Journal of Robotics and Control (JRC)
Volume 6, Issue 5, 2025
ISSN: 2715-5072, DOI: 10.18196/jrc.v6i5.28048

SHAP-Based Framework for Temporal Detection of Sensor Drift in Gas Sensor Arrays

Ridvan Firat Cinar [1*], Gokhan Yuksek [2], Timur Lale [3], Serdar Ekinci [4], Davut Izci [5], Alfian Ma'arif [6]
[1] Department of Computer Engineering, Batman University, Batman, Türkiye
[2, 3] Department of Electrical and Electronics Engineering, Batman University, Batman, Türkiye
[4] Department of Computer Engineering, Bitlis Eren University, Bitlis, Türkiye
[5] Department of Electrical and Electronics Engineering, Uludağ University, Bursa, 16059, Türkiye
[5] Applied Science Research Center, Applied Science Private University, Amman, 11931, Jordan
[6] Department of Electrical Engineering, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
Email: [1] ridvanfirat.cinar@batman.edu.tr, [2] gokhan.yuksek@batman.edu.tr, [3] timur.lale@batman.edu.tr,
[4] sekinci@beu.edu.tr, [5] davutizci@uludag.edu.tr, [6] alfianmaarif@ee.uad.ac.id
*Corresponding Author

***Abstract*—Sensor drift is a fundamental problem that undermines the long-term reliability of machine-learning models, as gradual changes in sensor response can silently degrade predictive performance. This challenge is particularly relevant in gas sensor arrays, where environmental variability and sensor aging strongly affect data quality. Ensuring effective monitoring of such drift is essential for industrial safety, environmental monitoring, and other applications that depend on stable sensing systems. To address this issue, we develop a SHAP (SHapley Additive Explanations)–based framework combined with an XGBoost classifier to detect drift at the feature level. The model is trained on structured sensor data, and SHAP values are computed batch-wise to capture temporal changes in feature relevance. Preprocessing enforces balanced class distributions and consistent concentration ranges, reducing potential confounding effects. Results from the Gas Sensor Array Drift dataset demonstrate that several sensor channels exhibit clear temporal variability in SHAP importance. Notably, these changes were observed even when accuracy and F1-score remained stable, highlighting that SHAP provides a complementary early-warning signal of drift. Case studies further confirm that the framework can capture both gradual and abrupt drift patterns. The research contribution of this study lies in providing an interpretable and time-resolved perspective on drift detection, enabling the localization of drift signals to specific sensors and features, independent of drift type or feature domain. This approach advances the state of the art by bridging explainable AI and drift monitoring. In conclusion, the framework supports practical deployment of adaptive sensing systems by making drift behaviour visible before conventional metrics degrade. Future work will focus on real-time scalability, broader datasets, and automated drift correction through recalibration or retraining.**

Keywords*—SHapley Additive Explanations (SHAP); Sensor Drift; Concept Drift; Model Monitoring; Time-Series Interpretability; Feature Importance Analysis.

I. INTRODUCTION

Machine learning is increasingly being used in sensor systems, particularly for the classification of gas types based on multi-sensor data [1]-[3]. A key problem in such systems is the so-called sensor drift phenomenon, in which the response of a sensor changes over time due to factors such as aging, humidity, or environmental conditions [4]-[8]. This drift can lead to a significant alteration in model performance, often without being captured by traditional performance metrics in a timely manner. Therefore, it is important to use explainable and time-aware methods to evaluate and maintain the reliability of machine learning models in real-world sensor applications.

SHAP (SHapley additive explanations) is an established method of explainable AI that allows the precise quantification of the influence of individual features on model decisions [9]-[12]. By analyzing SHAP values, one can not only determine how reliably a model functions, but also why it makes the decisions it does. This is particularly crucial for time-dependent data such as sensor-based measurements. However, the application of such methods requires careful data preparation and stable classification models [11].

In addition to model interpretability methods, a broad range of drift detection techniques has been developed. Statistical process control approaches employ control charts to identify abnormal fluctuations in data streams. Dedicated drift detection methods such as the drift detection method (DDM), early drift detection method (EDDM), and adaptive windowing algorithms have been widely applied to streaming data to capture concept drift in a timely manner. Distributional divergence metrics are also commonly used to quantify shifts in feature distributions across time [13]-[15]. These approaches provide formal statistical tools for identifying temporal change. In contrast, our framework leverages SHAP values not as a replacement but as a complementary tool, offering feature-level interpretability that links detected distributional changes to their impact on model decision mechanisms [16], [17].

A key contribution of this work is the use of SHAP values as temporal indicators to detect early-stage drift in sensor-based machine learning models. By computing the mean and variance of SHAP values across predefined time intervals (batches), it becomes possible to identify features whose influence on the model output varies significantly over time which is an important signal of feature-level drift. This approach highlights not only which features are unstable but also when they begin to diverge in their predictive relevance. To ensure meaningful and reliable results, the analysis is

constrained to representative concentration ranges and balanced class distributions, thus minimizing confounding effects that could obscure true drift patterns [18], [19].

In this work, the term drift refers to temporal changes in sensor response characteristics that alter the statistical relationship between features and model outputs. Such changes can originate from covariate shifts in feature distributions, concept drift affecting decision boundaries, or physical sensor degradation over time. Defining drift in this way ensures that both data-driven and device-related variations are taken into account in our analysis.

As part of the preprocessing, missing or inconsistent data entries are excluded from the dataset. Gas types are transformed into numerical class labels, and only concentration levels with enough samples are kept for analysis. In addition, features are grouped based on their corresponding sensors and attribute types to support meaningful interpretation of the SHAP outputs. These steps help to improve the clarity and reliability of the analysis [20]-[23].

In this work, a SHAP-centric analysis method is presented that allows a model-based, batch-wise assessment of sensor drift. Using XGBoost (eXtreme gradient boosting) [24], [25] classifiers and subsequent SHAP evaluation, drift-prone sensor features are identified over time and visually processed. This allows model reliability to be continuously monitored. The proposed method offers a complete solution for the clear evaluation of ML (machine learning) models in drift-prone environments [26]-[29].

