Bol. Soc. Paran. Mat.
©SPM – E-ISSN-2175-1188
SPM: www.spm.uem.br/bspm

(3s.) **v. 2026 (44) 2** : 1–12.
ISSN-0037-8712
doi:10.5269/bspm.80397

DAMM-JSD: A Novel Framework for Handling Missing Values in Concept Drift Detection for Streaming Data

Ushashree P. and R. B. V. Subramanyam

ABSTRACT: Detection of concept drift in streaming data is crucial task especially in the presence of missing values.In this paper a novel approach DAMM-JSD using Jensen-Shannon divergence,which is capable of robustly identifying distributional changes under different mechanisms of missing values such as MCAR,MAR,MNAR. Experimentation has been done on different benchmark datasets using different evaluation metrics and the proposed method DAMM-JSD has performed significanlty better to the existing base methods such as ADWIM and PH. The statistical significance of the method has been validated using Friedman and Nemenyi test. The method has the capability of identifying drift in an incomplete data streams and has greater time complexity compared to base methods.DAMM-JSD provides an robust and accurate solution for detecting concept drift and to address evolving data streams with inherent missing values

Keywords: ADWIN, data stream, concept drift, drift detection, missing values, Jensen–Shannon divergence, drift adaptation, imputation method.

Contents

1. Introduction

Concept drift is a term used to describe the change in the statistical properties of the dependent variable over a period of time,that would affect the performance of the models.This concept is quite probable to happen in non stationary environments where the data distribution is dynamic.This change in the properties of data would effect the relation between the dependent and independent variable and could lead to model degradation and effect the prediction accuracy. As the data distribution emerges, it would be difficult to identify the changing patterns in the models that have been trained on stationary dataset and would lead to degradation of the performance.To maintain predictive accuracy over time its important to maintain continuous monitoring and adaption of the model.

Concept drift is especially hard to address, since the model needs to be adjusted to new information and not overfit to the new data, and may degrade stability and the ability to make generalizations. In order to adapt, the models should react to the evolving data whilst maintaining what has been learned previously, through methods of means that respond to drift without deteriorating performance.The concept drift can be classified under different categories, such as gradual, abrupt, recurring, and mixed, each one requires a unique adaptation strategy. Gradual drift is the slow continuous changes and it can be handled by novel parameter updates. Sudden drift, which is frequently occasioned by external factors, can involve retraining on new data. Common in seasonal patterns, recurring drift is a problem where memory-based techniques are beneficial, but mixed drift is an even more complicated concept, since multiple types of drift are mixed. The adaptive models should be designed with a lot of care in their nature and rate of drift, the nature of the data stream and the needs of the application.

2020 *Mathematics Subject Classification:* 68T05, 62H30, 68T10.
Submitted November 27, 2025. Published March 14, 2026

As a matter of fact, the online learning systems are even more complicated by the absence of values. Generally, incompleteness in data streams occurs in general data streams as a result of numerous factors that may include machine failures, network transmission failures or random anomalies. These gaps influence not only the quality of the data but also the anomaly detection and concept drift detection since the missing value will skew the distributions, it will be hard to detect. The estimation of bias parameters and the reduction of the volume of useful data are also possible, the ability of a model to learn complex structures and make accurate predictions is limited.The ability to handle missing data is therefore a strong aspect of reliability in real time online learning systems. The fact that these systems need to change with the variation in data means that missing values may interfere with the learning process, and negatively impact performance. Imputation methods or direct imputation of missing data should therefore be effective, impartial and be able to offer consistent adjustment to concept drift.

In addition to technical constraints, adaptive systems must also be of practical ability like limited memory, processing speed or bandwidth. This is particularly significant in fields such as the IoT networks, financial markets and healthcare monitoring, where a great deal of streaming data needs to be processed in real-time. The joint problem of concept drift and missing data in these environments requires effective as well as lightweight and efficient solutions. The future work would therefore focus on coming up with adaptive techniques that would be able to handle the evolving data and incomplete information yet remain swift, scalable, and resource-efficient.

The correlation of concept drift with missing data also indicates that there is need to have a coherent method that all three process of detecting, adapting and data quality. New methods must exploit the relationship between drift detection and missing data instead of considering the two as different issues. To take some examples, risks of drift can be reduced by having better methods of filling in missing data, whereas the impact of missing data on model learning can be reduced by having good drift adaptation. This combination of strategies will be necessary to develop systems that will remain accurate, stable, and reliable in streaming data environments in the real world.

The problem of missing values in datasets can be solved in a number of ways. Simple methods like mean or mode imputation are common, and more complex methods such as multiple imputation by chained equations (MICE), MissForest and generative adversarial imputation networks (GAIN) are also possible. All these approaches have strengths and weaknesses, based on the character of the voided data and the situation of use. Although these methods have enhanced the quality of the data in most instances, it is an open challenge to combine the missing value treatment with adaptive learning in the presence of concept drift, especially in streaming applications.

The concept drift detection and adaptation is outlined through an extensive survey which is found in [1] including the evaluation techniques and classifying the algorithms based on the various kinds of drifts. Likewise, [2] proposed an incremental learning algorithm which is built on the evolving type-2 recurrent fuzzy neural networks, which can dynamically modify classifier in the streaming environment and exhibit a better adaptability to the changing data distributions.

