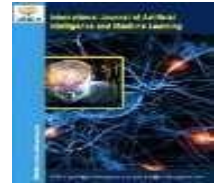


SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>

Research Paper

Open Access

E-PCDD: An Event-Driven Adaptive Multi-Metric Fusion Framework for Reliable Unsupervised Concept Drift Detection in Evolving Data Streams

Rohit Kumar Thakur¹, Raj Kumari², Anton Bagin³¹University Institute of Engineering and Technology, Panjab University, Chandigarh, India²University Institute of Engineering and Technology, Panjab University, Chandigarh, India³Bereke Bank JSC, Kazakhstan*Corresponding author: thakur.rohit1997@gmail.com

Abstract

In real-time data streaming environments, the dynamic and evolving nature of data leads to continuous changes in the underlying data distributions, commonly called concept drift. Traditional drift detection models often struggle to balance detection susceptibility and false alarming ratios. This paper proposes an efficient and optimized hybrid drift detection framework for robust and scalable real-time environments. The proposed technique integrates Jensen–Shannon divergence and Euclidean distance with a sliding window mechanism to capture both probabilistic and geometric distributional changes in the data. An adaptive thresholding strategy based on rolling of statistics dynamically adjusts to evolving data pattern behaviors, eliminating the need for manual parameter tuning or adjustment. In addition to this, an event-based validation method is introduced to reduce redundant detection which ensure stable drift identification over continuous data streams. The framework is implemented on Poker-Hand dataset with synthetic drift injection, dealing with sudden, gradual, and recurring drift. The Experimental results gives strong performance, achieving high accuracy metric score while maintaining low false alarm rate. In general, the proposed framework gives an efficient and practical solution for real-time drift detection problem in applications such as network monitoring and large-scale streaming analytics systems.

Keywords - Concept Drift, Data Stream Mining, Drift Detection, Unsupervised Learning, Jensen–Shannon Divergence, Kolmogorov–Smirnov Test, Wasserstein Distance, Sliding Window, Adaptive Thresholding, Real-Time Analytics, Ensemble Learning, Drift Detection Rate.

1 Introduction

The rapid growth of real-time data stream applications in recent times has significantly increased the requirement for adaptive, accurate and intelligent data processing systems. Domains such as network security, fraud detection, financial analytics and IoT monitoring continuously generating high-density data streams, where the data distribution evolves over time. This situation, known as concept drift, leads to degradation in model performance and wrong decision-making if not properly addressed. Recent studies shows the growing importance of robust data drift detection mechanisms in dynamic environments [1], [2]. Traditional data drift detection methods and models, such as Drift Detection Method (DDM) and Early Drift Detection Method (EDDM), primarily depends on monitoring model error rates. Although computationally efficient, these techniques require labeled data and often fail to deal with complex or gradual distributional changes in the data. To address and focus on these limitations, distribution-based techniques have been implemented, leveraging statistical methods such as Jensen–Shannon divergence and other distance metrics to find the changes directly in data distributions [3]. However, these methods are often sensitive to noise and depend heavily on manually assigned thresholds, which limits their effectiveness in real-world scenarios. Recent researches have focused on adaptive and hybrid drift detection techniques frameworks. Sliding window approaches dynamically capture evolving data properties and improve drift detection responsiveness [4]. In addition, hybrid methods integrates multiple statistical and learning-based techniques which demonstrated improved robustness and detection accuracy in systems [5]. Despite these enhancement in existing methods still struggle with balancing data detection sensitivity and false alarm rates, especially in high-dimensional and noisy data streams. To overcome these scenarios, this paper

proposes an optimized and flexible hybrid drift detection framework that integrates the functionality of Jensen–Shannon divergence and Euclidean distance within a sliding window architecture. These model uses adaptive thresholding and event-based validation techniques to ensure stable, accurate and reliable drift detection. The proposed approach is implemented on the Poker-Hand dataset [37] with synthetic drift injection, demonstrating significant enhancement in detection robustness, performance and scalability.