Sensor drift is a physical phenomenon caused by sensor aging, environmental conditions, or device instability, leading to gradual changes in raw sensor responses. In contrast, concept drift is a machine learning notion, defined as changes in the statistical relationship between input features and target outputs. In practice, sensor drift often induces concept drift from the perspective of predictive models, since altered sensor signals shift the feature–label mapping. Accordingly, our framework treats sensor drift as the physical root cause while analysing its manifestations in the form of concept drift through feature-level SHAP dynamics [30]-[33].

The research contribution is threefold. First, we propose a SHAP-based framework to quantify and monitor temporal drift at the feature level, enabling interpretable identification of unstable sensor channels. Second, we integrate batch-wise SHAP analysis with visualization techniques that reveal both the onset and progression of drift. Third, we demonstrate the practical value of this framework through case studies on the Gas Sensor Array Drift dataset, highlighting its ability to provide complementary early-warning signals beyond conventional performance metrics. By positioning SHAP within the broader landscape of drift detection, the study advances both methodological clarity and practical interpretability for dynamic sensing systems.

II. DATASET

A. Dataset Overview

Gas sensor array drift dataset referred in [34], publicly available from the UCI Machine Learning Repository, is a benchmark dataset specifically designed to evaluate the performance of classification algorithms against long-term sensor drift. It is collected over a period of 36 months (January 2007 – February 2011) at the ChemoSignals Laboratory, BioCircuits Institute, University of California, San Diego. The dataset comprises 13,910 measurements obtained from an array of 16 metal-oxide (MOX) gas sensors exposed to six different pure gases, acetaldehyde, acetone, ammonia, ethanol, ethylene, and toluene at various concentration levels ranging from 5 to 1000 ppm. Each measurement includes 128 features derived from time-series response signals, computed as 8 signal attributes per sensor (e.g., maximum, mean, variance, area under the curve, etc.). The dataset is organized into 10 batches, each corresponding to a distinct time period, thus emulating real-world temporal drift in sensor response characteristics. The primary target variable is the gas type, represented as a categorical label, while concentration and batch identifiers are also provided.

This dataset is frequently used in the context of domain adaptation, sensor calibration, and explainable machine learning, particularly for studying how sensor behavior changes over time and how models can adapt to or compensate for such drift. The dataset is made freely available for academic and research purposes and has become a standard benchmark in the field of chemo-sensory machine learning research. In this regard, studies such as [35]-[38] may be consulted for further insights into the diverse applications of this dataset.

B. Data Preprocessing

The dataset is pre-processed by encoding categorical gas labels into numerical class values and removing any missing entries to ensure consistency. Each of the 128 features, representing 8 statistical attributes across 16 gas sensors, is retained without reduction to preserve interpretability. Features are grouped by sensor and attribute type to support downstream explainability analysis. Additionally, data is optionally filtered by gas concentration to enable controlled drift studies. Batch identifiers are preserved to track temporal variation in sensor responses.

C. Model Training

A classification model is trained to identify different gas types based on sensor responses. XGBoost algorithm is used because it works well with structured data and can handle noisy or drifting inputs. The model is trained on 128 features from 16 sensors, each representing different response characteristics. Gas types are converted into numeric class labels for training. To control concentration effects, the data is filtered to fixed levels where needed. This setup enables us later to apply SHAP analysis and understand how each sensor feature contributed to the model's decision. XGBoost has been commonly applied in sensor-based studies due to its ability to handle structured data and maintain stable performance under varying input conditions [24].

III. SHAP Based Drift Analysis

Fig. 1 illustrates the overall workflow of the proposed SHAP-based drift monitoring framework. The process begins with dataset acquisition, preprocessing, and model training, followed by batch-wise SHAP computation and statistical evaluation. If drift is detected, the framework not only flags and localizes the affected features or sensors but also includes an optional feedback loop. This loop represents recalibration of unstable sensors or retraining of the model with updated data, ensuring long-term reliability in real-world deployment scenarios.

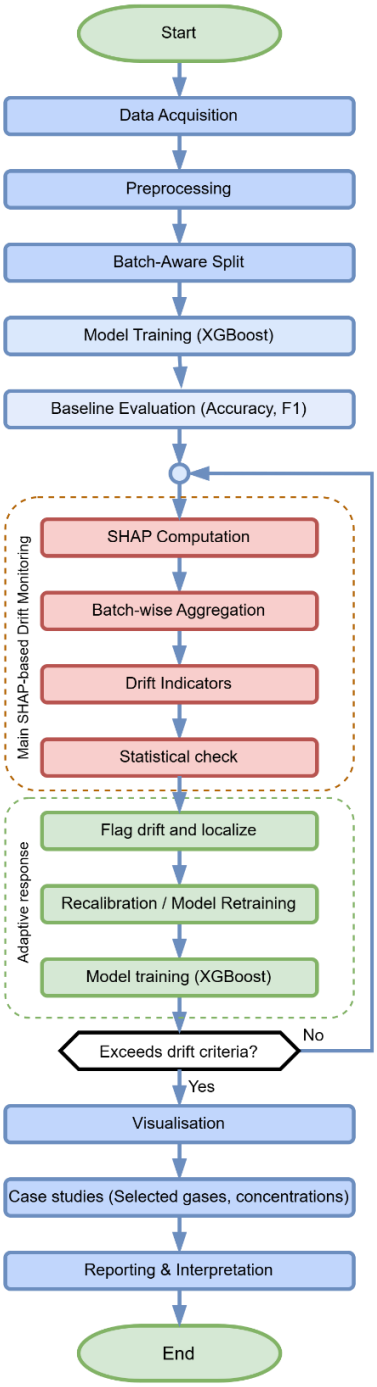


Fig. 1. Workflow of the proposed SHAP-based drift monitoring framework with optional feedback loop, illustrating dataset acquisition, preprocessing, model training, SHAP computation, statistical evaluation, drift localization, and visualization

The visualization and case study steps provide interpretable outputs that support both technical analysis and practical decision-making. As shown in Fig. 1, the framework outlines the full pipeline from dataset acquisition to visualization, including an optional feedback loop for model recalibration. It should be emphasized that Fig. 1 provides a high-level overview of the workflow, while implementation details such as parameter tuning and statistical thresholds are described in the methodology section. Since model retraining modifies the decision boundaries, SHAP computation is repeated after each iteration to ensure that drift monitoring remains accurate and interpretable.