[3] bring out real world applications of concept drift learning which include fraud detection, network security, and user profiling. They also discuss issues with implementing systems sensitive to drift and discuss appropriate adaptation methods, but the technical integration with missing values is not comprehensive. [4] discuss the concept drift based on data-driven decision support and provide the relevant adaptive learning architecture to run streaming data and present real-time case studies, but missing values issues are usually not considered in the framework.

A collection of drift adaptation and streaming learning algorithms are reported in both [5] and [6] and a variety of detection methods, adaptation methods and benchmark techniques are evaluated in both. [7] report an evolving type-2 fuzzy classifier in the continuous data case, with enhanced resistance to non-stationary conditions, and streaming example pclass is also presented in [8] with demonstrated efficient online adaptation. In the article by Ahmad et al. (2011), the Muse-RNN model presented in [9] demonstrates the advantages of self-evolving neural networks in drift adaptation as it presents evidence of structural adaptability to data streams.

[10] use fuzzy clustering to improve adaptive regression of drifting streams, based on the full availability of data, which is optimal clustering, whereas [11] infer fault detection to missing data using fuzzy stochastic systems, [12] lay the statistical foundation for missing data analysis, reviewing imputation and

inference techniques and GAIN [13] is a generative adversarial method of missing data imputation, which works well with batch data, but is not designed to work in adaptive or streaming settings.

MICE algorithm [14] and MissForest [15] algorithm are both popular to use flexible and accurate imputation, but both are effectively geared towards batch processing, not online or evolving data streams. [16] also present a temporal regularized matrix factorization, which can predict time-varying data better when it is surrounded by temporal dependencies, although adaptation to drift or missingness is not developed.

Recommendation systems having bias correction techniques are described in [17] and methods of review classification in which missing data are supported at the expense of robustness, but do not rely on evolving data or drift, are described in [18]. Denoising autoencoders [19] and MIDA [20] use deep learning to perform robust imputation, both showing good performance in a range of applications but in general do not consider how to address incomplete data. [21] and [22] give practical advice on analyzing missing data, but do not address the problem of incomplete data.

[23] proposed a drift detection mechanism via k-means space partitioning, which demonstrated improvement of accuracy in drift detection for streaming data, albeit under the assumption of fully observed data. [24] and [25] extend statistical methodologies for density estimation and change detection, yet omit strategies for handling incomplete data.Techniques such as regional density dissimilarity [26], kernel two-sample tests [27], QuantTree histograms [28], and incremental Kolmogorov-Smirnov tests [29] advance drift or change detection, but do not accommodate missing values in their primary designs.

[30], present a drift detection approach incorporating fuzzy distance estimations to address missing values in streaming data. This method leverages fuzzy logic to estimate distances between data segments, thereby improving drift detection in the presence of incomplete information. Their experiments indicate enhanced accuracy and adaptability over conventional methods. However, the technique's effectiveness depends on the selection and tuning of fuzzy parameters, which may affect performance across different data types and missingness patterns.

[31] discusses about various online strategies for handling drift in data streams further an adaptive optimization problem for adapting with missing features have been proposed by [32] but is computationaly heavier than the traditional methods.Author of [33] proposed an ensemble learning which resileient against missing data and adversial attacks but has been tested on time series forecasting itself.Refidiff is a diffusion based imputation method [34] that works on high dimensional data but lacks adaptation of drift.DMDI(deep missing-data imputation) is a framework proposed by [35] that is useful for intrution detection but is computationally costly and limited to cyber security datasets only.

Based on the comprehensive literature survey, it is evident that existing research in concept drift detection and missing value handling has primarily focused on these challenges in isolation. Most state-of-the-art drift detection algorithms assume complete data availability, while widely adopted imputation techniques are designed for static or batch datasets and fail to adapt effectively under evolving distributions. The absence of unified frameworks, limited adaptivity in real-world streaming scenarios, and susceptibility to distortion by missing values highlight critical gaps in this field. As a result, current solutions often fall short when confronted with the dual challenges of concept drift and incomplete data in dynamic environments.

To address these limitations, we propose DAMM-JSD, a unified and adaptive approach that integrates drift detection and missing value imputation within a single framework. By leveraging distribution-aware adaptive imputation and Jensen-Shannon Divergence, DAMM-JSD enables robust drift detection and reliable data completion in real-time streams, even under varying missingness mechanisms. This method offers significant improvements in accuracy, stability, and generalizability over existing benchmarks, paving the way for practical deployment in complex, non-stationary data environments

2. Methodology

In this paper we have proposed a method DAMM-JSD i,e Distribution Aware missing value handling with Jensen-Shanon divergence ,its a technique especially designed to handle the issue of missing values in streaming data especially in the presence of concept drift. Generally the traditional drift detection methods such as ADWIN,EDDM,DDM assume an ideal scenario such that the data streams are complete and directly focus on identifying the distributional changes only but here in our approach we incorporate missing value handling within drift detection process. Existing imputation techniques, including mean, mode, and PH, are typically static and lack adaptivity to evolving data distributions, often resulting in biased estimates and degraded performance in dynamic settings.

DAMM-JSD (Drift Adaptation for Missing data using Jensen–Shannon Divergence), a unified framework for concept drift detection in streaming data with missing values. DAMM-JSD integrates three core components:

Drift Detection : Jensen–Shannon Divergence (JSD) is computed between reference and current data windows to quantify distributional changes. When the JSD calculated surpasses a set threshold, a concept drift is identified.