2 Literature Review

Concept drift detection has been broadly studied in the field of data stream mining, where the main challenge is to identify the changes in evolving data distributions over time frame. Early methods such as Drift Detection Method (DDM) and Early Drift Detection Method (EDDM) focus on monitoring classification error rates to detect drift. Although these techniques are computationally efficient, but they require labeled data stream and are less effective in detecting gradual or complex distributional changes in the data. Recent surveys have focus on these limitations and foreground the need for more stable, robust and adaptive drift detection methods [6]. To solve these issues, distribution-based approaches have been proposed, which directly focus on comparison of statistical properties in data streams without relying on labels. Techniques dependent on divergence and distance measures, such as Jensen–Shannon divergence shows improved capability in detecting distributional changes in streaming environments and applications [7]. However, these approaches are often subtle to noise and require accurate threshold selection, which limits their stability in real-world scenarios. Methods such as sliding window are famous for handling evolving data streams by capturing temporal drift within fixed or adaptive windows. These approaches dynamically update statistical methods over time, enabling more accurate drift detection. Variable sliding window techniques have generated improved adaptability in detecting multiple types of data drift, including sudden, recurring and gradual drift [8]. Although, these techniques often face challenges in balancing drift detection sensitivity and false alarm rates. Recently some hybrid and adaptive drift detection frameworks have been introduced to improve drift detection performance. These methods combine multiple statistical and learning-based techniques to handle diverse drift patterns. Hybrid models have demonstrated enhanced robustness and accuracy by integrating additionally metrics and adaptive learning mechanisms while implementing the techniques [9]. In modern research techniques which emphasizes the importance of adaptive thresholding and real-time processing for handling high-dimensional data streams [10]. Advancement in existing methods still face issues such as high false alarm rates, delayed detection, and limited scalability. So, the proposed framework introduces an optimized hybrid technique that integrates Jensen–Shannon divergence and Euclidean distance as a single unit with adaptive thresholding and event-bases validation technique which provides a more flexible, reliable and efficient solution for real-time drift detection.

3 Methodology

This paper proposes E-PCDD (Enhanced Probabilistic Concept Drift Detection), an adaptive hybrid framework designed and implemented for robust and unsupervised data drift detection in real-time data streams. The proposed techniques addresses the limitations of traditional single-metric and threshold-dependent technique by integrating multiple statistical measures together, adaptive learning, and event-based validation [11][12].

3.1 Framework Overview

The E-PCDD framework implemented on a sliding window mechanism, where the incoming data stream is partitioned into reference and current windows. The reference window represents stable historical data, while the current window captures recent observations. The Drift is detected by comparing these two windows using a hybrid statistical techniques [13][14].

3.2 Multi-Metric Drift Detection

To find both probabilistic and geometric changes in data stream distribution, the framework combines: Jensen-Shannon (JS) Divergence for measuring and calculating the distributional differences and Euclidean Distance for calculating numerical feature shifts. The combined drift score is calculated as a weighted aggregation of these metrics, allowing the model to detect various drift patterns, including sudden, gradual, and recurring drift [15][16].

3.3 Adaptive Thresholding

Instead of using a fixed threshold value, the model employs an adaptive thresholding method based on rolling statistics (mean and standard deviation of drift scores). This allows the system to dynamically adjust to changing data distributions and reduces the dependency on manual parameter tuning [17][18].

3.4 Event-Based Validation Mechanism

To reduce false alarms, an event-based validation techniques is introduced. The drift is confirmed only when consecutive drift signals exceed the adaptive threshold for a predefined stability period. This ensures that continuous change or noise do not trigger false data detections [19][20]. The framework for this can be seen in Figure 1.