A. SHAP: Classic Shapley Estimation

The theoretical foundation of SHAP values derives from cooperative game theory, specifically the Shapley value formulation for distributing a model's output among its input features. In this view, each feature represents a "player" in a game contributing to the final prediction. The Shapley value for a feature i quantifies its average marginal contribution over all possible feature coalitions $S \subseteq F \setminus \{i\}$. The exact formulation is:

$$\phi_i = \sum_{S \subseteq F\{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} \times \left[f_{S\cup\{i\}}\left(x_{S\cup\{i\}}\right) - f_{S(x_S)} \right] \quad (1)$$

where, f_S denotes the model trained using only features in subset S, and x_S is the corresponding input. This approach, referred to as Shapley regression values, requires retraining the model on all subsets of features, making it computationally expensive. To mitigate this, approximations such as Shapley sampling values and quantitative input influence methods are employed, which estimate SHAP values using sampling rather than full model retraining. Importantly, these methods still maintain the additive feature attribution structure, aligning with the broader SHAP framework [9], [10].

B. SHAP In Drift Detection

In dynamic environments where sensor characteristics evolve over time, detecting distributional shifts which is also known as concept or sensor drift, is crucial for maintaining the model reliability. Traditional performance metrics may fail to provide early warnings of such drift, especially in the absence of labelled data from the target distribution. SHAP offers a feature-level attribution method that remains model-agnostic and locally faithful to predictions. By tracking how the importance of specific input features (e.g., sensor readings) fluctuates across time-indexed batches, SHAP can serve as a powerful tool to uncover subtle shifts in data distributions that may not yet affect overall accuracy [39]-[41].

SHAP computes the contribution of each feature to an individual prediction based on game-theoretic Shapley values. For a model f, the SHAP value ϕ_i for feature i is computed by averaging the marginal contribution of feature i across all possible feature subsets:

In the context of drift detection, we compute SHAP values over time $\phi_i^{(t)}$ for a specific feature i across time batches $t \in \{1,2,\ldots,T\}$. A growing variance or systematic shift in the mean of $\phi_i^{(t)}$ over time may signal drift in the influence of

feature i on model predictions. Thus, SHAP does not just flag changes in raw input distribution, but in how those changes impact the model's internal decision boundary that is offering a more insightful, model-aware drift signal.

To monitor sensor drift in a model-sensitive manner, SHAP values are used to track how the effect of each sensor feature changed over time. Since SHAP assigns a numerical contribution to each feature per prediction, we can analyse these contributions across batches to detect drift at the feature level. The expectation from a stable sensor is to show consistent SHAP values across time, while increasing variance or directional change in SHAP values suggests that the model's trust in that feature is shifting - a signal of potential drift. By grouping samples based on batch ID and computing the mean and variance of SHAP values per feature, we identify attributes and sensors most affected by temporal degradation. This method enables early detection of drift before it significantly harms classification accuracy and provides interpretability by revealing which sensors are becoming less reliable over time.

IV. Structural Analysis and Visualization of Drift Patterns

A. Structural Insights

To gain a better understanding of the dataset's structure and temporal dynamics, several exploratory visualizations are created. These figures aim to clarify class balance, concentration variability, and batch-level distribution shifts. Such visual inspections are essential before applying SHAP-based interpretability methods, as they reveal key structural properties of the data that could influence model behaviour and explainability. In particular, they help determine whether class imbalance or drift-like patterns may bias learning outcomes or affect feature importance assessments.

Fig. 2 shows the distribution of gas classes across the dataset. While all six gases are represented, the sample counts are uneven. Acetone and Ethylene are the most frequent, whereas Ammonia and Toluene have significantly fewer samples. This imbalance may affect the learning process by causing the model to dominant classes, potentially leading to weighted SHAP values and reduced interpretability for underrepresented gases.



Fig. 2. Distribution of gas classes in the Gas Sensor Array Drift dataset, showing imbalance across six gases with acetone and ethylene dominating while ammonia and toluene are underrepresented

Fig. 3 presents the range of gas concentrations using boxplots grouped by gas type. Ammonia displays a wide and variable concentration spread, while gases such as Toluene are mostly clustered at lower ppm levels. This variation suggests that sensor responses may differ not only by gas type but also by exposure range. For consistency in modelling and SHAP analysis, it is advisable to focus on concentration intervals that are sufficiently populated and evenly distributed across classes.

Fig. 4 visualizes the global structure of the feature space using principal component analysis (PCA), with samples colored by batch. Distinct clusters and directional shifts can be observed, indicating that the underlying data distribution changes over time. These temporal shifts provide early evidence of sensor drift, supporting the decision to incorporate batch-aware explanations into the modelling pipeline. Such separation further motivates the analysis of SHAP value trends across batches to detect early-stage drift behavior.

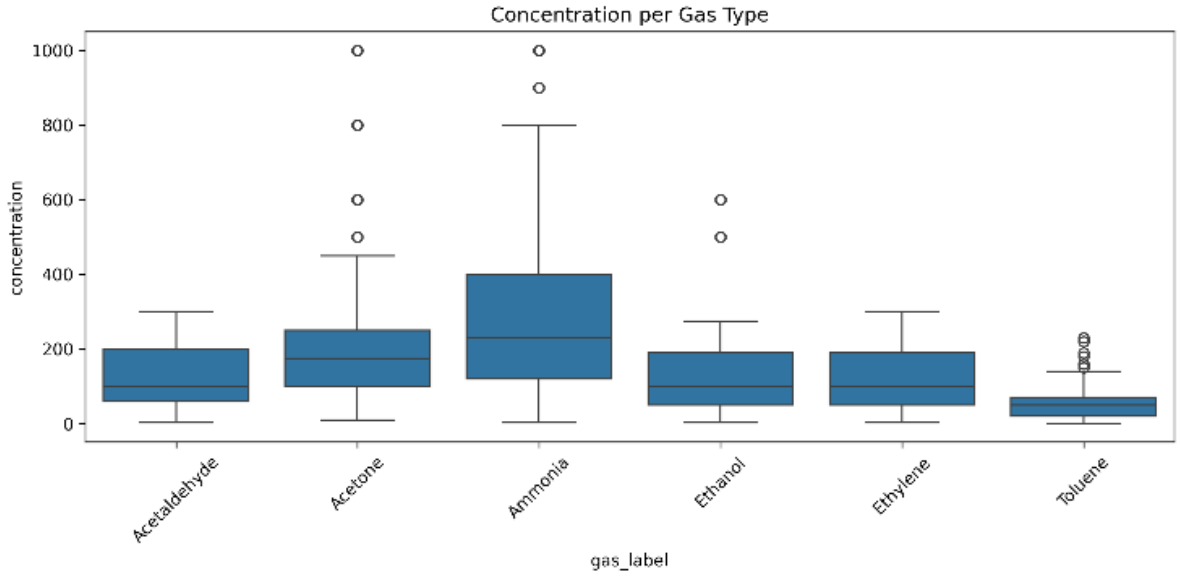


Fig. 3. Boxplot of concentration ranges for each gas type, indicating variability across gases (e.g., wide spread for ammonia, clustered low ppm values for toluene)

