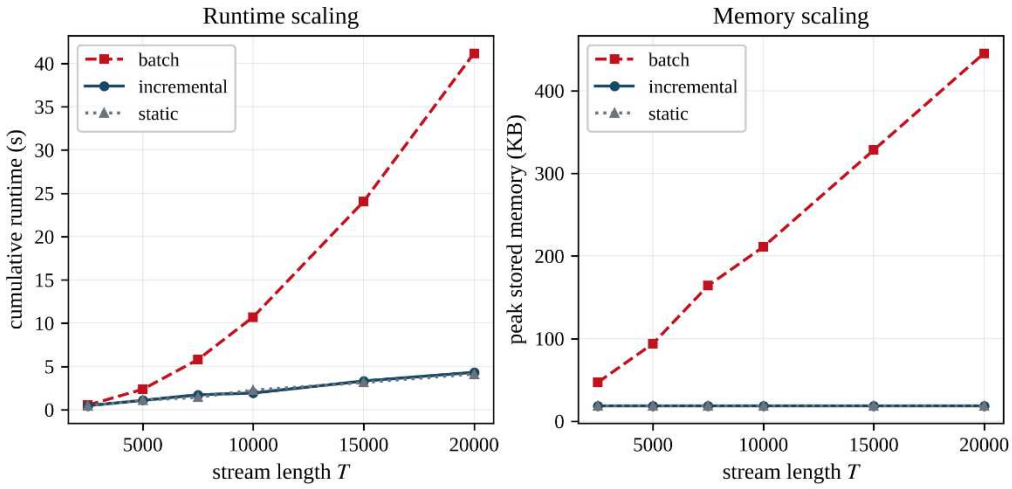



**Figure 9.** Computational scaling: cumulative runtime and peak memory versus stream length.

The ablation in Table 11 isolates the contribution of each component. Replacing the graded fuzzy similarity by a crisp tolerance is the most damaging change, lowering accuracy sharply and inflating the boundary rate, which shows that the granularity of the similarity carries much of the method's discrimination. Replacing the fuzzy rough neighborhood membership by a parametric Gaussian posterior leaves accuracy close to the full method on the synthetic streams, whose structure suits a parametric model, so the value of the instance-based representation is not a uniform accuracy gain but the non-parametric, noise-tolerant uncertainty structure it supplies to the three-way rule, a value that the real-stream results and the accuracy–coverage analysis make concrete. Removing the boundary region reduces the method to a two-way classifier: the forced accuracy is essentially unchanged, as expected, but the selective-accuracy advantage disappears, which is the clearest demonstration that the three-way rule contributes reliability rather than raw accuracy. Disabling drift adaptation lengthens recovery, and fixing the thresholds rather than letting them widen removes the transient rise in deferral around changes; both effects are in the predicted direction. Replacing the window by a fading scheme changes the boundary and recovery profile without dominating the window, confirming the two memory mechanisms as genuine alternatives. Figure 10 summarizes these effects visually, with the forced and selective accuracies side by side and the recovery times beneath.

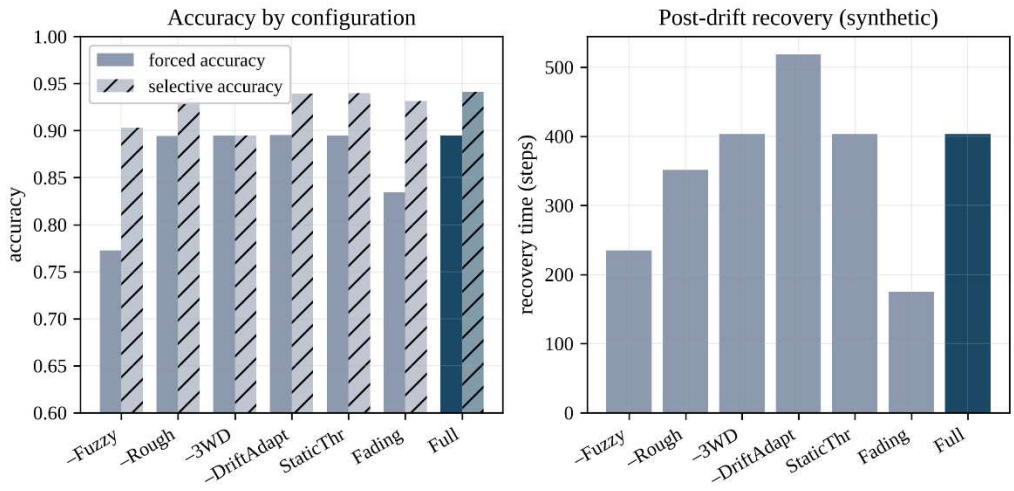



**Figure 10.** Ablation study: forced accuracy, selective accuracy, and post-drift recovery time.

The sensitivity analysis varies each hyperparameter around its default while holding the others fixed, on streams distinct from those used to set the defaults, and the results in Table 12 show that the framework is robust to all of its parameters within sensible ranges. Accuracy changes by small amounts as the window size, similarity bandwidth, neighborhood size and adaptation strength are varied over wide intervals, and no parameter exhibits a sharp optimum that would require careful tuning; the deferral cost behaves exactly as its decision-theoretic role predicts, trading coverage against selective accuracy monotonically, so it is a control to be set from the application's tolerance for deferral rather than a parameter to be optimized. The detector sensitivity trades detection delay against false alarms in the usual way, with the default sitting in the stable part of the range. Because the defaults were fixed a priori and never tuned on an evaluation stream, this analysis characterizes robustness rather than selecting parameters.