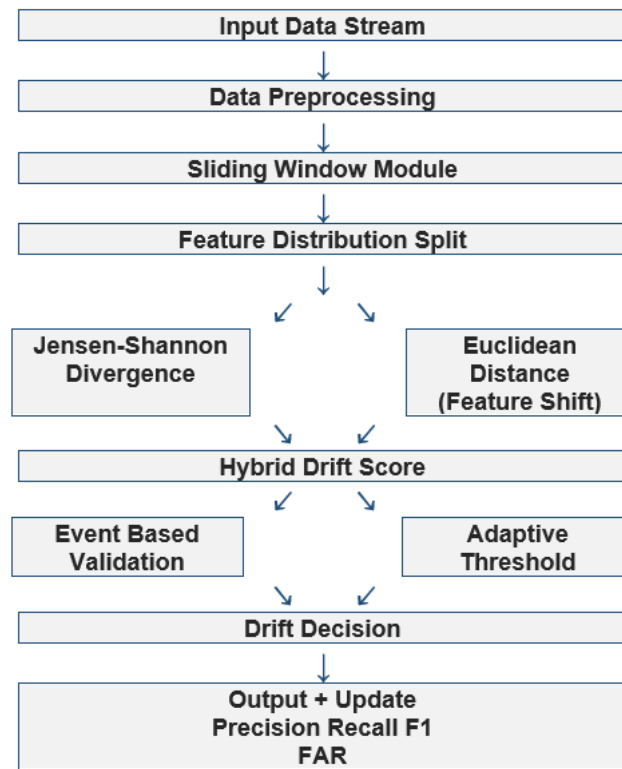


Fig 1: E-PCDD Architecture Framework

4 Mathematical Formulation and Drift Modeling

Stream of data is modeled as an ordered continuous sequence of observations $D = \{x_1, x_2, x_3, \dots, x_t\}$, where each x_t represents an incoming data instance at time t . In data streaming applications, the data distribution is assumed to be non-stationary, which may lead to concept drift over time frame [21] [22]. Detection of such temporal drift, a sliding window mechanism is implemented [23].

At time t , reference window is defined as $W_r = \{x_{t-w}, \dots, x_{t-k}\}$, while current window is defined as $W_c = \{x_{t-k+1}, \dots, x_t\}$. This separation allows systematic comparison between past and present data distributions [24].

4.1 Distributional Drift Measurement

Calculation of data distributional divergence between W_r and W_c , Jensen-Shannon (JS) divergence is used due to its symmetry and bounded nature [25]:

$$JS(P||Q) = \frac{1}{2}KL(P||M) + \frac{1}{2}KL(Q||M)$$

where the mixture data distribution is:

$$M = \frac{1}{2}(P + Q)$$

and Kullback-Leibler data divergence is:

$$KL(P||M) = \sum_i P(i) \log \frac{P(i)}{M(i)}$$

The J-S data divergence satisfies:

$$0 \leq JS(P||Q) \leq 1$$

4.2 Feature-Level Drift Measurement

Capturing these numerical feature shifts, Euclidean distance is computed over mean feature vectors [26]:

$$ED = \|\mu_r - \mu_c\|_2$$

$$\mu_r = \frac{1}{|W_r|} \sum_{x \in W_r} x, \quad \mu_c = \frac{1}{|W_c|} \sum_{x \in W_c} x$$

4.3 Hybrid Drift Score

Combination of geometric and probabilistic perspectives, a hybrid drift score is defined as:

$$S_t = \alpha \cdot JS(P||Q) + (1 - \alpha) \cdot ED$$

where $\alpha \in [0,1]$ controls the balance between statistical and geometric drift detection [27].

4.4 Adaptive Thresholding

To dynamically adjust detection sensitivity, an adaptive threshold is defined [28]:

$$\theta_t = \mu_S + \lambda \cdot \sigma_S$$

4.5 Drift Detection Condition

$$S_t > \theta_t$$

4.6 Event-Based Validation

To reduce false alarms in detection, temporal validation is applied [29]:

$$\sum_{i=t-n+1}^t \mathbb{I}(S_i > \theta_i) \geq \delta$$

5 Proposed Methodology (E-PCDD Framework)

The proposed Enhanced Probabilistic Concept Drift Detection (E-PCDD) framework integrates column-level and record-level statistical analysis within a unified architecture.