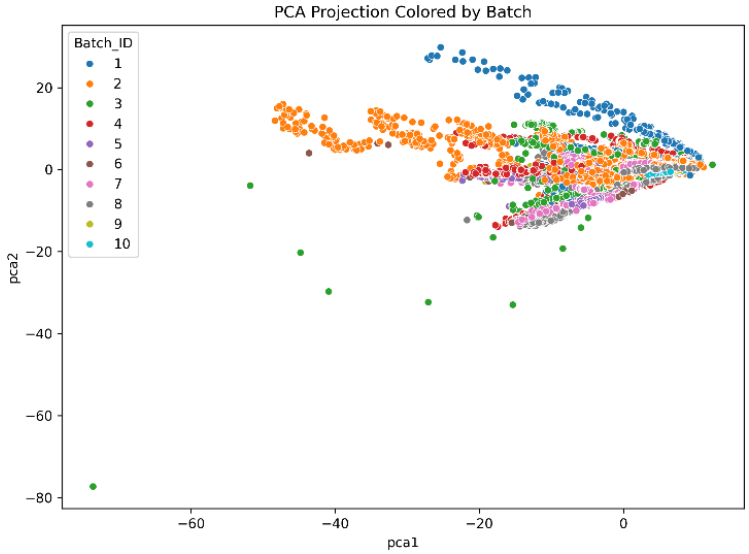


Fig. 4. Principal component analysis (PCA) projection of the dataset colored by batch, highlighting temporal shifts and clustering that reflect drift in the data distribution

B. Preselection and Focusing on Drifting Points

As shown in Fig. 5, the heatmap presents the average SHAP drift variance across sensors and gas types, aggregated over all available concentration levels. Higher values indicate greater temporal instability in a sensor's contribution to model predictions, suggesting sensitivity to sensor drift. Notably, Sensor 6 and Sensor 2 exhibit high average drift for gases such as Acetaldehyde, Ethanol, and Toluene, with maximum values exceeding 7.0. In contrast, several sensors (e.g., Sensor 7, Sensor 14, Sensor 16) remain consistently stable across gases. Toluene appears relatively sparse in the dataset across many concentration levels, resulting in insufficient data to compute reliable SHAP statistics for some sensor combinations, which are left blank in the matrix. This global overview helps prioritize model monitoring or sensor recalibration efforts.

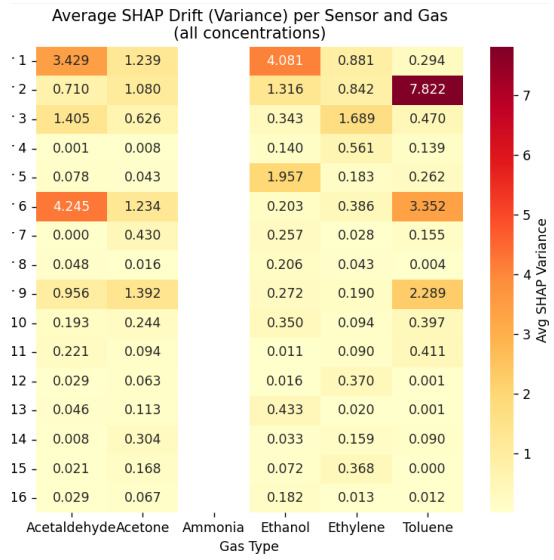


Fig. 5. Heatmap of average SHAP drift variance across sensors and gas types, identifying drift-prone channels (e.g., Sensors 2 and 6) and stable ones (e.g., Sensors 7, 14, 16)

While the previous section highlighted potential temporal inconsistencies in the data, these patterns remained largely qualitative in nature. In this section, SHAP analysis is applied to evaluate the evolution of the influence of individual features quantitatively. The key idea is that sensor drift may not only manifest in raw feature values but also in how these features contribute to the model's output.

To reveal such changes, SHAP values are calculated across batches for a fixed concentration level and class label. By examining the mean and variance of these values over time, features with fluctuating predictive relevance are identified, indicating the presence of drift. In this way, not only shifts in the input data but also changes in the model's internal behavior can be detected, offering a more detailed understanding of sensor dynamics and the stability of the system over extended periods.

Due to the high dimensionality of the dataset, which includes six gas types, multiple concentration levels, and 128 features across 10 time-based batches, it is impractical to analyze every possible combination. Instead, this study focuses on a small number of selected cases that are representative of the overall structure. A fixed concentration level and a single gas class are used in each example to isolate drift effects. Additionally, feature selection is guided by sensor structure and prior visual inspection, allowing for a more targeted and interpretable drift analysis. This approach strikes a balance between clarity and technical relevance while maintaining applicability to other parts of the dataset.

V. CASE STUDY

A. Case I: SHAP-based Drift Analysis for Acetaldehyde (Class 1) at 100 ppm Concentration

This study aims to detect sensor drift with interpretable artificial intelligence methods (especially SHAP) in the context of gas classification problems. Within the scope of the study, the processed version of the gas sensor array drift dataset obtained from UCI is used, and the analyses are limited to observations with only 100 ppm concentration. Acetaldehyde (Class 1) is targeted as a class, and the variability of samples belonging to this class on different production batches (Batch ID 1–10) is evaluated. In

particular, the behavior of the features that are effective in the classification decisions of the model over time is monitored with SHAP values, and in this way, the drift effect is interpreted qualitatively.