Comparisons across benchmarks are placed on a statistical footing in Tables Table 13 and Table 14. Treating each dataset–seed combination as a problem instance gives twenty-two instances, and the Friedman test decisively rejects the hypothesis that the methods are equivalent. The average-rank ordering, visualized as a critical-difference diagram in Figure 11, places the adaptive random forest first, the sliding-window instance learner second and the proposed framework third, with the three separated by less than the critical difference and therefore not significantly distinguishable on rank; the framework in turn outranks the remaining methods, significantly so for the simple Hoeffding tree, the two naive Bayes variants and the static fuzzy rough classifier. The paired Wilcoxon tests sharpen this picture: the framework beats the five weaker baselines on

every one of the twenty-two instances or all but one, with the maximal effect size, and is statistically level with the adaptive random forest on accuracy while trailing the sliding-window learner by about two parts in a thousand. The framework therefore matches the strongest baselines on forced accuracy and adds the deferral capability that none of them provides. The stress testing in Table 15 completes the picture by probing the failure region: as label noise, drift magnitude and class imbalance are increased, the framework degrades gracefully on the first two but, like the baselines, loses minority-class recall under strong imbalance, and at the highest noise level the deferral rate climbs toward one third as the score sits near one half for many instances, which is the over-deferral failure mode discussed in further.

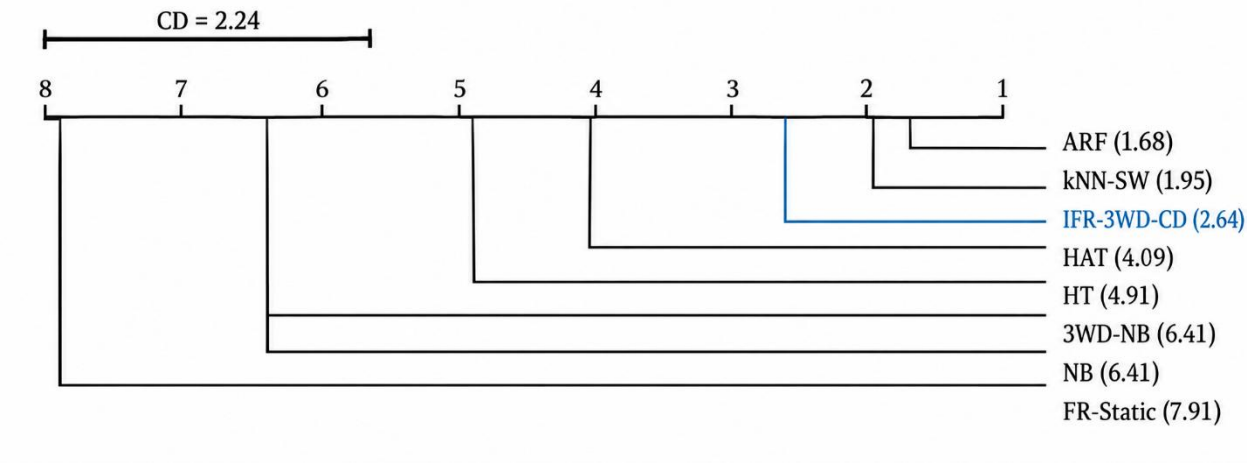



**Figure 11.** Critical-difference diagram of average accuracy ranks across problem instances.

**Table 5.** Prequential accuracy by method and benchmark with standard deviations.

| Method | SEA-Abrupt | SEA-Gradual | SEA-Recurring | HYP-Incremental | Electricity | Weather |
|---|---|---|---|---|---|---|
| Gaussian NB | 0.868±0.001 | 0.870±0.002 | 0.856±0.001 | 0.849±0.021 | 0.731 | 0.690 |
| Hoeffding Tree | 0.884±0.002 | 0.886±0.004 | 0.867±0.000 | 0.863±0.007 | 0.774 | 0.737 |
| Hoeffding Adaptive Tree | 0.896±0.006 | 0.895±0.004 | 0.894±0.007 | 0.867±0.004 | 0.813 | 0.735 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Adaptive Random Forest | 0.927±0.002 | 0.921±0.002 | 0.928±0.003 | 0.869±0.005 | 0.856 | 0.778 |
| Sliding-window k-NN | 0.917±0.002 | 0.916±0.002 | 0.915±0.002 | 0.884±0.001 | 0.779 | 0.790 |
| Fuzzy-rough (static) | 0.842±0.002 | 0.846±0.003 | 0.836±0.002 | 0.777±0.050 | 0.659 | 0.733 |
| Three-way NB | 0.868±0.001 | 0.870±0.002 | 0.856±0.001 | 0.849±0.021 | 0.731 | 0.690 |
| IFR-3WD-CD (proposed) | 0.914±0.002 | 0.910±0.001 | 0.914±0.001 | 0.885±0.002 | 0.800 | 0.764 |