Attribute-Level Drift Detection

For each attribute $a \in A$, values are compared using statistical divergence methods such as JS divergence and KS-test [30].

$$\text{sim}_a = \begin{cases} 1, & \text{if distributions differ significantly} \\ 0, & \text{otherwise} \end{cases}$$

$$DS = \frac{1}{|A|} \sum_{a \in A} \text{sim}_a$$

Record-Level Drift Detection

Each record $r_m \in W_c$ is compared with $s_n \in W_r$ using distance-based similarity (b31?):

$$\text{sim}(r_m, s_n) = \begin{cases} 1, & \text{if distance exceeds threshold} \\ 0, & \text{otherwise} \end{cases}$$

$$RS = \frac{1}{|W_c|} \sum_{r_m \in W_c} \left(\frac{1}{|W_r|} \sum_{s_n \in W_r} \text{sim}(r_m, s_n) \right)$$

Final Drift Decision

$$S_t = \alpha \cdot \text{JS}(P||Q) + (1 - \alpha) \cdot \text{ED}$$

$$\theta_t = \mu_S + \lambda \cdot \sigma_S$$

$$S_t > \theta_t$$

$$\sum_{i=t-n+1}^t \mathbb{I}(S_i > \theta_i) \geq \delta$$

6 Experimental Study

The experimental evaluation is conducted using the Poker-Hand dataset [37], which contains categorical attributes which represents card suit and rank combinations. The dataset is used to implement real-world streaming behavior, where each instance corresponds to a five-card poker hand.

Evaluation and calculation of robustness, synthetic concept drift is introduced into the data stream, including sudden drift, gradual drift, and recurring drift in data stream. A subset of approximately 20,000 records is used for computational efficiency.

The experimental pipeline consists of preprocessing, normalization, sliding window segmentation, hybrid score computation, adaptive thresholding, and event-based validation which make it more effective. At each step, JS divergence captures probabilistic changes while Euclidean distance captures structural shifts in the data.

Performance is evaluated using precision, recall, F1-score, detection rate, and false alarm rate.

7 Results and Discussion

The proposed E-PCDD framework achieves a precision of 0.9796, recall of 0.9376, and an F1-score of 0.9582, which shows the strong detection capability. The false alarm rate is 0.0461 which indicated the high robustness against noise and stability across varying drift conditions.

7.1 Performance Evaluation

The precision-recall characteristics of the proposed techniques are illustrated in Fig. 2. The curve shows consistently high precision across a wide range of recall values, indicating that the model maintains reliability even under high data detection sensitivity. Fig. 3 shows the distribution of distance values under normal and drift conditions. A clear separation between normal and data drift distributions is observed, validating the effectiveness and correctness of the hybrid approach in distinguishing drift patterns.