Missing Data Handling :DAMM-JSD does not discard data with missing values, but, depending on the online imputation or estimation, makes sure that incomplete data is not wasted.

Adaptation :When drift is detected, the underlying base learner (e.g. Adaptive random Forest or AdaBoost) is incrementally updated or retrained. The adaptation enables the model to adapt fast to new concepts in the data stream.

DAMM-JSD is able to mitigate these threats by incorporating a statistical divergence-based imputation plan into the drift-detection process itself. The approach constantly models the distributions of observed attributes of past and present windows of data stream. Jensen Shannon divergence (JSD) is used to measure the changes across the windows in DAMM-JSD, as it satisfies the properties of symmetricity, boundedness and stability which allows it to be used to measure changes in probabilistic distributions. Through the collective consideration of distributional shifts and the availability of missing data, DAMM-JSD allows robust and flexible identification of concept drift as well as has missing values imputed in a way that is consistent with the changing stream of data.

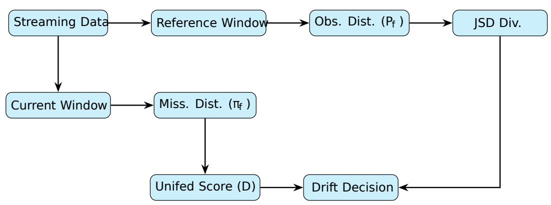


Figure 1: DAMM-JSD architecture for unified drift detection and missing value handling.

Figure 1 shows us the structure of the suggested DAMM-JSD framework that will smoothly incorporate the concept drift detection with missing values modeling in the streaming data contexts. The system is based on the streaming data separation into two major segments: a reference window, containing the historically constant data, and a current window, comprising of the most recent observations. The framework evaluates two aspects that are complementary (the distribution of values of an observed attribute and the patterns of missingness across features) on a per-window basis.

The distribution of observed features is obtained out of the reference window which offers a stationary profile of the data. At the same time, the analysis of the current window provides an approximation of the distribution of the missing values which is dynamic in the context of incompleteness of the data in real-life situations. The distributions are used to calculate measures of divergence, and JensenShannon Divergence(JSD) is used to measure the statistical distance between the past and present information. JSD has the benefits of symmetry and boundedness, which makes it especially applicable to detecting changes in distributional characteristics of changing data streams.

The discrepancies calculated on both the noticed data and the missingness are then combined into a single metric which summarizes the total extent of alteration of the data stream. This rating is compared with a statistically established limit. The occurrence of concept drift is detected by the framework and an adaptation phase is initiated when the unified score goes beyond the threshold. In the process of adaptation, DAMM-JSD dynamically fills the missing values with distribution-wise estimators based on the most topical and current values, which guarantees the integrity and consistency of further learning. The modified data is then gradually fed to the predictive model and allows it to continue learning without requiring a full retraining.

This holistic architecture makes sure that the drift detection and the handling of missing value is dealt with simultaneously and in context to reduce false alarms and to reduce error propagation even in difficult non-stationary environments. Figure 1 has described the modular workflow in which the framework demonstrates the ability to provide strong and adaptable operation in real-life streaming applications.

Algorithm 1 DAMM-JSD: Drift Detection and Adaptive Missing Value Imputation

---

Require: Data stream $\mathcal{D}$, window sizes w_{ref}, w_{cur}, threshold ϵ
Ensure: Updated model $\mathcal{M}$

Initialize W_{ref}, W_{cur}, model $\mathcal{M}$, parameters θ
for each data point (x_t, y_t) **do**
  Add (x_t, y_t) to W_{cur}
  if $|W_{\text{cur}}| \geq w_{\text{cur}}$ **then**
    Estimate distributions P_{cur}, P_{ref}
    Compute JSD = JSD($P_{\text{cur}} \parallel P_{\text{ref}}$)
    if JSD $> \epsilon$ **then**
      Impute missing values in W_{cur} using f_θ
      Update θ with gradient descent
      Update model $\mathcal{M}$ with imputed data
      Set $W_{\text{ref}} \leftarrow W_{\text{cur}}$, reset W_{cur}
    end if
  end if
end for

The DAMM-JSD algorithm (Algorithm 1) describes the architecture of the framework and the procedure of drift detection and automatic missing values imputation in streaming data. Buffering of incoming data is done into current and reference windows at every time step, where distributions of the observed-feature and missingness are estimated. The algorithm computes the Jensen-Shannon Divergence (JSD) between these distributions generating a single score, which provides the extent of change in the stream of data. Once this score exceeds a predetermined threshold, the system detects concept drift and enters a two-step process of adaptation: missing values are filled in with distribution-sensitive estimators, and the predictive model adapts incrementally with the adapted data. This will make the drift detection approach as well as the imputation context-dependent and adaptive to the changing data environment, which will keep the performance robust without the need to re-train it entirely.