After training the XGBoost classifier as a model, SHAP values are calculated for each sample and the contribution vector belonging to the target class (Class 1) is extracted. These values are then grouped on a batch basis and converted to average SHAP values, and the five features with the highest variance are selected. Line graphs (drift traces) are created for these five features, which include the average SHAP values over time (Fig. 6). In addition, a heat map is also created that visualizes the average SHAP values of 16 features, each corresponding to the maximum response channel of different sensors. Thus, a detailed drift analysis is performed both at the time-dependent feature level and at the sensor level.

The feature with the highest variance, feature_36, exhibited a remarkable stability-deterioration transition in the drift trace graph. The SHAP effect, which is horizontal in the range of Batch 1–6, entered a significant increasing trend as of Batch 7 and peaked in Batch 10. This suggests that the model began to find this feature increasingly important over time and that the sensor produced increasingly dominant signals in different batches. A similar clear increase in positivity is observed in the sensor channel corresponding to this feature in the heat map; especially the transition to red tones in the range of Batch 8–10 produced a consistent result.

Feature_44 presents a drift characteristic that changes direction over time. In the SHAP graph, it reaches the highest positive value around Batch 3, then gradually decreases and becomes negative in Batch 9. This change implies that the model initially evaluated this feature as a positive discriminator, but over time it became less confident and even had an adverse effect in distinguishing the class. The transitions of the sensor to which this feature corresponds in the heat map from light blue to dark blue between Batch 4–9 are exactly parallel to the SHAP decrease observed in the graph.

Despite having a medium level of variance, feature_101 exhibits a very sharp negative direction change, especially after Batch 5. This sudden deviation suggests that the sensor may have undergone a physical change or that there is a serious systematic shift in the data. The SHAP value of the sensor corresponding to this feature in the heatmap supports this break by suddenly turning to blue tones after Batch 6. The model evaluated this feature as unreliable over time.

Feature_109 is a feature that shows bidirectional drift, switching between both positive and negative SHAP values. The positive SHAP value, which peaks in Batch 3, turns negative between Batches 6–8. This transformation reveals that the model's approach to this feature has evolved in the opposite direction over time. When the heatmap is examined, it is seen that the colour transitions in the sensor corresponding to this feature change from red to dark blue, and the graph completely overlaps with this behaviour.

Feature_16, which has lower variance compared to others, draws a steady decline curve spreading over time. The SHAP effect, which is positive starting from Batch 1, decreased steadily and almost decreased to zero in Batch 10. This suggests that there may be gradual wear and tears on the sensor performance or a long-term loss of stability. The sensor corresponding to this feature in the heat map also exhibited a gradual transition from a warm colour to a neutral tone in a similar manner.

In general, the patterns observed in both the drift trace graphs and the SHAP heat map are reliable with each other. The change in the effect of the features selected based on variance on model decisions over time is clearly seen; the sensor-based counterpart of these effects can also be monitored in parallel via the heat map in Fig. 7. The color transitions in the heat map progressed synchronously with the graphical changes; this situation demonstrated both the reliability of the SHAP-based drift analysis and the effect of sensor behaviors on the model decision mechanism. The findings reveal that SHAP-based analysis of variance is an effective method to evaluate model reliability, especially in long-term distributions. Key findings from Case I indicate that specific features such as feature_36 and feature_44 exhibited strong temporal drift signals, with consistent confirmation at the sensor level. The alignment of SHAP line traces with heatmap transitions supports the validity of the approach. These results highlight the sensitivity of SHAP-based monitoring to both gradual and abrupt drift effects in acetaldehyde at 100 ppm.

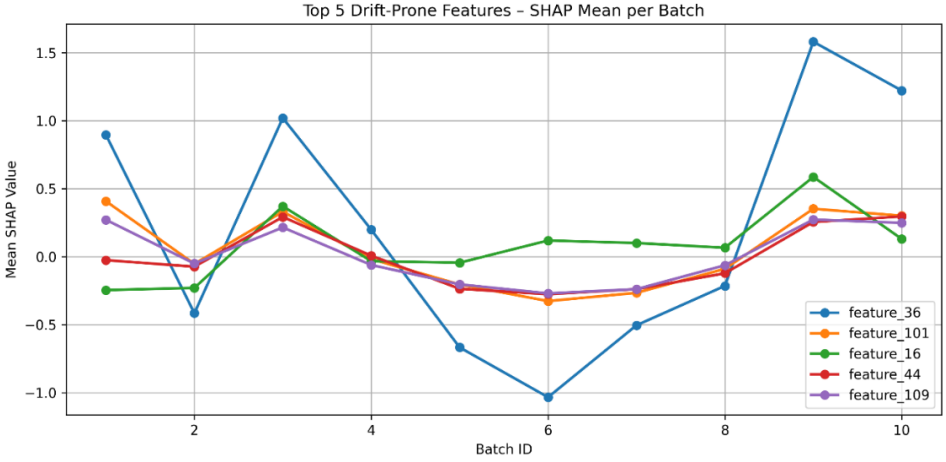


Fig. 6. Line plots of the top 5 drift-prone features (Class 1, 100 ppm), illustrating temporal changes in SHAP importance across batches

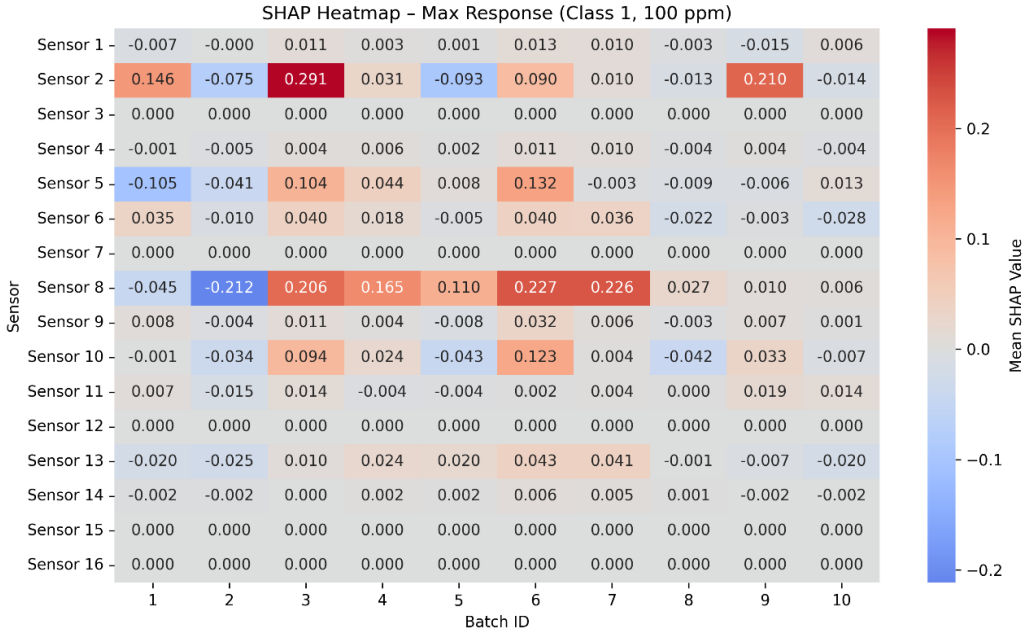


Fig. 7. Heatmap of SHAP values for maximum response channels (Class 1, 100 ppm), showing sensor-level drift patterns consistent with feature-level SHAP traces