**Table 6.** Macro-averaged $F_1$ by method and benchmark (forced predictions).

| **Method** | **SEA-Abrupt** | **SEA-Gradual** | **SEA-Recurring** | **HYP-Incremental** | **Electricity** | **Weather** |
|---|---|---|---|---|---|---|
| Gaussian NB | 0.847±0.002 | 0.848±0.002 | 0.834±0.002 | 0.849±0.021 | 0.697 | 0.654 |
| Hoeffding Tree | 0.870±0.002 | 0.872±0.005 | 0.851±0.001 | 0.863±0.007 | 0.766 | 0.668 |
| Hoeffding Adaptive Tree | 0.884±0.008 | 0.884±0.004 | 0.883±0.008 | 0.867±0.004 | 0.806 | 0.675 |
| Adaptive Random Forest | 0.920±0.002 | 0.913±0.002 | 0.921±0.003 | 0.869±0.005 | 0.851 | 0.719 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Sliding-windowk-NN | 0.909±0.002 | 0.907±0.002 | 0.907±0.002 | 0.884±0.001 | 0.772 | 0.741 |
| Fuzzy-rough (static) | 0.811±0.003 | 0.816±0.005 | 0.804±0.003 | 0.777±0.051 | 0.584 | 0.604 |
| Three-way NB | 0.847±0.002 | 0.848±0.002 | 0.834±0.002 | 0.849±0.021 | 0.697 | 0.654 |
| IFR-3WD-CD (proposed) | 0.905±0.002 | 0.900±0.001 | 0.905±0.002 | 0.885±0.002 | 0.792 | 0.694 |

**Table 7.** Drift adaptation on the synthetic streams with known change points.

| **Benchmark** | **Method** | **Detection delay** | **Recovery (steps)** | **Accuracy drop** |
|---|---|---|---|---|
| SEA-Abrupt | IFR-3WD-CD | 145 | 479 | 0.105 |
| | −DriftAdapt | – | 603 | 0.115 |
| | Adaptive Random Forest | – | 1008 | 0.152 |
| | Sliding-window k-NN | – | 778 | 0.128 |
| SEA-Gradual | IFR-3WD-CD | 195 | 149 | 0.069 |
| | −DriftAdapt | – | 104 | 0.061 |
| | Adaptive Random Forest | – | 343 | 0.093 |
| | Sliding-window k-NN | – | 141 | 0.060 |
| SEA-Recurring | IFR-3WD-CD | 109 | 582 | 0.135 |
| | −DriftAdapt | – | 848 | 0.156 |
| | Adaptive Random Forest | – | 862 | 0.176 |

| | | | | |
|---|---|---|---|---|
| | Sliding-window k-NN | – | 893 | 0.175 |
| HYP-Incremental | IFR-3WD-CD | – | – | – |
| | −DriftAdapt | – | – | – |
| | Adaptive Random Forest | – | – | – |
| | Sliding-window k-NN | – | – | – |

**Table 8.** Three-way decision behavior: forced accuracy, boundary rate, coverage, selective accuracy.

| **Benchmark** | **Method** | **Forced acc.** | **Boundary rate** | **Coverage** | **Selective acc.** | **POS prec.** | **NEG prec.** |
|---|---|---|---|---|---|---|---|
| SEA-Abrupt | Three-way NB | 0.868 | 0.242 | 0.758 | 0.931 | 0.950 | 0.926 |
| | IFR-3WD-CD (proposed) | 0.914 | 0.170 | 0.830 | 0.947 | 0.948 | 0.946 |
| SEA-Gradual | Three-way NB | 0.870 | 0.241 | 0.759 | 0.932 | 0.951 | 0.928 |
| | IFR-3WD-CD (proposed) | 0.910 | 0.171 | 0.829 | 0.946 | 0.948 | 0.945 |
| SEA-Recurring | Three-way NB | 0.856 | 0.252 | 0.748 | 0.927 | 0.950 | 0.921 |
| | IFR-3WD-CD (proposed) | 0.914 | 0.172 | 0.828 | 0.947 | 0.949 | 0.946 |
| HYP-Incremental | Three-way NB | 0.849 | 0.349 | 0.651 | 0.928 | 0.928 | 0.927 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | IFR-3WD-CD (proposed) | 0.885 | 0.330 | 0.670 | 0.946 | 0.946 | 0.946 |
| Electricity | Three-way NB | 0.731 | 0.150 | 0.850 | 0.762 | 0.857 | 0.736 |
| | IFR-3WD-CD (proposed) | 0.800 | 0.323 | 0.677 | 0.896 | 0.894 | 0.896 |
| Weather | Three-way NB | 0.690 | 0.348 | 0.652 | 0.767 | 0.589 | 0.827 |
| | IFR-3WD-CD (proposed) | 0.764 | 0.443 | 0.557 | 0.882 | 0.872 | 0.883 |

