PITFALLS OF UNLABELED DISAGREEMENT-BASED DRIFT DETECTION IN STREAMING TREE ENSEMBLES

Lara Sá Neves, Afonso Lourenço, Goreti Marreiros
GECAD, ISEP, Polytechnic of Porto, Portugal
{lspsn,fonso,mgt}@isep.ipp.pt

Lizy K. John
The University of Texas at Austin, USA
ljohn@ece.utexas.edu

ABSTRACT

Detecting concept drift in high-speed data streams remains challenging, particularly when models must operate on unlabeled data and avoid false alarms caused by benign shifts. While disagreement-based uncertainty has shown promise in neural networks, its adaptation to ensembles of incremental decision trees (IDTs) remains largely unexplored. We investigate this approach by constructing batch-specific disagreement measures via label flipping in ensemble members and evaluating their effectiveness for drift detection in tabular data streams. Our experiments show that, although this method performs well in ensembles of multi-layer perceptrons (MLPs), it consistently underperforms loss-based detectors when applied to IDTs. We attribute this behavior to the intrinsic rigidity of IDTs: learning primarily through structural expansion, with limited parameter adaptation, restricts model plasticity and prevents disagreement from reliably reflecting learning potential. Recent work on restructuring IDTs using their intrinsic decomposition into non-overlapping rules offers a promising direction for improving adaptability.

1 INTRODUCTION

Handling change in high-speed data streams is challenging due to heavy concept drifts. Effective monitoring algorithms should (R1) operate on unlabeled deployment data to detect model deterioration and (R2) resist non-deteriorating shifts with few samples. While existing data-based drift detectors perform well on unlabeled data (R1) (Xuan et al., 2021; Wan & Wang, 2021), they often generate false positives when shifts are benign (R2). Many methods track changes in classifier posterior distributions (Lindstrom et al., 2013; Lughofer et al., 2016; Lu et al., 2025), which can indicate uncertainty. However, such estimates may be unreliable when models continuously adapt to evolving streams. To address this, we propose batch-specific uncertainty, which is more practical than prequential metrics. Streaming models should focus on reliability in the current distribution rather than hypothetical generalization. Similarly to transductive reasoning, measuring how conflicting information in a batch affects the model rather than relying on accumulated past uncertainty.

A prominent example is the model disagreement framework (Yu & Aizawa, 2019; Jiang et al., 2021; Rosenfeld & Garg, 2023; Ginsberg et al., 2022). To date, it has been studied mainly with expressive neural networks trained in large batches, which struggle on tabular streams due to slow convergence, overwritten weights, and limited inductive advantage (Sahoo et al., 2017). For tabular data, ensembles of incremental decision trees (IDTs) remain state-of-the-art, leveraging fast online convergence and tree replacement via loss-based drift detectors (Bifet & Gavaldà, 2007; Gama et al., 2004), with extensions incorporating unlabeled data through self-training, unsupervised drift detection, and active learning (Gomes et al., 2025). This raises a key question: can the disagreement framework be adapted for tree-based streaming ensembles? To implement this, we exploit the fact that in binary classification, arbitrarily flipping labels for each ensemble component can create diverse, disagreeing representations, a simple yet effective way to design a true disagreeing critic (Rosenfeld & Garg, 2023; Ginsberg et al., 2022; Pagliardini et al., 2022; Chuang et al., 2020). Surprisingly, we find this strategy performs poorly across nearly all evaluated streams for ensembles of IDTs, but not multi-layer perceptrons (MLPs). We hypothesize that disagreement among IDTs fails to provide reliable signals of concept change, not due to flaws in the detection logic, but because the underlying learners lack the plasticity needed for a disagreement critic to capture their learning potential.