B. Case II: SHAP-based Drift Analysis for Toluene (Class 4) at 200 ppm Concentration

In this second case study, the reflection of the temporal changes in sensor performance in high concentration samples of toluene gas (Class 4, 200 ppm) on the model interpretation is analyzed through SHAP values. The analysis only includes observations of this class and concentration, and the data is divided into 8 different batches (Batch ID 1–8). The aim is to observe which features the model evaluates under these conditions, to what extent and in what ways, and to reveal potential drift effects. In this context, 5 features with high variance are selected and the time-dependent trends of SHAP values are examined, and a SHAP heat map is created over the features of the maximum response channels corresponding to the sensors.

As displayed in Fig. 8 the top 5 drifted features by order, feature_26, feature_18, feature_1, feature_31 and feature_33, are visualized with a line graph showing the changes in SHAP effects over time. Feature_26 stands out as a particularly striking example: this feature, which has a very high impact of +1.1 in Batch 1, gradually shifts to negative until Batch 5 and drops to -1.1. This indicates that the relevant sensor channel changes from a positive to a negative discriminator in terms of the model over time. Other features show more limited but regular deviations; feature_18, for example, is positive in Batch 1 and declines to -0.3 in Batches 5–6, while feature_33, on the contrary, exhibits an increasing trend that increases its importance over time.

SHAP heat map only includes features corresponding to sensor channels with maximum response, and thus, drift behavior at the sensor level can be analyzed (Fig. 9). Sensor 1 exhibits significant fluctuations across batches: it has a negative SHAP effect of -0.371 in Batch 1, increased to +0.123 in Batch 2, then turns negative again and became positive again in Batch 8. This sensor is directly related to feature_1 and exhibits a behaviour parallel to the directional changes observed in the line graphs. Similarly, Sensor 5 stands out with a sudden SHAP increase (+0.275) in Batch 6–7, supporting the dominance of feature_31 and feature_33 over time. The SHAP effect of most of the other sensors remain close to zero, indicating that the drift is concentrated only on certain sensors/channels.

A high level of consistency is observed between the SHAP lines of the top 5 features and the heatmap. Especially features that show SHAP direction changes with time, such as feature_26 and feature_1, are supported by the direction changes of the related sensors (e.g. Sensor 1) on the heatmap. Sharp transitions on the SHAP heatmap (e.g. Batch 1→2 for Sensor 1 or Batch 6 for Sensor 5) and SHAP deviations on the line graphs occurred simultaneously, indicating the extent to which the model relies on different sensor channels over time. Key findings from Case II show that high-concentration toluene samples produced notable directional shifts in SHAP values, particularly for feature_26 and feature_1. Sensor-level analysis confirmed these fluctuations, indicating localized drift in specific channels. The study demonstrates how SHAP can capture complex and bidirectional drift behavior in high concentration scenarios.

C. Comparison with Conventional Drift Detection Methods

To contextualize the proposed SHAP-based framework, Table I contrasts it with widely used drift detection approaches such as ADWIN, DDM, and distributional divergence metrics. The comparison highlights strengths and limitations of each method and clarifies the added value of SHAP for feature-level interpretability.

As shown in Table I, conventional detectors provide reliable statistical guarantees of global shifts but remain blind to the underlying causes. In contrast, the SHAP-based framework offers model-aware, feature-level insights, enabling early warnings and targeted diagnosis. Thus, the proposed method should be seen as complementary rather than competitive with existing approaches.

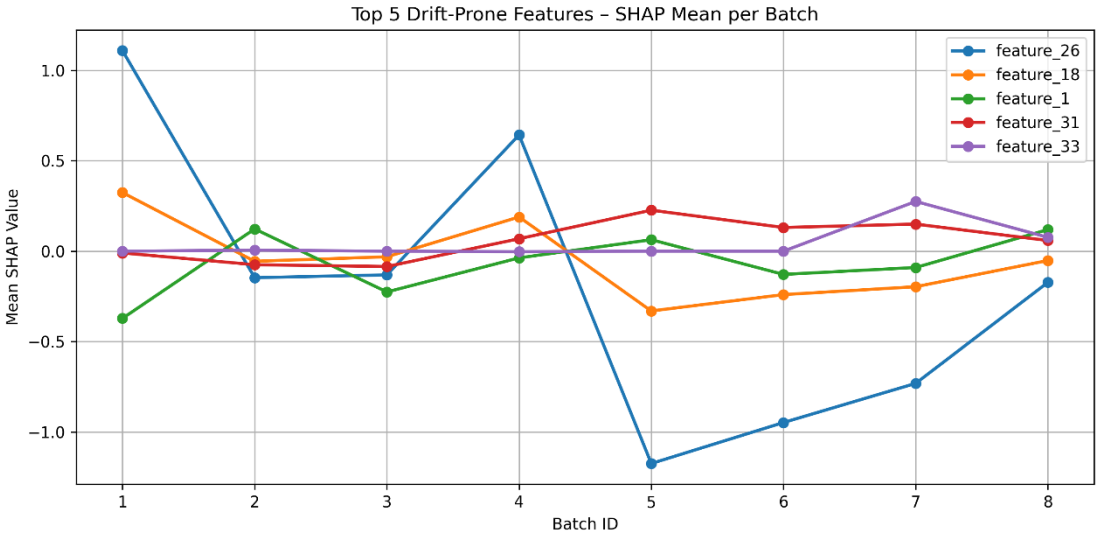


Fig. 8. Line plots of the top 5 drift-prone features (Class 4, 200 ppm), highlighting both gradual and abrupt changes in SHAP contributions over time