**Table 9.** Decision stability on abrupt-drift stream measured by switch rate.

| Method | Forced acc. | Switch rate |
|---|---|---|
| IFR-3WD-CD | 0.907 | 0.0007 |
| −DriftAdapt | 0.904 | 0.0005 |
| Fuzzy-rough (static) | 0.842 | 0.0001 |
| Hoeffding Adaptive Tree | 0.876 | 0.0007 |
| Sliding-window k-NN | 0.908 | 0.0000 |

**Table 10.** Computational cost: runtime, per-instance time, and working memory.

| Method | Runtime (s) | Per-instance (ms) | Memory (KB) | Memory basis |
|---|---|---|---|---|
| Gaussian NB | 3.5 | 0.115 | 29 | model |
| Hoeffding Tree | 1.1 | 0.038 | 474 | model |
| Hoeffding Adaptive Tree | 2.6 | 0.087 | 323 | model |

| Method | Runtime (s) | Per-instance (ms) | Memory (KB) | Memory basis |
|---|---|---|---|---|
| Adaptive Random Forest | 25.9 | 0.870 | 18410 | model |
| Sliding-window k-NN | 16.5 | 0.549 | 444 | window |
| Fuzzy-rough (static) | 6.7 | 0.222 | 24 | window |
| Three-way NB | 3.3 | 0.109 | 29 | model |
| IFR-3WD-CD (proposed) | 6.4 | 0.215 | 24 | window |

**Table 11.** Ablation study removing each component: accuracy, boundary rate, recovery.

| Configuration | Forced acc. | Selective acc. | Boundary rate | Recovery | Per-inst. (ms) |
|---|---|---|---|---|---|
| −Fuzzy (crisp similarity) | 0.772 | 0.903 | 0.584 | 235 | 0.203 |
| −Rough (Gaussian posterior) | 0.894 | 0.935 | 0.284 | 351 | 0.282 |
| $-$3WD (two-way) | 0.895 | 0.895 | 0.000 | 403 | 0.213 |
| −DriftAdapt | 0.895 | 0.940 | 0.209 | 518 | 0.221 |
| Static thresholds | 0.895 | 0.940 | 0.215 | 403 | 0.211 |
| Fading memory | 0.834 | 0.931 | 0.476 | 175 | 0.476 |
| Full IFR-3WD-CD | 0.895 | 0.941 | 0.226 | 403 | 0.215 |

**Table 12.** Parameter sensitivity on SEA-Abrupt showing accuracy and selective accuracy spans.

| Parameter | Range | Accuracy span | Δ | Selective acc. span |
|---|---|---|---|---|
| Window size w | 200–1600 | 0.888–0.920 | 0.032 | 0.945–0.946 |
| Bandwidth $\sigma_R$ | 1–4 | 0.914–0.915 | 0.001 | 0.946–0.947 |
| Neighbourhood k | 5–50 | 0.911–0.914 | 0.003 | 0.936–0.947 |

| Parameter | Range | Accuracy span | Δ | Selective acc. span |
|---|---|---|---|---|
| Adapt. strength κ | 0–1 | 0.914–0.914 | 0.000 | 0.945–0.947 |
| Deferral cost $c_B$ | 0.1–0.4 | 0.914–0.914 | 0.000 | 0.939–0.950 |
| Detector δ | 0.01–0.1 | 0.914–0.914 | 0.000 | 0.946–0.946 |

**Table 13.** Average accuracy ranks across twenty-two problem instances with Friedman test.

| Method | Mean rank |
|---|---|
| Adaptive Random Forest | 1.68 |
| Sliding-window k-NN | 1.95 |
| IFR-3WD-CD (proposed) | 2.64 |
| Hoeffding Adaptive Tree | 4.09 |
| Hoeffding Tree | 4.91 |
| Three-way NB | 6.41 |
| Gaussian NB | 6.41 |
| Fuzzy-rough (static) | 7.91 |

**Table 14.** Paired Wilcoxon signed-rank comparison against each baseline on accuracy.

| Baseline | Mean acc. diff. | p-value | Effect size | W/L/T |
|---|---|---|---|---|
| Three-way NB | +0.0474 | 4.8e-07 | +1.00 | 22/0/0 |
| Adaptive Random Forest | -0.0082 | 1.1e-01 | -0.55 | 5/17/0 |
| Fuzzy-rough (static) | +0.0811 | 4.8e-07 | +1.00 | 22/0/0 |
| Hoeffding Adaptive Tree | +0.0170 | 9.1e-06 | +0.91 | 21/1/0 |
| Hoeffding Tree | +0.0302 | 4.8e-07 | +1.00 | 22/0/0 |
| Gaussian NB | +0.0474 | 4.8e-07 | +1.00 | 22/0/0 |
| Sliding-window k-NN | -0.0023 | 9.3e-03 | -0.64 | 4/18/0 |