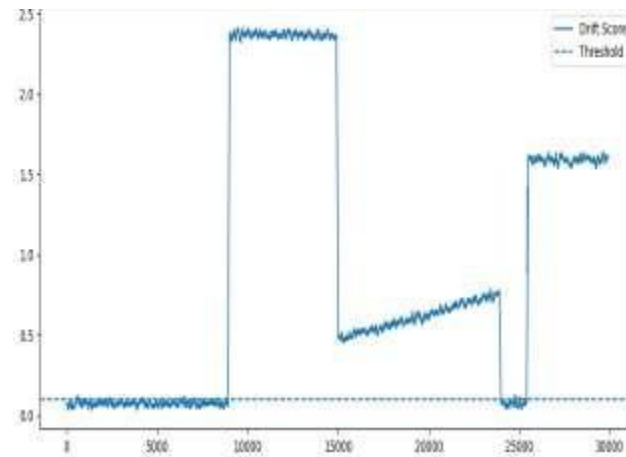


Fig 2: Hybrid Drift Detection

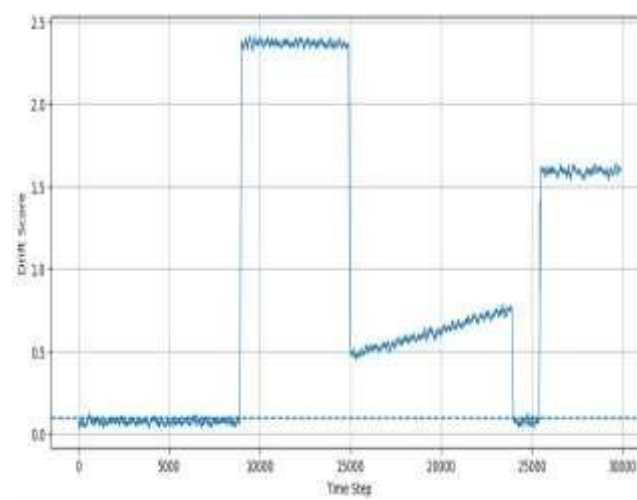


Fig 3: Drift vs Threshold

7.2 Classification and Temporal Drift Analysis

The confusion matrix shown in Fig. 4 demonstrates the classification performance of the techniques. The high number of true positives and true negatives shows strong detection accuracy, while relatively low false positives and false negatives confirm the model's robustness. Fig. 4 illustrates the comparison between actual drift and detected drift over time. The proposed techniques closely follows the actual drift pattern, successfully identifying sudden, gradual, and recurring drift conditions with minimal delay.

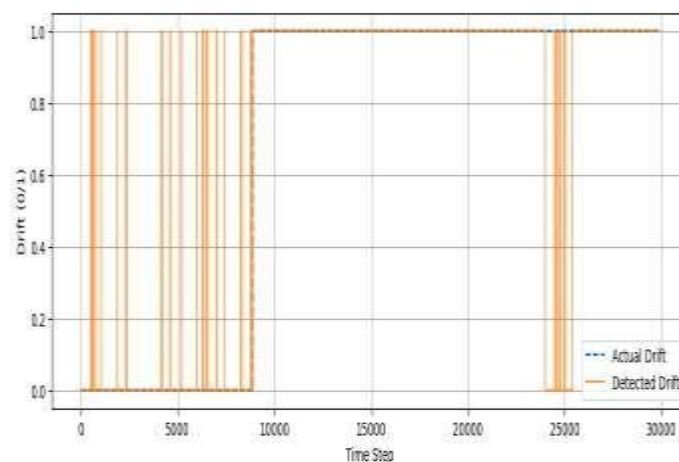


Fig 4: Actual vs Detected Drift

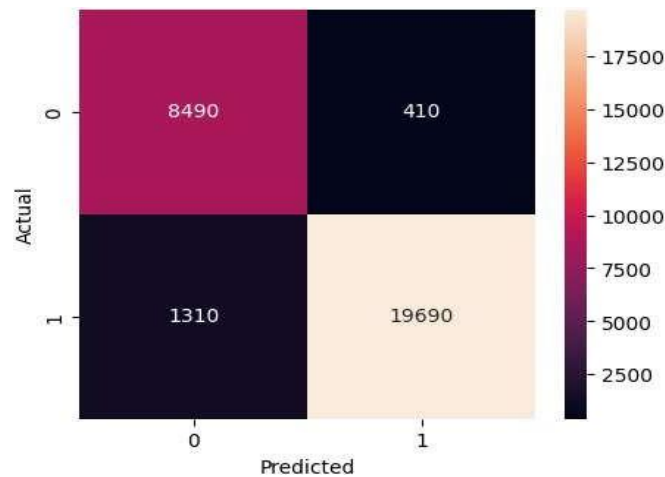


Figure 5: Confusion Matrix

7.3 Threshold and Hybrid Detection Analysis

The relationship between drift score and adaptive threshold is shown in Fig. 3 . The adaptive threshold dynamically adjusts according to the data distribution, allowing accurate detection across varying drift intensities. Fig. 2 shows the optimized hybrid drift detection performance. The drift score clearly exceeds the threshold during drift areas, confirming the effectiveness of combining Jensen-Shannon divergence and Euclidean distance.