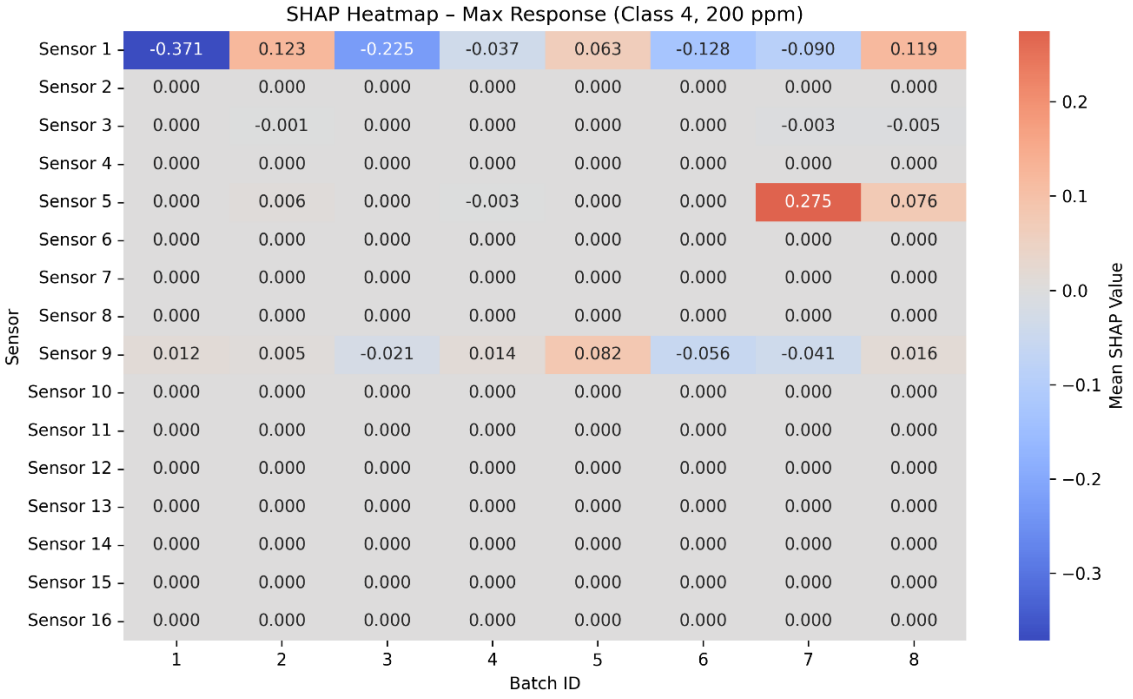


Fig. 9. Heatmap of SHAP values for maximum response channels (Class 4, 200 ppm), confirming sensor-level drift signatures in line with feature-level trends

TABLE I. COMPARISON OF SHAP-BASED DRIFT MONITORING WITH CONVENTIONAL DRIFT DETECTION METHODS

Method type	Strengths	Limitations	Added value of SHAP-based framework
ADWIN / DDM (window-based detectors)	Online, adaptive, fast; statistical guarantees on distributional change	Global signal only, no feature attribution; may trigger late if accuracy degrades slowly	SHAP identifies which features drift and why the model shifts, providing interpretability
KL / Wasserstein divergence (distributional metrics)	Quantifies magnitude of distributional shift; simple to compute	Blind to model impact; sensitive to noise; requires feature distribution estimation	SHAP highlights shifts that matter for the model decision, filtering out irrelevant noise
Performance monitoring (accuracy, F1)	Directly reflects end-task degradation; easy to track	Insensitive to early drift; no localization of cause	SHAP provides early-warning signals before accuracy drops, and points to drifting sensors
SHAP-based (this study)	Model-aware, feature-level, interpretable; localizes drift; early-warning even if metrics stable	Model-dependent; computational cost; currently lacks formal statistical thresholds	Complements conventional methods by bridging drift detection with explainable AI

VI. CONCLUSION AND DISCUSSION

In this study, a SHAP-based interpretability framework was developed to detect temporal drift in gas sensor array models. The analysis demonstrated that SHAP values provide complementary early-warning signals of feature-level instability, even when conventional performance metrics remain stable. Case studies on different gases and concentrations showed that the method can capture both gradual and abrupt drift patterns, highlighting its potential for monitoring model reliability in dynamic environments.

Beyond these results, a key contribution of the proposed framework is its ability to make anomalous drift behaviour visible at the feature level via SHAP analysis. Unlike traditional approaches that primarily rely on changes in global accuracy or distributional statistics, our method tracks shifts in feature contributions that underpin model decisions. This enables more sensitive detection of drift phenomena, whether stemming from covariate shifts, concept drift, or sensor degradation, all before conventional metrics noticeably change. By decoupling detection from specific drift types or feature domains, the framework offers an interpretable and practical route to monitor long-term sensor reliability.

With these promising results, several limitations should be acknowledged. The analysis was restricted to selected gas classes and concentration levels, which may limit generalizability across the entire dataset. The evaluation relied primarily on variance and mean shifts in SHAP values without extensive statistical testing, and the model dependency of SHAP may introduce artifacts that do not directly correspond to sensor degradation. In addition, data balancing and filtering, while useful for clarity, may have masked certain types of real drift.

In conclusion, the framework supports practical deployment of adaptive sensing systems by making drift behaviour visible before conventional metrics degrade. Future work will focus on real-time scalability, broader datasets, and automated drift correction through recalibration or retraining. Future work will also integrate statistical significance testing, such as trend analysis and change-point detection, to provide stronger validation of SHAP-based drift indicators. In addition, comparative benchmarks with established drift detectors (e.g., ADWIN, DDM, KL/Wasserstein divergence) will be conducted to quantify performance differences and further contextualize the added value of the proposed framework.

REFERENCES

[1] B. Zong, S. Wu, Y. Yang, Q. Li, T. Tao, and S. Mao, "Smart Gas Sensors: Recent Developments and Future Prospective," *Nanomicro Lett*, vol. 17, no. 1, p. 54, Dec. 2025, doi: 10.1007/s40820-024-01543-w.