**Table 15.** Stress testing under varied label noise, drift magnitude, and class imbalance.

| Axis | Level | Method | Accuracy | G-mean | Boundary rate | Recovery |
|---|---|---|---|---|---|---|
| noise | noise=0.00 | IFR-3WD-CD | 0.963 | 0.953 | 0.151 | 286 |
| | | −DriftAdapt | 0.961 | 0.951 | 0.138 | 538 |
| | | ARF | 0.980 | 0.976 | 0.000 | 904 |
| | noise=0.05 | IFR-3WD-CD | 0.914 | 0.896 | 0.172 | 552 |
| | | −DriftAdapt | 0.913 | 0.893 | 0.159 | 688 |
| | | ARF | 0.926 | 0.912 | 0.000 | 1090 |
| | noise=0.10 | IFR-3WD-CD | 0.866 | 0.841 | 0.195 | 487 |
| | | −DriftAdapt | 0.865 | 0.840 | 0.182 | 516 |
| | | ARF | 0.874 | 0.854 | 0.000 | 1538 |
| | noise=0.20 | IFR-3WD-CD | 0.767 | 0.734 | 0.327 | 598 |
| | | −DriftAdapt | 0.768 | 0.735 | 0.302 | 604 |
| | | ARF | 0.769 | 0.736 | 0.000 | 1703 |
| magnitude | mag=low | IFR-3WD-CD | 0.914 | 0.903 | 0.176 | 190 |
| | | −DriftAdapt | 0.914 | 0.903 | 0.168 | 190 |
| | | ARF | 0.930 | 0.921 | 0.000 | 194 |
| | mag=medium | IFR-3WD-CD | 0.913 | 0.893 | 0.173 | 720 |
| | | −DriftAdapt | 0.913 | 0.894 | 0.159 | 676 |
| | | ARF | 0.925 | 0.909 | 0.000 | 1401 |
| | mag=high | IFR-3WD-CD | 0.915 | 0.893 | 0.168 | 532 |
| | | −DriftAdapt | 0.910 | 0.887 | 0.157 | 864 |
| | | ARF | 0.928 | 0.911 | 0.000 | 973 |
| imbalance | pos~0.45 | IFR-3WD-CD | 0.914 | 0.896 | 0.172 | 552 |
| | | −DriftAdapt | 0.913 | 0.893 | 0.159 | 688 |
| | | ARF | 0.926 | 0.912 | 0.000 | 1090 |
| | pos~0.25 | IFR-3WD-CD | 0.921 | 0.852 | 0.122 | 130 |

| Axis | Level | Method | Accuracy | G-mean | Boundary rate | Recovery |
|---|---|---|---|---|---|---|
| | | −DriftAdapt | 0.922 | 0.852 | 0.116 | 130 |
| | | ARF | 0.925 | 0.864 | 0.000 | 1011 |
| | pos~0.12 | IFR-3WD-CD | 0.925 | 0.807 | 0.104 | 151 |
| | | −DriftAdapt | 0.925 | 0.808 | 0.098 | 151 |
| | | ARF | 0.929 | 0.826 | 0.000 | 407 |

## 5. Discussion

The experiments were organized around seven questions, and the results speak to each of them. On the incremental maintenance of the fuzzy rough approximation, the framework computes its decision score over a bounded window rather than the accumulated history, and the complexity analysis, corroborated by the measured scaling in Table 10 and the efficiency figure, shows that this keeps the per-instance cost independent of the elapsed stream length while a periodic batch reconstruction grows superlinearly and its memory without bound. The maintenance is therefore online in the strict sense rather than a batch method applied at intervals. On the maintenance of the three-way regions under drift, the region composition tracked over the stream shows the positive and negative regions dominating during stable periods and the boundary region swelling around each change before contracting again, so the partition is not static but responds to the current state of the stream. On the response to drift, the separation of detection from adaptation was deliberate: the detector supplies only a reset trigger, and the comparison with the configuration in which adaptation is disabled isolates the effect of that trigger together with the threshold widening, which is a faster return to the pre-change accuracy on most of the induced changes.

The value of the explicit boundary region is visible in the gap between forced and selective accuracy and in the accuracy–coverage relationship. Committing only to the instances the model places confidently, and deferring the rest, raises the accuracy of the committed decisions appreciably above the accuracy obtained when every instance is forced into a verdict, and the accuracy–coverage curve of the proposed score lies above those of the baseline confidences across the range of coverages, which means that at any chosen level of deferral the committed decisions are more reliable. This is the practical content of the three-way option: it does not improve the forced classification itself, which is why the framework is only competitive and not dominant on

raw accuracy, but it exposes a reliable confidence ordering that a binary classifier does not provide, and it allows a user to trade coverage for reliability in a controlled way. The deferral is not free, and a larger boundary region is not better in itself; the point is that the trade-off is made explicit and is governed by an interpretable cost.