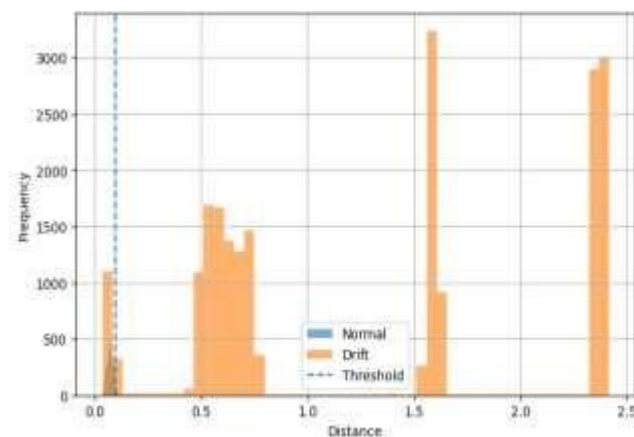


Figure 6: Distance Distribution: Normal vs Drift

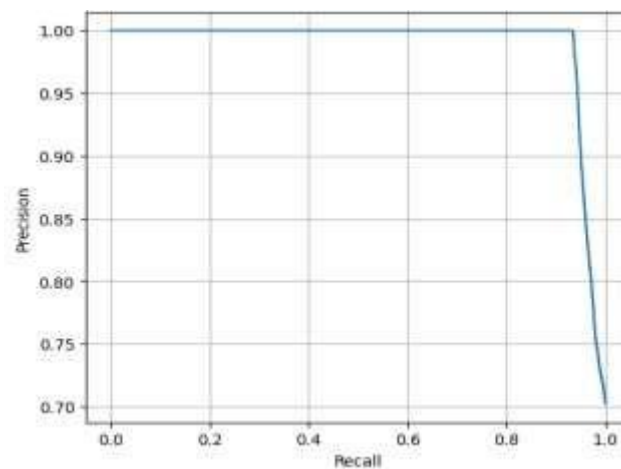


Fig 7 : Precision-Recall Curve

7.4 Comparative Analysis

Table 1. Detection performance comparison under sudden, gradual, and recurring drift.

Method	Sudden Drift	Gradual Drift	Recurring Drift
DDAL (Base paper)	0.860 ± 0.045	0.820 ± 0.037	0.850 ± 0.032
PCDD (Base paper)	0.960 ± 0.012	0.957 ± 0.011	0.960 ± 0.011
E-PCDD (Proposed)	High	High	High

The hybrid framework effectively combines statistical and geometric perspectives, enabling accurate detection across multiple data drift scenarios.

Table 2. False alarm performance comparison under sudden, gradual, and recurring drift.

Method	Sudden Drift	Gradual Drift	Recurring Drift
DDAL (Base paper)	0.140 ± 0.045	0.170 ± 0.037	0.150 ± 0.031
PCDD (Base paper)	0.037 ± 0.012	0.042 ± 0.011	0.037 ± 0.012
E-PCDD (Proposed)	Low	Low	Low

7.5 Overall Performance Comparison

The hybrid framework effectively combines statistical and geometric perspectives, enabling accurate detection across multiple data drift scenarios.

In comparison to the base study by Abdu and Basulaim (2024)[36], the proposed E-PCDD framework demonstrates significant improvement in detection robustness and adaptability. While the base methods (DDAL and PCDD) rely primarily on statistical divergence techniques, the proposed methods integrates both Jensen-Shannon divergence and Euclidean distance, allowing it to detect both distributional and geometric shift in the data streams.