2 THEORY

For a stream of drifting concepts D_i, learners update θ_t incrementally to minimize risk on D_t:

$$\theta_t := \text{Alg}_t(\theta_{t-1}, \mathcal{L}_t), \mathcal{L}_t = \sum_{i=1}^{t} \mathbb{E}_{(x,y)\sim D_i}[\ell(y, h_{\theta_{t-1}}(x))]. \tag{1}$$

Storing all past data is impractical. Tree-based learners address this via approximations, e.g., incremental information gain with Hoeffding bounds, expanding only when differences between best and second-best splits are significant (Domingos & Hulten, 2000). This operation can be described as:

Lemma 1 (Incremental labeled update). For $h \in \mathcal{H}$ and history model $h_{\theta_{t-1}}$:

$$\varepsilon_{D_t}(h) = \varepsilon_{D_t}(h, h_{\theta_{t-1}}) + \varepsilon_{D_t}(h_{\theta_{t-1}}), \tag{2}$$

where $\varepsilon_{D_t}(h, h_{\theta_{t-1}})$ denotes the one-hot disagreement. However, this bound is insufficient under drifting unlabeled distributions. Since manual labeling is infeasible in true streams, meaningful error bounds need a notion of distributional distance, e.g. $\mathcal{H}\Delta\mathcal{H}$-divergence (Kifer et al., 2004).

Lemma 2 (Drift-based update). Assuming a binary hypothesis class capable of discriminating $\mathcal{D}_{t-1}$ and $\mathcal{D}_t$ (Ben-David et al., 2010), i.e., whose $\mathcal{H}\Delta\mathcal{H}$ class contains all pairwise exclusive-ors:

$$\varepsilon_{D_t}(h) \leq \varepsilon_{D_t}(h, h_{\theta_{t-1}}) + \varepsilon_{D_{t-1}}(h_{\theta_{t-1}}) + \frac{1}{2}\Delta(h_{\theta_{t-1}}), \tag{3}$$

While vacuous in practice, this bound suggests that (1) the conservative splitting and parent hyper-rectangles of $h_{\theta_{t-1}}$ act as a regularizer for h; (2) bias is minimized only if $h_{\theta_{t-1}}$ is localized around $\mathcal{D}_t$; and (3) useful drift detectors must account for both data and model complexity (Fig. 1) : if $\mathcal{D}_{t-1}/\mathcal{D}_t$ are similar, the bound is small and $h_{\theta_{t-1}}$ can be reused; otherwise, h is updated via pruning, regrowing, or ensemble modification.

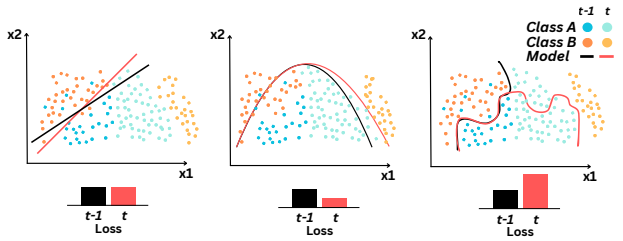


Figure 1: Drift detection across complexities: (left) loss-based false Lta on over-regularized model, (center) data-based false positive for a true matching model complexity, (right) both successful in overly complex model.

This motivates bounding error relative to the previous model rather than the entire hypothesis class, as the true labeling function y^* and drifted distribution $\mathcal{D}t$ are not adversarial. Hence, detection can exploit $\Delta(h_{\theta_{t-1}})$ with alternative hypotheses to obtain more practical bounds under drift:

Lemma 3 (Disagreement-based update). Let $h^* = \arg\max_{h' \in \mathcal{H}'} \Delta(h_{\theta_{t-1}}, h')$, $\mathcal{H}'$ per $h_{\theta_{t-1}}$:

$$\varepsilon_{D_t}(h) \leq \varepsilon_{D_t}(h, h_{\theta_{t-1}}) + \varepsilon_{D_{t-1}}(h_{\theta_{t-1}}) + \frac{1}{2}\Delta(h_{\theta_{t-1}}, h^*), \tag{4}$$

While h^* is intractable, it motivates maximizing $\Delta(h_{\theta_{t-1}}, h')$ to identify parts of the input space most affected by drift. In binary ensembles, this can be as simple as flipping labels (Rosenfeld & Garg, 2023; Ginsberg et al., 2022) (Fig. 2): under-regularized models fail to capture drift, correctly regularized models balance disagreement, and overly complex models overfit new regions. Using this discrepancy while preserving $\varepsilon_{\mathcal{D}_{t-1}}(h_{\theta_{t-1}})$ allows functional regularization, while graceful forgetting and pruning outdated nodes improve adaptability and free capacity.