We formalize the DAMM-JSD approach for streaming data $\mathcal{D} = \{(x_t, y_t)\}_{t=1}^{T}$, where $x_t \in \mathbb{R}^d$ and $y_t \in \mathcal{Y}$. Concept drift is defined as a change in the joint distribution:

$$P_t(x, y) \neq P_{t-1}(x, y). \tag{2.1}$$

Drift Detection: DAMM-JSD detects drift by computing the Jensen-Shannon Divergence (JSD) between distributions from two consecutive windows:

$$\text{JSD}(P_t \parallel P_{t-1}) = \frac{1}{2} D_{\text{KL}}(P_t \parallel M) + \frac{1}{2} D_{\text{KL}}(P_{t-1} \parallel M), \tag{2.2}$$

where $M = \frac{1}{2}(P_t + P_{t-1})$ and D_{KL} is the Kullback-Leibler divergence. If JSD$(P_t \parallel P_{t-1})$ exceeds a threshold, concept drift is declared.

Adaptive Imputation: When missing values are present, DAMM-JSD uses a model f_θ to estimate the missing components based on observed data:

$$\hat{x}_t^m = f_\theta(x_t^o), \tag{2.3}$$

where $\hat{x}_t^m$ denotes the imputed values and x_t^o the observed features. The model parameters θ are updated incrementally to minimize reconstruction error.

Model Update: Upon drift detection, the predictive model $\mathcal{M}$ is updated using the newly imputed and observed data:

$$\mathcal{M}_t = \mathrm{Adapt}(\mathcal{M}_{t-1}, \hat{x}_t, y_t). \tag{2.4}$$

Theoretical Properties:

- *JSD Non-negativity:* $\mathrm{JSD}(P \parallel Q) \geq 0$; zero only if $P = Q$.
- *JSD Symmetry:* $\mathrm{JSD}(P \parallel Q) = \mathrm{JSD}(Q \parallel P)$.
- *JSD Boundedness:* $0 \leq \mathrm{JSD} \leq 1$.
- *Imputation Convergence:* Given smooth loss and bounded updates, model parameters θ converge to an optimal estimator over time.

Significance: DAMM-JSD can be used to provide reliable learning with changing distributions and incomplete data because it merges drift detection and adaptive imputation. JSD yields consistent interpretable drift indications, and adaptive imputation preserves the integrity of data to robust updates in the model. Such synergy reduces false alarms and propagation of errors, which promotes correct real-time decision-making in realistic streaming settings.

3. Results

We compared the output of our proposed approach DAMM-JSD on two baselines, that is, ADWIN and PH in terms of conventional performance metrics, i.e., recall, F1-score, and stability. Further we have determined statistical significance using both Friedman tests followed by the Nemenyi post-hoc analysis. The experimental validation of DAMM-JSD is carried out on a variety of datasets i.e both real and synthetic datasets to assess its performance under varying levels of missingness and concept drift. We have experimented our method on the following three datasets:: Appliances, Gaussian, and Hyperplane.The Appliances dataset is a real-world time-series collection consisting of 19,735 instances with 28 features, capturing household energy consumption patterns. The target variable, "Appliances," is modeled as a binary classification problem (low vs. high energy usage). Due to natural fluctuations in consumption, the dataset has both missing values and non-stationary behavior, making it a more suitable for evaluating adaptive learning methods.

The Gaussian dataset is synthetic and stationary, composed of 5,000 instances and 5 features. Binary class labels are determined through a linear threshold and gradual drift is simulated by the smooth varying of the feature distributions.The Hyperplane data is a non-stationary synthetic dataset with 5,000 instances and 20 features. The datasets are labeled as binary, and a moving hyperplane determines binary labels, so that they become sudden drift events (Appliances), gradual drift events (Gaussian), and sudden drift events (Hyperplane), foreign to the full scope, on which DAMM-JSD is evaluated under varying domains and drift conditions. Table 1 summarizes the most important attributes of the datasets that we employed in our experiments.

Table 1: Experimental setup for DAMM-JSD evaluation

Dataset	Type	Inst.	Feat.	Class Labels	Drift Type	Miss. Rate
Appliances	Real-world (Time-series)	19,735	28	Binary (High/Low)	Natural + Missingness	10%–30%
Gaussian	Synthetic (Stationary)	5,000	5	Binary (0/1)	Gradual Drift	10%–30%
Hyperplane	Synthetic (Non-stationary)	5,000	20	Binary (0/1)	Sudden Drift	10%–30%

In order to systematically test the DAMM-JSD framework in the face of missing values there were three canonical mechanisms injected into the datasets of MCAR (Missing Completely At Random), MAR (Missing At Random) and MNAR(Missing Not At Random). This multi-dimensional variation enables us to stress-test the flexibility of the suggested technique with respect to missing information filling as well as drift detection simultaneously, and they often appear together in practice in streaming applications.