The ablation clarifies where the method's behaviour comes from and, equally importantly, where it does not. Removing the graded fuzzy similarity in favor of a crisp tolerance is the most damaging single change, which confirms that the granularity of the similarity, not merely the neighborhood structure, carries much of the method's discrimination. Replacing the fuzzy rough neighborhood membership by a parametric Gaussian posterior is revealing in the opposite direction, and the split between synthetic and real streams is the key to reading it. On the synthetic streams, whose class-conditional structure is close to the linear-Gaussian ideal, the parametric variant matches the fuzzy rough membership almost exactly, in both forced and selective accuracy. On the real streams the picture reverses: the fuzzy rough membership holds its selective accuracy near 0.90 on electricity and 0.88 on Weather, whereas the parametric variant falls to about 0.83 and 0.78 respectively, a gap of six to ten points that appears precisely where the data depart from the parametric ideal. The reading is that the fuzzy rough representation is not a universal accuracy improver but a non-parametric, noise-tolerant alternative whose benefit is realized on data that are neither linear nor Gaussian, and whose main contribution within the framework is the quality of the uncertainty ordering it supplies to the three-way rule rather than a uniform gain in forced accuracy. Removing the boundary region collapses the method to a two-way classifier and removes the selective-accuracy advantage without changing the forced accuracy, which is consistent with the role just described. Disabling drift adaptation lengthens recovery, and fixing the thresholds rather than widening them removes the transient increase in deferral around changes; both effects are in the expected direction. Replacing the window by a fading scheme recovers faster but at a lower forced accuracy, confirming that windowing and fading are distinct alternatives rather than one dominating the other.

Several failure modes deserve to be stated plainly, because average performance conceals them. The change detector is the weakest link under heavy noise: when the label-noise rate is high the error signal that the detector monitors is dominated by noise, the smaller concept changes produce error increases that are not separable from that noise, and the detector either fires late or misses

them, so that the adaptation it triggers arrives after the window has already partly flushed the old concept on its own. The threshold-widening mechanism is a partial safeguard here, since it responds to the continuous intensity signal rather than to a discrete alarm, but it cannot fully compensate for a missed detection. The aggressive window reset is itself a source of risk: shrinking the window to a small fraction after an alarm accelerates re-learning but raises the variance of the decision score immediately afterwards, and when a detection is both late and followed by a reset the net effect on recovery can be negative rather than positive, which the stress testing makes visible at the larger drift magnitudes. The boundary region can also over-expand: under high noise or sparse evidence the score sits near one half for many instances, the deferral rate climbs, and coverage falls to a point where the method defers more than is useful, so the deferral cost must be set with the expected noise level in mind. Finally, class imbalance erodes the minority-class recall, as it does for the baselines, and although the geometric-mean results show the framework degrades gracefully, the three-way rule does not by itself solve imbalance and would need to be combined with an imbalance-aware mechanism for strongly skewed streams.

These observations bound the scope of the contribution. The fuzzy relation is evaluated against every window instance, so the per-instance cost, while independent of the stream length, is linear in the window size and the dimensionality, and the method will become expensive on very wide, high-dimensional streams where the window must be large; approximate nearest-neighbor indexing or online feature selection in the spirit of recent streaming rough-set work (Sun et al., 2024; Shu et al., 2024) would be the natural remedy and is left to future work. The framework inherits the limitations of whichever change detector it uses, and a different detector would shift the detection delays and false-alarm behaviour reported here. The decision score is an approximation to the full-history fuzzy rough membership, restricted to a window, and this is appropriate under drift but means the method is not intended for stationary streams where forgetting is unnecessary. The benchmarks, though standard, are themselves limited: the synthetic streams make the drift explicit at the cost of realism, and the real streams lack documented change points, so the detection and recovery metrics are reported only where ground truth exists. The development and evaluation are binary, and although the construction extends to several classes in a one-against-the-rest manner, the behaviour of the boundary region in a fully multi-class three-way setting is not studied here. These are limitations of the present study rather than refutations of the approach, and none of them is claimed to be already solved.

The directions they suggest are concrete. Scaling to high-dimensional and wide streams through approximate similarity search and adaptive feature selection is the most pressing. Beyond it, the framework could be extended to imbalanced streams with cost-sensitive or resampling components, to multi-label and fully multi-class streams where the three-way structure becomes richer, and to the delayed- and scarce-label regimes that are common in practice and in which the deferral region acquires an additional meaning as a query criterion for active or human-in-the-loop labelling. Distributed and federated maintenance of the fuzzy rough membership across several streams, adaptation to adversarially induced drift, and deployment on resource-constrained edge devices, where the constant per-instance cost is an asset, are further avenues. Each of these is an open problem, not a solved one. Figure 12 places the present framework on a longer research trajectory toward uncertainty-aware streaming decision systems.