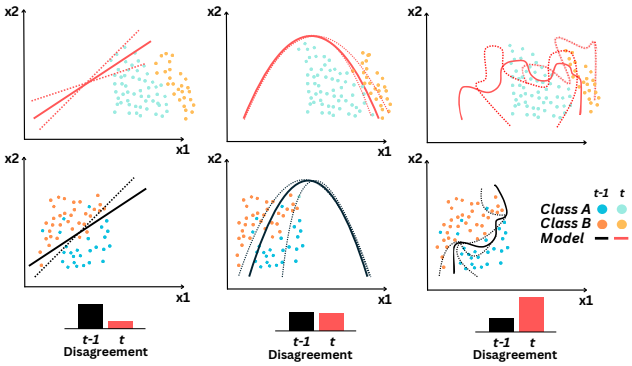


Figure 2: Disagreement-based drift across complexities: (left) hardly induced in far input space, (center) even, (right) easily induced in under-regularized far input space.

3 METHOD

For each batch, the data is split in two consecutive sub-windows, Q and R. Two copies of the ensemble g, denoted g_Q and g_R, are trained to remain consistent with past distributions P while being exposed to flipped versions of the pseudo-labeled Q and R, respectively (Fig. 3). Pairwise disagreements among base learners form the distributions D_Q and D_R, capturing the impact of new data on predictive consistency. A Kolmogorov-Smirnov (KS) test between D_Q and D_R is used to detect significant concept drift. The Q–R split naturally balances convergence and detection latency, as overly small windows may yield noisy estimates.

To achieve expressive adaptation, without relying on overly large windows that delay detection, we adopt Oza's ensemble backbone, with the Poisson parameter λ governing resampling (Oza & Russell, 2001). However, rather than using $\lambda = 1$, instances are exploited more aggressively under underfitting, using $\lambda(\epsilon) = \epsilon\lambda_{\max}$, where $\epsilon \in \langle 0, 1\rangle$ denotes the current error (Korycki & Krawczyk, 2022). Thus, accelerating convergence to more reliable estimates.

Algorithm 1 Disagreement framework

Initialize ensemble g on past data P
while *stream has new batch* **do**
  $Q', R' \leftarrow$ pseudo-label and flip in Q, R
  $g_Q, g_R \leftarrow$ copies of g
  Train g_Q on Q', g_R on R'
  for *each ensemble* $g_X \in \{g_Q, g_R\}$ **do**
    for *each pair of models* (g_a, g_b) *in* g_X **do**
      $d_{a,b} = \frac{1}{K}\sum_{i=1}^{K} I[g_a(x_i) \neq g_b(x_i)]$
    $D_X \leftarrow$ collection of all $d_{a,b}$
  if *KS_test*(D_Q, D_R) *rejects* H_0 **then**
    Drift detected

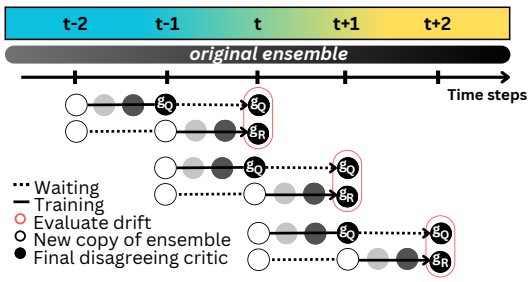


Figure 3: Windowed disagreement.

4 EXPERIMENTS

We evaluate IDT & MLP ensembles with 6 loss-based: $\text{HDDM}_{\text{A\&W}}$ (Pesaranghader & Viktor, 2016), ADWIN (Bifet & Gavaldà, 2007), PH (Mouss et al., 2004), DDM (Gama et al., 2004), EDDM (Baena-García et al., 2006); and 5 data-based: BNDM (Xuan et al., 2021), CSDDM (Wan & Wang, 2021), D3 (Sethi & Kantardzic, 2015), IBDD (Souza et al., 2020), OCDD (Gözüaçık & Can, 2021).