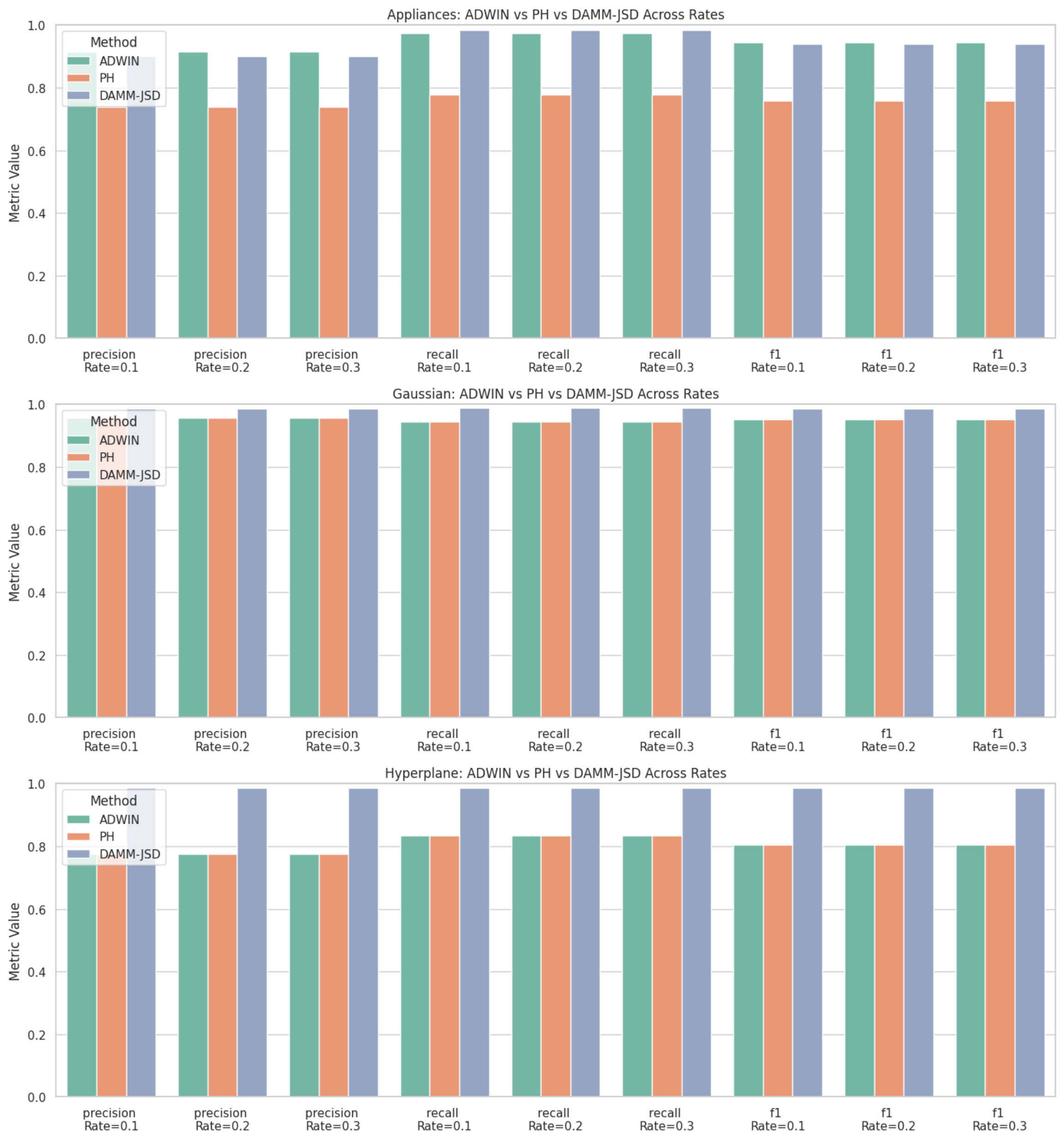


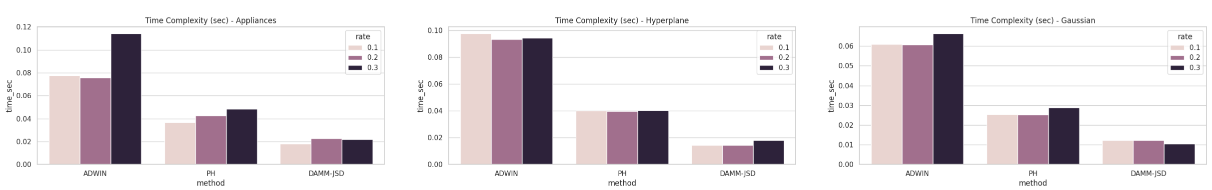
Figure 2: Bar chart comparison of precision, recall, and F1-score across missingness rates (0.1, 0.2, 0.3) for ADWIN, PH, and DAMM-JSD on Appliances, Gaussian, and Hyperplane datasets.

Table 2: Comparison of the performance of DAMM-JSD, ADWIN, and PH at different rates of missingness.

Dataset	Rate	Method	Precision	Recall	F1-score
	0.1	ADWIN	0.914	0.975	0.943
	0.1	DAMM-JSD	0.901	0.983	**0.940**
	0.1	PH	0.739	0.777	0.758
	0.2	ADWIN	0.914	0.975	0.943
Appliances	0.2	DAMM-JSD	0.901	0.983	**0.940**
	0.2	PH	0.739	0.777	0.758
	0.3	ADWIN	0.914	0.975	0.943
	0.3	DAMM-JSD	0.901	0.983	**0.940**
	0.3	PH	0.739	0.777	0.758
	0.1	ADWIN	0.957	0.945	0.951
	0.1	DAMM-JSD	0.985	0.989	**0.987**
	0.1	PH	0.957	0.945	0.951
	0.2	ADWIN	0.957	0.945	0.951
Gaussian	0.2	DAMM-JSD	0.985	0.989	**0.987**
	0.2	PH	0.957	0.945	0.951
	0.3	ADWIN	0.957	0.945	0.951
	0.3	DAMM-JSD	0.985	0.989	**0.987**
	0.3	PH	0.957	0.945	0.951
	0.1	ADWIN	0.776	0.833	0.804
	0.1	DAMM-JSD	0.986	0.985	**0.986**
	0.1	PH	0.776	0.833	0.804
	0.2	ADWIN	0.776	0.833	0.804
Hyperplane	0.2	DAMM-JSD	0.986	0.985	**0.986**
	0.2	PH	0.776	0.833	0.804
	0.3	ADWIN	0.776	0.833	0.804
	0.3	DAMM-JSD	0.986	0.985	**0.986**
	0.3	PH	0.776	0.833	0.804

