for MLOps Feature Monitoring

Nusrat Begum[1] Thanapon Noraset[1] Suppawong Tuarob[1]
[1]Faculty of Information and Communication Technology, Mahidol University, Thailand
nusrat.beg@student.mahidol.edu

1. Motivation

- ML models in production suffer from **distribution drift**: input feature distributions shift over time, silently degrading predictions.
- Existing unsupervised detectors signal *that* drift occurred but not *which* features drifted or *what* to do about it.
- **Gap:** No unsupervised detector simultaneously provides statistical rigor, feature-level explainability, and automated remediation.

Research Question:

Can adversarial validation be extended with permutation testing and feature attribution to provide statistically sound, explainable, and actionable drift detection across diverse drift types?

2. EADD Framework

Step 1: Data Windowing
Reservoir sampling $W_{ref} = 500$
Sliding window $W_{cur} = 200$

↓

Step 2: Adversarial Classifier
LightGBM distinguishes W_{ref} vs W_{cur}
AUC-ROC > 0.7
$\Rightarrow$ suspect drift

↓

Step 3: Permutation Test
Shuffle labels $B = 50$ times
$p < 0.01 \Rightarrow$ confirm drift

↓

Step 4: Attribution & Prescription
TreeSHAP feature ranking
Classify: univariate / subset / multivariate
Automated remediation action

Key Innovation

EADD is the **first unsupervised drift detector** that provides:

1. **Statistical rigor** via permutation testing (zero false alarms)
2. **Feature-level explainability** via SHAP attribution
3. **Actionable prescriptions** mapping drift patterns to MLOps fixes

3. Experiment 1 — Temporal Drift Types

Objective: Can EADD detect all four temporal drift patterns?

Drift Type	EADD	D3
Abrupt	5/5	5/5
Gradual	5/5	0/5
Incremental	5/5	0/5
Recurring	5/5	5/5
Overall	**20/20**	**10/20**

$\Rightarrow$ EADD: **100%** detection rate D3: **50%** (fails on gradual & incremental)

4. Experiment 2 — 13 Real-World Datasets

Objective: Benchmark against D3 on diverse real-world streams.

Dataset	EADD MTD	D3 MTD
Insects Abrupt	173	152.5
Insects Gradual	9,371	9,481
Insects Incr. Abrupt	**31**	160
Insects Incr. Balanced	102	104.5
Insects Reoccurring	**131**	157
SineClusters	**139**	229
WaveformDrift2	**119**	169
Avg. (13 datasets)	**1,661**	1,725

$\Rightarrow$ **0% missed detection rate** MTD 3.7% faster than D3

5. Experiment 3 — Explainability

Objective: Does SHAP correctly identify which features drifted?

Scenario	AUC	Top Feature	Correct?
Univariate	0.801	F3 (49.7%)	Yes
Subset	0.767	F5, F2, F7	Yes
Multivariate	0.786	$\leq$14.1% each	Yes

$\Rightarrow$ **100% feature attribution accuracy** across all three drift patterns

6. Experiment 4 — False Alarm Robustness

Objective: Do statistical noise patterns cause false alarms?

Noise Type	EADD FA	D3 FA
Gaussian	0	10.4
Autocorrelated	0	87.4
Heteroscedastic	0	10.4
Correlated	0	10.0
Total	**0**	**118.2**

$\Rightarrow$ EADD: **zero false alarms** D3: up to 87.4 on autocorrelated data
Mann–Whitney $p = 0.0101$ (statistically significant difference)

Summary of Results

Metric	EADD	D3
Temporal coverage	**4/4**	2/4
Real-world MDR	**0%**	7.7%
Avg. MTD (samples)	**1,661**	1,725
Feature attribution	**3/3**	—
False alarm total	**0**	118.2

7. Conclusions & Future Work

- EADD transforms drift detection from a **binary alarm** into an **actionable diagnostic tool** with feature-level attribution and remediation prescriptions.
- **Strengths:** 100% temporal coverage, zero false alarms, correct feature identification in all tested scenarios, competitive detection speed.
- **Limitations:** Computational overhead of retraining LightGBM per monitoring cycle; prescription rules are currently heuristic.
- **Future directions:**
  - Online incremental discriminator updates
  - Feedback-loop integration with CI/CD retraining
  - Extension to multi-modal data streams

References

1. J. Lu et al., "Learning under concept drift: A review," *IEEE TKDE*, 2018.
2. O. Gözüaçık et al., "Unsupervised concept drift detection with a discriminative classifier," *ACM CIKM*, 2019.
3. B. Lukats et al., "Unsupervised concept drift detection from deep learning representations in real-time," *CIKM*, 2025.
4. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *NeurIPS*, 2017.
5. G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," *NeurIPS*, 2017.

JCSSE 2026 — 23rd International Joint Conference on Computer Science and Software Engineering · June 24–27, 2026 · Bangkok, Thailand