We use 12 synthetic streams from 7 SOA generators: SEA (rotating boundaries), Hyperplane (10 features), Stagger (feature distribution changes), Anomaly Sine (contextual drifts), RBF (centroid shifts), and Agrawal (classification changes). Each contains 90,000 instances with five 15,000-instance drifts, both abrupt and recurring. We adopt prequential evaluation and report Mean Time to Detection (MTD), Detection Accuracy (DA), and False Alarms (FA), counting alarms outside the defined detection window as false positives, with 7,500 and 9,000 instances for abrupt and gradual drifts, respectively (Fig. 4). All hyperparameters for ensembles and drift detectors, including both loss- and data-based methods, were set according to recommended ranges in the original papers and tuned using a weighted min-max normalization: $0.5 \times \text{DA} + 0.3 \times (1 - \text{FA}) + 0.2 \times (1 - \text{MTD})$. For ensembles, we use as base classifiers: Hoeffding tree (Domingos & Hulten, 2000), Hoedffing Adaptive Tree (Bifet & Gavalda, 2009), and Extremely Fast Decision Tree (Manapragada et al., 2018) for IDTs, and standard feedforward networks for MLPs, with all ensembles configured to contain 100 learners.

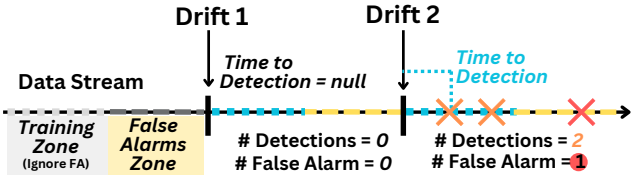


Figure 4: Evaluation metrics: Detection window.

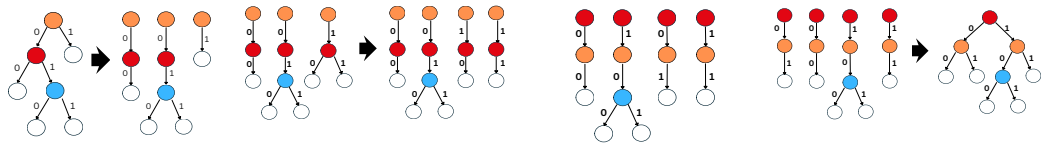
While ensembles of MLPs show good behavior, disagreement-based uncertainty from IDTs performs consistently poorly across nearly all evaluated streams (Table 1). It exhibits substantially delayed detections and, in several settings, a non-trivial number of false alarms, particularly when compared to loss-based baselines. These results indicate that disagreement signals derived from ensembles of IDTs are often too weak or too noisy to serve as reliable drift indicators.

Table 1: MTD(FA) results for gradual (G) and abrupt (A) drifts, in $\otimes$ Disagreement-based, $\diamond$ Data-based, and ∇ Loss-based detectors.