4. Discussion

The results in Table 2 and Figure 2 demonstrate that DAMM-JSD consistently achieves high precision, recall, and F1-score across all datasets and missingness rates. DAMM-JSD outperforms PH by a wide margin and is competitive with or superior to ADWIN, especially in scenarios with higher missingness rates and concept drift. This confirms the robustness and effectiveness of DAMM-JSD in challenging streaming data environments.



(a) Appliances (b) Hyperplane (c) Gaussian

Figure 3: Time Complexity (sec) for the three datasets under different missingness rates.

Figures 3 compare the time complexity of DAMM-JSD, ADWIN, and PH across the Appliances, Hyperplane, and Gaussian datasets, respectively, under varying missingness rates. Compared to PH, DAMM-JSD is much better and is competitive or superior to ADWIN in cases of increased concept drift and missingness rates. This proves the efficiency and power of DAMM-JSD in tests of high-stress streaming data.

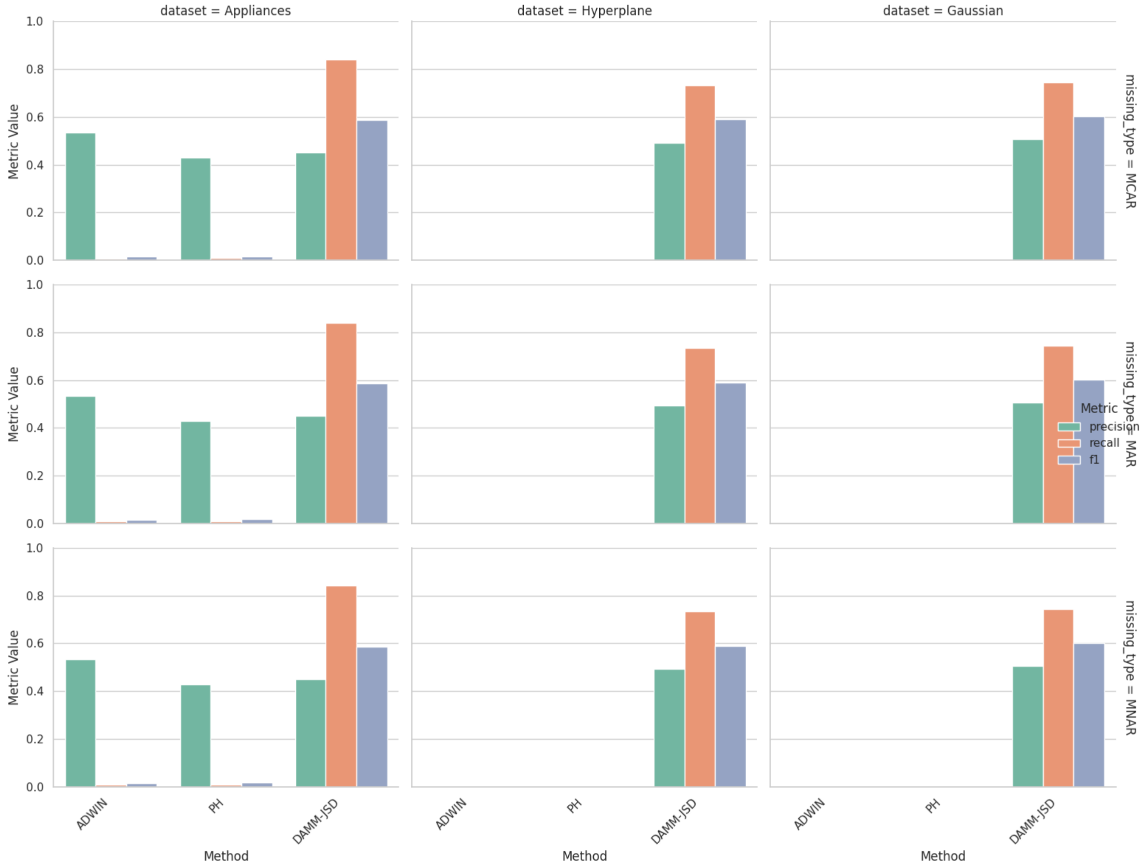


Figure 4: Precision, recall, and F1-score for ADWIN, PH, and DAMM-JSD across three datasets (Appliances, Hyperplane, Gaussian) and missingness mechanisms (MCAR, MAR, MNAR). Each panel shows the comparative performance of the methods for a specific dataset and missingness type.

Figure 4 presents a detailed comparison of ADWIN, PH, and DAMM-JSD across all datasets and canonical missingness mechanisms (MCAR, MAR, MNAR), for the key classification metrics.The multi-panel design allows measuring the individual performance of each approach to certain missing data problems directly. DAMM-JSD has always had high recall and F1 scores in all the types of missingness and datasets which shows its strength and versatility. This image is a complement to the aggregate and statistical analyses because it offers finer evidence on the superiority of DAMM-JSD.