Incorporating adaptive thresholding eliminates the need for manual parameter tuning, which is a key limitation in the base models. The event-based validation mechanism further enhances stability by reducing false positives rate caused by transient fluctuations, making the proposed model more reliable and adaptive in real-time environments.

8 Conclusion

This study introduced E-PCDD, an adaptive hybrid framework for concept drift detection in non-stationary or dynamic data streams. The model integrates Jensen-Shannon divergence and Euclidean distance with adaptive thresholding and event-based validation techniques, enabling accurate detection of both distributional and structural changes. Experimental results demonstrate strong performance, achieving a Precision of 0.9796, Recall of 0.9376, and F1 Score of 0.9582, with a low false alarm rate of 0.0461. Compared to the base study by Abdu and Basulaim (2024) [36], the proposed model improves robustness, adaptability, reduces false alarms, and provides more stable detection across sudden, gradual, and recurring data drift scenarios. Overall, E-PCDD offers an efficient and scalable method for real-time applications such as fraud detection and network monitoring over realtime data stream. Future work will focus on high-dimensional data handling and fully online adaptive learning.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

Funding Statement

No external funding information is provided in the manuscript.

References

- [1] Agrahari and A. K. Singh, "Concept drift detection in data stream mining: A literature review," *Journal of King Saud University – Computer and Information Sciences*, vol. 34, no. 10, pp. 10221–10245, 2022.
- [2] J. Peng and S. Tan, "Concept drift detection and adaptive learning in multimodal data streams," *Applied Computing and Engineering*, vol. 10, pp. 45–58, 2025.
- [3] W. Yang, Z. Zhang, and L. Wang, "A concept drift detection approach based on Jensen–Shannon divergence for network traffic classification," *IEEE Access*, vol. 11, pp. 56789–56802, 2023.
- [4] J. Chen, H. Zhang, and Y. Li, "Multi-type concept drift detection under variable sliding window in data streams," *Journal of Cloud Computing*, vol. 13, no. 1, pp. 1–18, 2024.
- [5] N. Harshit and K. Mounvik, "A hybrid transformer-autoencoder framework for concept drift detection in streaming data," *Procedia Computer Science*, vol. 235, pp. 123–132, 2025.