	Method	T	RBF	RBF2	SEA0	SEA1	SEA2	SineA	Sine4	SineL	Hyp0	Hyp1
$\otimes$	MLPs	G	1137(4)	1383(4)	843(3)	1475(1)	2427(1)	643(3)	2863(1)	1747(4)	1573(2)	2187(1)
		A	820(6)	910(4)	365(1)	980(0)	1620(1)	410(1)	1980(0)	810(0)	685(1)	490(0)
	IDTs	G	2267(6)	1333(15)	1167(4)	3700(2)	3300(3)	2100(2)	6600(0)	3900(3)	1533(10)	2467(2)
		A	3133(14)	2840(12)	2025(0)	1775(6)	2275(5)	1600(13)	1367(0)	1200(0)	1400(17)	2000(18)
$\diamond$	BNDM	G	321(18)	1825(22)	1938(11)	1387(17)	3221(3)	566(19)	180(17)	180(18)	2029(9)	2029(9)
		A	152(53)	89(48)	173(0)	186(3)	175(12)	65(0)	133(3)	132(3)	1893(40)	1893(40)
	CSDDM	G	244(36)	95(52)	1448(10)	1245(4)	240(31)	801(11)	851(12)	381(6)	137(24)	108(19)
		A	73(80)	86(96)	1085(4)	304(18)	156(17)	55(6)	246(14)	203(37)	565(74)	419(59)
	D3	G	1662(10)	434(16)	505(14)	486(17)	231(14)	436(5)	978(8)	421(4)	639(14)	452(12)
		A	123(5)	121(3)	728(44)	583(50)	547(40)	129(0)	129(0)	129(0)	1036(53)	1005(56)
	IBDD	G	92(60)	139(52)	254(39)	379(37)	168(39)	343(11)	158(20)	101(11)	232(31)	232(31)
		A	60(154)	60(147)	86(113)	57(100)	61(90)	59(18)	231(3)	212(0)	64(108)	64(108)
	OCDD	G	90(67)	76(71)	249(71)	249(71)	249(71)	68(71)	495(18)	208(45)	249(71)	249(71)
		A	189(182)	188(182)	149(190)	149(190)	149(190)	62(83)	138(62)	138(58)	153(190)	153(190)
∇	ADWIN	G	1333(5)	3733(1)	4017(1)	1700(0)	5117(0)	2633(1)	4650(0)	7325(0)	2083(2)	3600(2)
		A	1870(7)	1533(5)	320(2)	1680(0)	1750(2)	275(0)	500(0)	470(0)	190(6)	360(3)
	DDM	G	5752(0)	1336(0)	4052(0)	4292(0)	5160(0)	585(1)	4417(0)	3129(0)	3091(0)	3858(0)
		A	1839(4)	-(0)	486(0)	731(0)	1321(0)	317(0)	280(0)	269(0)	336(0)	1209(0)
	EDDM	G	2029(5)	4270(0)	2605(0)	1994(0)	7545(0)	515(2)	-(0)	-(0)	3063(0)	3008(0)
		A	974(37)	-(0)	429(8)	43(0)	3688(4)	622(0)	2026(0)	2004(0)	530(29)	611(8)
	$HDDM_A$	G	1037(3)	3589(3)	1648(0)	1604(1)	1857(1)	3705(0)	1144(0)	3520(1)	2995(2)	1710(0)
		A	169(4)	1781(8)	71(0)	257(3)	499(6)	5(0)	73(0)	64(0)	33(2)	61(2)
	$HDDM_W$	G	1018(3)	2518(10)	1840(0)	1008(3)	286(8)	865(0)	2247(0)	2284(0)	3358(0)	1487(5)
		A	1390(19)	878(4)	159(0)	250(2)	48(8)	186(2)	18(5)	12(7)	65(2)	57(1)
	PH	G	1363(7)	1474(3)	2060(1)	1937(0)	2884(0)	6466(1)	2062(0)	1225(0)	1434(4)	1769(3)
		A	1364(7)	1562(6)	249(1)	144(0)	184(1)	1916(0)	114(0)	103(0)	152(7)	170(8)

5 Conclusions

Taken together, our results hint at a fundamental limitation in current drift research: increasingly sophisticated model-dependent detection mechanisms cannot compensate for rigid base learners. Disagreement estimates derived from IDTs fail to provide reliable signals of concept change, not due to flaws in the detection logic itself, but because the underlying learners lack the plasticity required for uncertainty to reflect learning potential. IDTs converge quickly online thanks to their few trainable parameters, but this efficiency comes at the cost of severe rigidity. Unlike MLP systems, which adapt through both parameter updates and activation dynamics (Lourenço et al., 2025a), IDTs rely almost exclusively on irreversible structural growth driven by locally optimal split decisions, resulting in history-dependent models dominated by outdated inductive biases (Lourenço et al., 2025b). Traditional attempts to address this limitation frame plasticity primarily as capacity management, via subtree pruning (Nowak Assis et al., 2025; Manapragada et al., 2018), rather than as the ability of the current parameters to serve as a meaningful starting point for further learning. As a consequence, both disagreement-based drift detections exhibits brittle, stream-specific behavior, often failing outside narrow settings. To circumvent this, recent work on restructuring incremental decision trees with their intrinsic, non-overlapping rules (Schreckenberger et al., 2020; Heyden et al., 2024; Zhao et al., 2025) (Fig. 5) offer a promising path forward by partially relaxing this rigidity.