Table 3: Friedman and Nemenyi post-hoc statistical comparison of DAMM-JSD, ADWIN, and PH.

Metric	Friedman	p-value	DAMM-JSD	ADWIN	PH
Accuracy	50.40	0.0000	1.00	2.33	2.67
Stability	28.80	0.0000	1.33	2.00	2.67
Overall	28.80	0.0000	1.33	2.00	2.67

The statistical significance of the observed differences between the evaluated methods is reported in Table 3 by the p-values. In particular, p-values of near zero point can be interpreted as the fact that the differences in performance observed between DAMM-JSD, ADWIN, and PH could hardly have been due to the effect of random chance. The average rank also supports the high-performance of DAMMJSD, as it regularly acquires the first position among all metrics, ADWIN and PH come next. The findings as in Table 3 attest to the fact that DAMM-JSD is far superior to the baseline methods in terms of accuracy, consistency, and general efficiency.

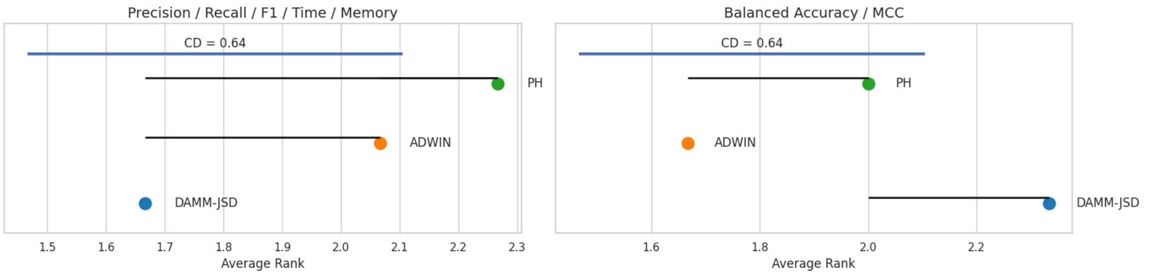


Figure 5: Critical Difference diagrams for average ranks of DAMM-JSD, ADWIN, and PH across multiple metrics.

The figure 5 show the statistically significant methods based on the Nemenyi post-hoc test, and the critical value (CD = 0.64) is represented by a horizontal bar. DAMM-JSD is regularly ranked higher and the visual distinction underscores the immense difference that the competing approaches have. This figure supports the statistical tables by providing an easy-to-understand visual report on the comparative outcomes. In general, the experimental analysis demonstrates that DAMM-JSD is more effective than the existing drift detection techniques (ADWIN and PH) under various settings of data (datasets), missingness (rates and mechanisms). The Friedman and Nemenyi post-hoc test will also validate the fact that these improvements are statistically significant and the efficiency measures will prove the computing superiority of DAMM-JSD. The visualizations also support these findings and underline the proposed method strength, stability, and feasibility of application in real-world streaming context with missing data.

5. Conclusion

In this paper,we discussed the impact of missing values on concept drift detection.We have proposed a method which is an integrated solution that not only detects the drift but also helps to handles with missing values and adapts accordingly.Our proposed drift detection algorithm has three main steps.The first one is to calculate the difference of distributional changes based on windows and that exceeds the threshold states the drift detection. The second step using an imputation method replaces missing values for efficient learning and in the final step when drift is detected the model is retrained incrementally so that it adapts to the new concept. The method has been tested on different datasets by introducing missingness and it has performed better in the evaluation metrics such as accuracy,stability and efficiency in comparison with other methods.The method has been further tested through statistical tests and the results showed the reliability of DAMM-JSD under various environments.

Our Proposed method exhibited a better results in comparison with other methods and it could be further we wish to examine the behavior in non stationary streams of data where there is no specific pattern of missing values.Further we wish to extend the method to handle multi-class and and multi label which are generally to be found in real world applications. Additionally, deployment and validation in real world application would have practical impact on the methods potentially.

References

1. Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., Zhang, G., *Learning under concept drift: a review*. IEEE Trans. Knowl. Data Eng., (2018).

2. Pratama, M., Lu, J., Lughofer, E., Zhang, G., Er, M. J., *An incremental learning of concept drifts using evolving type-2 recurrent fuzzy neural networks*. IEEE Trans. Fuzzy Syst., 25(5), 1175–1192, (2016).

3. Žliobaitė, I., Pechenizkiy, M., Gama, J., *An overview of concept drift applications*. In: Chapter 4, Springer, pp. 91–114, (2016).

4. Lu, J., Liu, A., Song, Y., Zhang, G., *Data-driven decision support under concept drift in streamed big data*. Complex Intell. Syst., 6(1), 157–163, (2020).

5. Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., Bouchachia, A., *A survey on concept drift adaptation*. ACM Comput. Surv., 46(4), 1–37, (2014).

6. Ditzler, G., Roveri, M., Alippi, C., Polikar, R., *Learning in nonstationary environments: a survey*. IEEE Comput. Intell. Mag., 10, 12–25, (2015).

7. Pratama, M., Lu, J., Zhang, G., *Evolving type-2 fuzzy classifier.* IEEE Trans. Fuzzy Syst., 24(3), 574–589, (2015).

8. Pratama, M., Anavatti, S. G., Joo, M., Lughofer, E. D., *pclass: an effective classifier for streaming examples.* IEEE Trans. Fuzzy Syst., 23(2), 369–386, (2015).