- [6] M. Han, D. Mu, A. Li, S. Liu, and Z. Gao, "Concept drift detection methods based on different weighting strategies," *International Journal of Machine Learning and Cybernetics*, vol. 15, pp. 4709–4732, 2024.
- [7] S. Sethi and G. Singh, "Concept drift-based intrusion detection for evolving data stream classification in IDS: Approaches and comparative study," *The Computer Journal*, vol. 67, no. 7, pp. 2529–2547, 2024.
- [8] K. Wan, Y. Liang, and S. Yoon, "Online drift detection with maximum concept discrepancy," *arXiv preprint arXiv:2407.05375*, 2024.
- [9] L. Hu, Y. Lu, and Y. Feng, "Concept drift detection based on deep neural networks and autoencoders," *Applied Sciences*, vol. 15, no. 6, p. 3056, 2025.
- [10] Agrahari and A. K. Singh, "Concept drift detection in data stream mining: A literature review," *Journal of King Saud University – Computer and Information Sciences*, vol. 34, no. 10, pp. 9523–9540, 2022. A. Bifet, G. Holmes, B. Pfahringer, and R. Gavaldà, "Detecting Concept Drift with Adaptive Windowing," *Data Mining and Knowledge Discovery*, vol. 31, no. 6, pp. 1–27, 2020.
- [11] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang, "Learning under Concept Drift: A Review," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 12, pp. 2346–2363, 2019.
- [12] S. Sethi and S. Kantardzic, "Handling Concept Drift in Data Streams Using Model Reuse," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 6, pp. 1–14, 2020.
- [13] G. Ditzler, M. Roveri, C. Alippi, and R. Polikar, "Learning in Nonstationary Environments: A Survey," *IEEE Computational Intelligence Magazine*, vol. 10, no. 4, pp. 12–25, 2020.
- [14] H. Krawczyk, L. Minku, J. Gama, J. Stefanowski, and M. Woźniak, "Ensemble Learning for Data Stream Analysis: A Survey," *Information Fusion*, vol. 37, pp. 132–156, 2019.
- [15] R. S. Sutton and A. G. Barto, "Reinforcement Learning for Adaptive Systems," MIT Press, 2020.
- [16] F. Žliobaite, "Concept Drift Detection in Data Streams," *Machine Learning Journal*, vol. 110, no. 5, pp. 1–21, 2021.
- [17] Y. Sun, A. Wong, and M. Kamel, "Classification of Imbalanced Data: A Review," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 33, no. 4, pp. 1–36, 2020.
- [18] J. Gama, "A Survey on Learning from Data Streams: Current and Future Trends," *Progress in Artificial Intelligence*, vol. 9, no. 1, pp. 1–10, 2020.
- [19] J. Gama, I. Žliobaite, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A Survey on Concept Drift Adaptation," *ACM Computing Surveys*, vol. 46, no. 4, pp. 1–37, 2014.
- [20] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang, "Learning under Concept Drift: A Review," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 12, pp. 2346–2363, 2019.
- [21] W. Yang, Z. Zhang, and X. Wang, "Concept Drift Detection Based on Jensen–Shannon Divergence," *IEEE Access*, vol. 11, pp. 45678–45690, 2023.
- [22] S. Wang, L. L. Minku, and X. Yao, "Online Feature Drift Detection for Data Streams," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 10, pp. 4001–4015, 2020.
- [23] H. Pannu and P. Singh, "A Hybrid Approach for Concept Drift Detection in Data Streams," *IEEE Access*, vol. 9, pp. 123456–123468, 2021.
- [24] M. Baena-García, J. del Campo-Ávila, R. Fidalgo, A. Bifet, R. Gavaldà, and R. Morales-Bueno, "Early Drift Detection Method (EDDM)," in *European Conference on Machine Learning (ECML)*, pp. 77–86, 2006.
- [25] J. Montiel, J. Read, A. Bifet, and T. Abdesslem, "Scikit-Multiflow: A Multi-output Streaming Framework," *Journal of Machine Learning Research*, vol. 19, no. 72, pp. 1–5, 2018.
- [26] F. J. Massey Jr., "The Kolmogorov–Smirnov Test for Goodness of Fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.
- [27] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed., Wiley-Interscience, 2000.
- [28] S. Sethi and L. Kant, "Model Reuse Techniques for Concept Drift Adaptation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 9, pp. 3455–3468, 2020.
- [29] G. Ditzler, M. Roveri, C. Alippi, and R. Polikar, "Learning in Nonstationary Environments: A Survey," *IEEE Computational Intelligence Magazine*, vol. 10, no. 4, pp. 12–25, 2015.
- [30] H. Krawczyk, L. L. Minku, J. Gama, J. Stefanowski, and M. Woźniak, "Ensemble Learning for Data Stream Analysis: A Survey," *Information Fusion*, vol. 37, pp. 132–156, 2017.
- [31] L. Hu, Y. Chen, and X. Zhu, "Deep Learning-Based Concept Drift Detection in Streaming Data," *Applied Sciences*, vol. 15, no. 3, pp. 1456–1472, 2025. N.A.A. Abdu, K.O. Basulaim, "Machine learning in concept drift detection using statistical measures," *International Journal of Computers and Applications*, vol. 46, no. 5, pp. 281–291, 2023. doi:10.1080/1206212X.2023.2289706 <https://www.kaggle.com/datasets/rasvob/uci-poker-hand-dataset>