(a) Disconnect subtree (b) Desired branch splits (c) Move splits to root (d) Subtree rebuild

Figure 5: Restructuring IDTs with their intrinsic, non-overlapping rules that fully partition the space.

ACKNOWLEDGMENTS

Work funded by Portuguese Foundation for Science and Technology under the UT Austin Portugal Program, Ph.D. scholarship PRT/BD/18497/2024 and project doi.org/10.54499/UID/00760/2025.

REFERENCES

Manuel Baena-García, José del Campo-Ávila, Raúl Fidalgo, Albert Bifet, Ricard Gavaldà, and Rafael Morales-Bueno. Early drift detection method. In *Fourth international workshop on knowledge discovery from data streams*, volume 6, pp. 77–86, 2006.

Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*, 79:151–175, 2010.

Albert Bifet and Ricard Gavalda. Adaptive learning from evolving data streams. In *International symposium on intelligent data analysis*, pp. 249–260. Springer, 2009.

Albert Bifet and Ricard Gavaldà. Learning from time-changing data with adaptive windowing. In *Proceedings of the 7th SIAM International Conference on Data Mining*, 2007. doi: 10.1137/1.9781611972771.42.

Ching-Yao Chuang, Antonio Torralba, and Stefanie Jegelka. Estimating generalization under distribution shifts via domain-invariant representations. *arXiv preprint arXiv:2007.03511*, 2020.

Pedro Domingos and Geoff Hulten. Mining high-speed data streams. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 71–80, 2000.

João Gama, Pedro Medas, Gladys Castillo, and Pedro Rodrigues. Learning with drift detection. In *Brazilian Symposium on Artificial Intelligence*, pp. 286–295. Springer, 2004.

Tom Ginsberg, Zhongyuan Liang, and Rahul G Krishnan. A learning based hypothesis test for harmful covariate shift. *arXiv preprint arXiv:2212.02742*, 2022.

Heitor Murilo Gomes, Jesse Read, Maciej Grzenda, Bernhard Pfahringer, and Albert Bifet. Sleade: Disagreement-based semi-supervised learning for sparsely labeled evolving data streams. *IEEE Transactions on Knowledge and Data Engineering*, 2025.

Ömer Gözüaçık and Fazli Can. Concept learning using one-class classifiers for implicit drift detection in evolving data streams. *Artificial Intelligence Review*, 54(5):3725–3747, 2021.

Marco Heyden, Heitor Murilo Gomes, Edouard Fouché, Bernhard Pfahringer, and Klemens Böhm. Leveraging plasticity in incremental decision trees. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 38–54. Springer, 2024.

Yiding Jiang, Vaishnavh Nagarajan, Christina Baek, and J Zico Kolter. Assessing generalization of sgd via disagreement. *arXiv preprint arXiv:2106.13799*, 2021.

Daniel Kifer, Shai Ben-David, and Johannes Gehrke. Detecting change in data streams. In *VLDB*, volume 4, pp. 180–191. Toronto, Canada, 2004.

Łukasz Korycki and Bartosz Krawczyk. Instance exploitation for learning temporary concepts from sparsely labeled drifting data streams. *Pattern Recognition*, 129:108749, 2022.

Patrick Lindstrom, Brian Mac Namee, and Sarah Jane Delany. Drift detection using uncertainty distribution divergence. *Evolving Systems*, 4:13–25, 2013.

Afonso Lourenço, João Gama, Eric P Xing, and Goreti Marreiros. Bridging streaming continual learning via in-context large tabular models. *arXiv preprint arXiv:2512.11668*, 2025a.