9. Das, M., Pratama, M., Savitri, S., Jie, Z., *Muse-rnn: a multilayer self-evolving recurrent neural network for data stream classification.* Proc. 19th IEEE Int. Conf. Data Min., (2019).

10. Song, Y., Lu, J., Lu, H., Zhang, G., *Fuzzy clustering-based adaptive regression for drifting data streams.* IEEE Trans. Fuzzy Syst., 28(3), 544–557, (2020).

11. Ning, Z., Yu, J., Pan, Y., Li, H., *Adaptive event-triggered fault detection for fuzzy stochastic systems with missing measurements.* IEEE Trans. Fuzzy Syst., 26(4), 2201–2212, (2017).

12. Little, R. J., Rubin, D. B., *Statistical Analysis with Missing Data.* Wiley, Hoboken, New Jersey, (2019).

13. Yoon, J., Jordon, J., Schaar, M., *GAIN: missing data imputation using generative adversarial nets.* Proc. Int. Conf. Mach. Learn., pp. 5689–5698, (2018).

14. van Buuren, S., Groothuis-Oudshoorn, K., *MICE: multivariate imputation by chained equations in R.* J. Stat. Softw., pp. 1–68, (2010).

15. Stekhoven, D. J., Bühlmann, P., *MissForest: non-parametric missing value imputation for mixed-type data.* Bioinformatics, 28(1), 112–118, (2011).

16. Yu, H.-F., Rao, N., Dhillon, I. S., *Temporal regularized matrix factorization for high-dimensional time series prediction.* Proc. Adv. Neural Inf. Process. Syst., pp. 847–855, (2016).

17. Schnabel, T., Swaminathan, A., Singh, A., Chandak, N., Joachims, T., *Recommendations as treatments: debiasing learning and evaluation.* Proc. Int. Conf. Mach. Learn., pp. 1670–1679, (2016).

18. García-Laencina, P. J., Sancho-Gómez, J.-L., Figueiras-Vidal, A. R., *Pattern classification with missing data: a review.* Neural Comput. Appl., 19(2), 263–282, (2010).

19. Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P.-A., *Extracting and composing robust features with denoising autoencoders.* Proc. Int. Conf. Mach. Learn., pp. 1096–1103, (2008).

20. Gondara, L., Wang, K., *MIDA: multiple imputation using denoising autoencoders.* Proc. Pacific-Asia Conf. Knowl. Discov. Data Min., Springer, pp. 260–272, (2018).

21. Allison, P. D., *Missing Data.* Sage Publ., Thousand Oaks, California, (2010).

22. Enders, C. K., *Applied Missing Data Analysis.* Guilford Press, (2010).

23. Liu, A., Zhang, G., Lu, J., *Concept drift detection via equal intensity k-means space partitioning.* IEEE Trans. Cybern., Early Access, pp. 1–14, (2020).

24. Silverman, B. W., *Density Estimation for Statistics and Data Analysis.* Routledge, (2018).

25. Bu, L., Alippi, C., Zhao, D., *A pdf-free change detection test based on density difference estimation.* IEEE Trans. Neural Netw. Learn. Syst., 29(2), 324–334, (2018).

26. Liu, A., Lu, J., Zhang, G., *Accumulating regional density dissimilarity for concept drift detection in data streams.* Pattern Recognit., 76, 256–272, (2018).

27. Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., Smola, A., *A kernel two-sample test.* J. Mach. Learn. Res., 13, 723–773, (2012).

28. Boracchi, G., Carrera, D., Cervellera, C., Maccio, D., *QuantTree: histograms for change detection in multivariate data streams.* Proc. Int. Conf. Mach. Learn., pp. 638–647, (2018).

29. dos Reis, D. M., Flach, P., Matwin, S., Batista, G., *Fast unsupervised online drift detection using incremental Kolmogorov–Smirnov test.* Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., pp. 1545–1554, (2016).

30. Liu, A., Lu, J., Zhang, G., *Concept drift detection: dealing with missing values via fuzzy distance estimations.* IEEE Trans. Fuzzy Syst., (2020).

31. Lima, A. M. S., Sousa, E. P. M., *Handling missing values in data streams: an overview.* Anais XXXIX Simpósio Bras. Banco Dados, pp. 750–756, (2024).

32. Bertsimas, D., Delarue, A., Pauphilet, J., *Adaptive optimization for prediction with missing data.* Mach. Learn., 114(5), 124, (2025).

33. Kamilin, M. H. B., Yamaguchi, S., *Stackade ensemble learning for resilient forecasting against missing values, adversarial attacks, and concept drift.* Appl. Sci. (Basel), 15(16), 8859, (2025).

34. Ahamed, M. A., Ye, Q., Cheng, Q., *RefiDiff: refinement-aware diffusion for efficient missing data imputation.* arXiv preprint, arXiv:2505.14451, (2025).

35. Tahir, M., Abdullah, A., Udzir, N. I., Kasmiran, K. A., *A novel approach for handling missing data to enhance network intrusion detection system.* Cyber Secur. Appl., 3, 100063, (2025).

Ushashree P.,
Department of Computer Science & Engineering,
National Institute of Technology Warangal,
India.
E-mail address: ushashree.sgs@gmail.com

and

R. B. V. Subramanyam,
Department of Computer Science & Engineering,
National Institute of Technology Warangal,
India.
E-mail address: rbvs66@nitw.ac.in