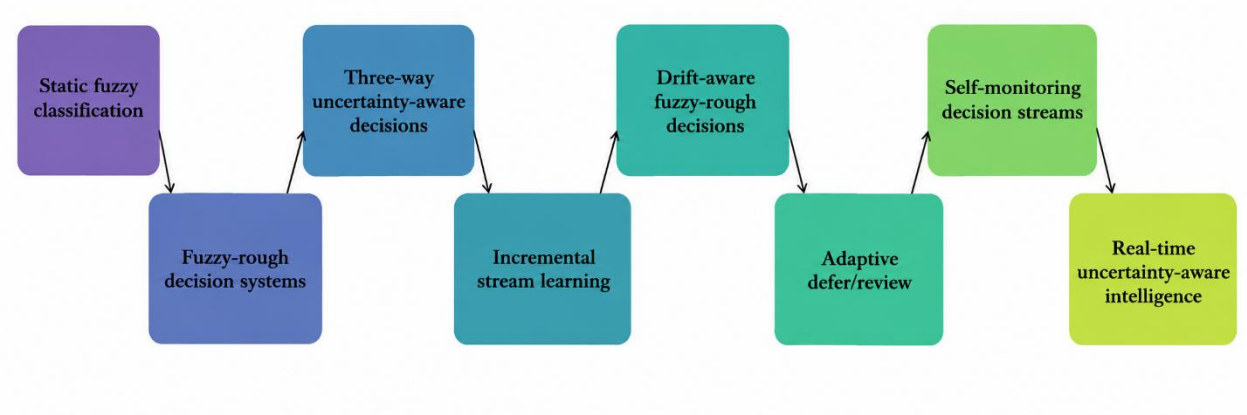



**Figure 12.** Conceptual roadmap toward 2030 for uncertainty-aware streaming decision systems.

## 6. Conclusions

This paper set out to address a coupling that is specific to evolving data streams: the data are uncertain, so forcing every instance into a binary verdict discards information about how far that verdict can be trusted, and the data are non-stationary, so the reference on which any verdict rests ages and cannot be recomputed from scratch at each step. The proposed framework, IFR-3WD-CD, meets both difficulties within one mechanism. A vaguely quantified fuzzy rough membership, maintained incrementally over a bounded, online-standardized window, supplies a noise-tolerant and non-parametric measure of how well an instance fits each class; a decision-theoretic three-way rule turns that measure into an accept, defer or reject action, with the boundary region as a deliberate, cost-weighted category rather than an error; and a drift layer keeps detection and

adaptation separate, using a change detector only to trigger selective forgetting while a bounded drift-intensity signal widens the boundary region exactly while the stream is unstable. The per-instance cost is independent of the elapsed stream length, in contrast to the quadratic total cost and unbounded memory of periodic batch reconstruction, and this is derived from the update rules and confirmed empirically. Evaluated under a strictly chronological prequential protocol that admits no future information, on controlled benchmarks that isolate abrupt, gradual, recurring and incremental drift and on two real streams, and against eight baselines including strong adaptive ensembles, the framework ranks among the leading methods on forced accuracy, being statistically indistinguishable from the best ensemble across benchmarks while adding a deferral capability the ensemble lacks, and its committed decisions are substantially more reliable than its forced ones, exceeding even the strongest baseline's forced accuracy on the real streams once the hardest instances are deferred. The ablation shows that each component earns its place, with the graded fuzzy similarity carrying much of the discrimination and the three-way rule providing a reliable confidence ordering that a binary classifier does not; the boundary region expands after drift and contracts as the model re-stabilizes; and the statistical analysis places the framework among the leading methods across benchmarks. The study also reports where the method fails, under heavy noise, aggressive resetting, and strong imbalance, so that its claims are bounded rather than universal. The central message is that, in evolving streams, uncertainty and change are best handled together: a fuzzy rough representation can express the uncertainty, a three-way rule can act on it by deferring what is truly ambiguous, and incremental maintenance with drift-aware adaptation is what makes the combination usable on a stream rather than only on a static sample. What the framework establishes, concretely, is when an uncertainty-aware boundary is useful under drift, how quickly the incremental model recovers once a concept has moved, and at what computational cost the decision structure can be kept current over an unbounded stream.

## Declarations

**Ethics approval.** This study did not involve human participants, animals, or any personally identifiable data. All experiments were conducted using publicly available benchmark datasets and synthetically generated data streams. Therefore, ethics committee approval and informed consent were not required.

**Consent to participate.** Not applicable.

**Consent for publication.** Not applicable.

**Competing interests.** The author declares no competing interests.

**Funding.** This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

**Data availability.** The data generated and/or analyzed during this study are available from the corresponding author upon reasonable request.

**Code availability.** The implementation, experiment scripts, and figure-generation code that reproduce all results are available from the author on reasonable request.