Afonso Lourenço, João Rodrigo, João Gama, and Goreti Marreiros. Dfdt: Dynamic fast decision tree for iot data stream mining on edge devices. *arXiv preprint arXiv:2502.14011*, 2025b.

Pengqian Lu, Jie Lu, Anjin Liu, and Guangquan Zhang. Early concept drift detection via prediction uncertainty. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 19124–19132, 2025.

Edwin Lughofer, Eva Weigl, Wolfgang Heidl, Christian Eitzinger, and Thomas Radauer. Recognizing input space and target concept drifts in data streams with scarcely labeled and unlabelled instances. *Information Sciences*, 355:127–151, 2016.

Chaitanya Manapragada, Geoffrey I Webb, and Mahsa Salehi. Extremely fast decision tree. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1953–1962, 2018.

Hayet Mouss, Djamel Mouss, Nadia Mouss, and Linda Sefouhi. Test of page-hinckley, an approach for fault detection in an agro-alimentary production system. *2004 5th Asian Control Conference (IEEE Cat. No. 04EX904)*, 2:815–818, 2004.

Daniel Nowak Assis, Jean Paul Barddal, and Fabricio Enembreck. Behavioral insights of adaptive splitting decision trees in evolving data stream classification. *Knowledge and Information Systems*, pp. 1–32, 2025.

Nikunj C Oza and Stuart Russell. Experimental comparisons of online and batch versions of bagging and boosting. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 359–364, 2001.

Matteo Pagliardini, Prakhar Gupta, Martin Jaggi, Thomas Hofmann, and Maxim Tatarchenko. Agree to disagree: Diversity through disagreement for better transferability. *arXiv preprint arXiv:2202.04414*, 2022.

Ali Pesaranghader and Herna L Viktor. Fast hoeffding drift detection method for evolving data streams. In *Joint European conference on machine learning and knowledge discovery in databases*, pp. 96–111. Springer, 2016.

Elan Rosenfeld and Saurabh Garg. (almost) provable error bounds under distribution shift via disagreement discrepancy. In *Advances in Neural Information Processing Systems*, volume 36, pp. 28761–28784, 2023.

Doyen Sahoo, Quang Pham, Jing Lu, and Steven CH Hoi. Online deep learning: Learning deep neural networks on the fly. *arXiv preprint arXiv:1711.03705*, 2017.

Christian Schreckenberger, Tim Glockner, Heiner Stuckenschmidt, and Christian Bartelt. Restructuring of hoeffding trees for trapezoidal data streams. In *2020 International Conference on Data Mining Workshops (ICDMW)*, pp. 416–423. IEEE, 2020.

Tegjyot Singh Sethi and Mehmed Kantardzic. Don't pay for validation: Detecting drifts from unlabeled data using margin density. *Procedia Computer Science*, 53:103–112, 2015.

Vinicius M.A. Souza, Farhan A. Chowdhury, and Abdullah Mueen. Unsupervised drift detection on high-speed data streams. In *Proceedings - 2020 IEEE International Conference on Big Data, Big Data 2020*, 2020. doi: 10.1109/BigData50022.2020.9377880.

Jones Sai Wang Wan and Sheng De Wang. Concept drift detection based on pre-clustering and statistical testing. *Journal of Internet Technology*, 22, 2021. ISSN 20794029. doi: 10.3966/160792642021032202020.

Junyu Xuan, Jie Lu, and Guangquan Zhang. Bayesian nonparametric unsupervised concept drift detection for data stream mining. *ACM Transactions on Intelligent Systems and Technology*, 12, 2021. ISSN 2157-6904. doi: 10.1145/3420034.

Qing Yu and Kiyoharu Aizawa. Unsupervised out-of-distribution detection by maximum classifier discrepancy. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9518–9526, 2019.

Ruirui Zhao, Yaqian You, Jianbin Sun, João Gama, and Jiang Jiang. Online learning from drifting capricious data streams with flexible hoeffding tree. *Information Processing & Management*, 62 (6):104221, 2025.