**Author contributions.** M.A.A. contributed to conceptualization, methodology, data curation, formal analysis, investigation, and visualization. A.N. contributed to literature review, data extraction, and manuscript drafting. All authors reviewed and approved the final manuscript.

**References**

Agrahari, S., & Singh, A. K. (2022). Concept drift detection in data stream mining: A literature review. Journal of King Saud University – Computer and Information Sciences, 34(10), 9523–9540.

An, S., Liu, J., Wang, C., & Zhao, S. (2021). A relative uncertainty measure for fuzzy rough feature selection. International Journal of Approximate Reasoning, 139, 130–142.

Deng, D., Shen, W., Deng, Z., Li, T., & Liu, A. (2025). An ensemble learning model based on three-way decision for concept drift adaptation. Tsinghua Science and Technology, 30(5), 2029–2047.

Elwell, R., & Polikar, R. (2011). Incremental learning of concept drift in nonstationary environments. IEEE Transactions on Neural Networks, 22(10), 1517–1531.

Guo, D., Jiang, C., Sheng, R., & Liu, S. (2022). A novel outcome evaluation model of three-way decision: A change viewpoint. Information Sciences, 607, 1089–1110.

Harries, M., Wales, N. S., & others. (1999). Splice-2 comparative evaluation: Electricity pricing.

Hu, C., Qi, Y., & Huang, X. (2026). An incremental algorithm for three-way approximations with two-dimensional variation of neighborhood decision systems. Information Sciences, 755, 123764.

Huang, Q., Tang, J., Huang, Y., & Cui, M. (2025). Dynamic dominance-based three-way decision approach in incomplete hybrid ordered data under variation of attribute values. Applied Soft Computing, 186, 114112.

Hulten, G., Spencer, L., & Domingos, P. (2001). Mining time-changing data streams. Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 97–106. https://doi.org/10.1145/502512.502529

Ji, X., Duan, W., Peng, J., & Yao, S. (2025). Fuzzy rough set attribute reduction based on decision ball model. International Journal of Approximate Reasoning, 179, 109364.

Liang, D., Wang, M., & Xu, Z. (2021). A novel approach of three-way decisions with information interaction strategy for intelligent decision making under uncertainty. Information Sciences, 581, 106–135.

Liang, P., Lei, D., Gao, X., Hu, J., & Chin, K. S. (2024). A sequential three-way decision model for classification with multilevel information gain and regret value optimization. Information Sciences, 658, 120041.

Paim, A. M., & Enembreck, F. (2025). Adaptive random tree ensemble for evolving data stream classification. Knowledge-Based Systems, 309, 112830.

Palangetić, M., Cornelis, C., Greco, S., & Słowiński, R. (2022). Granular representation of OWA-based fuzzy rough sets. Fuzzy Sets and Systems, 440, 112–130.

Qian, W., Zhou, Y., Qian, J., & Wang, Y. (2022). Cost-sensitive sequential three-way decision for information system with fuzzy decision. International Journal of Approximate Reasoning, 149, 85–103.

Shen, W., Deng, Z., Li, T., Liu, K., Xia, D., & Deng, D. (2026). A novel incremental Gaussian mixture model based on fuzzy three-way decision for concept drift adaptation. Pattern Recognition, 176, 113181.

Shu, T., Lin, Y., & Guo, L. (2024). Online hierarchical streaming feature selection based on adaptive neighborhood rough set. Applied Soft Computing, 152, 111276.

Street, W. N., & Kim, Y. (2001). A streaming ensemble algorithm (SEA) for large-scale classification. Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 377–382. https://doi.org/10.1145/502512.502568

Suárez-Cetrulo, A. L., Quintana, D., & Cervantes, A. (2023). A survey on machine learning for recurring concept drifting data streams. Expert Systems with Applications, 213, 118934.

Sun, Y., & Zhu, P. (2024). Online group streaming feature selection based on fuzzy neighborhood granular ball rough sets. Expert Systems with Applications, 249, 123778.

Theerens, A., & Cornelis, C. (2022). Fuzzy quantifier-based fuzzy rough sets. In Proceedings of the 17th Conference on Computer Science and Intelligence Systems (Annals of Computer Science and Information Systems, Vol. 30, pp. 269–278).

Yang, S., Han, M., Zhu, S., Yang, W., Dai, Z., Li, J., & Ding, J. (2025). A survey of processing methods for different types of concept drift. Data & Knowledge Engineering, 161, 102525.

Yang, X., Chen, H., Li, T., & Luo, C. (2022). A noise-aware fuzzy rough set approach for feature selection. Knowledge-Based Systems, 250, 109092.

Yao, Y. (2023). The Dao of three-way decision and three-world thinking. International Journal of Approximate Reasoning, 162, 109032.

Zhang, C., Lu, Z., & Dai, J. (2025). Incremental attribute reduction for dynamic fuzzy decision information systems based on fuzzy knowledge granularity. Information Sciences, 689, 121467.