[2] M. Liu, R. Ren, X. Zhou, S. Zhu, and T. Wang, "From gas sensing to AI–gas sensing," *Chemical Communications*, vol. 61, no. 55, pp. 9996–10010, Jul. 2025, doi: 10.1039/D5CC01291K.

[3] A. Krayden, M. Avraham, H. Ashkar, T. Blank, S. Stolyarova, and Y. Nemirovsky, "TinyML-Based Real-Time Drift Compensation for Gas Sensors Using Spectral–Temporal Neural Networks," *Chemosensors*, vol. 13, no. 7, p. 223, Jun. 2025, doi: 10.3390/chemosensors13070223.

[4] E. Llobet, "Gas sensors using carbon nanomaterials: A review," *Sens Actuators B Chem*, vol. 179, pp. 32–45, Mar. 2013, doi: 10.1016/j.snb.2012.11.014.

[5] A. Vergara, S. Vembu, T. Ayhan, M. A. Ryan, M. L. Homer, and R. Huerta, "Chemical gas sensor drift compensation using classifier ensembles," *Sens Actuators B Chem*, vol. 166–167, pp. 320–329, May 2012, doi: 10.1016/j.snb.2012.01.074.

[6] K. Jiang *et al.*, "Gas Sensor Drift Compensation Using Semi-Supervised Ensemble Classifiers with Multi-Level Features and Center Loss," *ACS Sens*, vol. 10, no. 4, pp. 2906–2918, Apr. 2025, doi: 10.1021/acssensors.4c03655.

[7] A. Dries and U. Rückert, "Adaptive concept drift detection," *Proc West Mark Ed Assoc Conf*, vol. 1, pp. 232–243, 2009, doi: 10.1137/1.9781611972795.21.

[8] M. Rataj, X. Zhang, G. Hovakimyan, and J. M. Bravo, "Evolving Strategies in Machine Learning: A Systematic Review of Concept Drift Detection," *Information*, vol. 15, no. 12, p. 786, Dec. 2024, doi: 10.3390/INFO15120786.

[9] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," *Adv Neural Inf Process Syst*, vol. 30, pp. 4766–4775, Nov. 2017.

[10] L. Merrick and A. Taly, "The Explanation Game: Explaining Machine Learning Models Using Shapley Values," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12279, pp. 17–38, 2020, doi: 10.1007/978-3-030-57321-8_2.

[11] C. Molnar, *Model-Agnostic Methods*. Lulu.com, 2020.

[12] Y. Lee, Y. Lee, E. Lee, and T. Lee, "Explainable Artificial Intelligence-Based Model Drift Detection Applicable to Unsupervised Environments," *Computers, Materials & Continua*, vol. 76, no. 2, pp. 1701–1719, Aug. 2023, doi: 10.32604/cmc.2023.040235.

[13] P. Naumann, J. Kauffmann, and G. Montavon, "Wasserstein Distances Made Explainable: Insights into Dataset Shifts and Transport Phenomena," *arXiv preprint arXiv:2505.06123*, 2025.

[14] C. Coil, K. Faber, B. Sniezynski, and R. Corizzo, "Distance-based change point detection for novelty detection in concept-agnostic continual anomaly detection," *J Intell Inf Syst*, pp. 1–39, Jun. 2025, doi: 10.1007/s10844-025-00949-1.

[15] C. Shen, H. Liu, J. Wang, Z. Yang, and C. Hai, "Kullback–Leibler Divergence-Based Distributionally Robust Chance-Constrained Programming for PV Hosting Capacity Assessment in Distribution Networks," *Sustainability*, vol. 17, no. 5, p. 2022, Feb. 2025, doi: 10.3390/su17052022.

[16] B. Zimmermann and M. Boussard, "Improving Drift Detection by Monitoring Shapley Loss Values," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13364, pp. 455–466, 2022, doi: 10.1007/978-3-031-09282-4_38.

[17] L. Baier, N. Kühl, G. Satzger, M. Hofmann, and M. Mohr, "Handling Concept Drifts in Regression Problems - the Error Intersection Approach," *Wirtschaftsinformatik*, pp. 210–224, Mar. 2020, doi: 10.30844/WI_2020_C1-BAIER.

[18] N. U. Edakunni, U. Tekriwal, and A. Jain, "Explaining Drift using Shapley Values," *arXiv preprint arXiv:2401.09756*, 2024.

[19] R. Klinkenberg, "Learning drifting concepts: Example selection vs. example weighting," *Intelligent Data Analysis*, vol. 8, no. 3, pp. 281-300, 2004, doi: 10.5555/1293831.1293836.

[20] F. Bayram, B. S. Ahmed, and A. Kassler, "From concept drift to model degradation: An overview on performance-aware drift detectors," *Knowl Based Syst*, vol. 245, Jun. 2022, doi: 10.1016/j.knosys.2022.108632.

[21] S. Ramírez-Gallego, B. Krawczyk, S. García, M. Woźniak, and F. Herrera, "A survey on data preprocessing for data stream mining: Current status and future directions," *Neurocomputing*, vol. 239, pp. 39–57, May 2017, doi: 10.1016/J.NEUCOM.2017.01.078.

[22] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*, in Intelligent Systems Reference Library, vol. 72. Cham: Springer International Publishing, 2015, doi: 10.1007/978-3-319-10247-4.

[23] A. Farhangfar, L. Kurgan, and J. Dy, "Impact of imputation of missing values on classification error for discrete data," *Pattern Recognit*, vol. 41, no. 12, pp. 3692–3705, Dec. 2008, doi: 10.1016/J.PATCOG.2008.05.019.

[24] T. Chen and C. Guestrin, "XGBoost," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016, doi: 10.1145/2939672.2939785.

[25] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artif Intell Rev*, vol. 54, no. 3, pp. 1937–1967, 2021.

[26] A. Tsymbal, "The problem of concept drift: definitions and related work," *Computer Science Department, Trinity College Dublin*, vol. 106, no. 2, p. 58, 2004.
[27] Q. Liu, X. Hu, M. Ye, X. Cheng, and F. Li, "Gas Recognition under Sensor Drift by Using Deep Learning," *International Journal of Intelligent Systems*, vol. 30, no. 8, pp. 907–922, Aug. 2015, doi: 10.1002/int.21731.
[28] H. Zheng and A. Paiva, "Assessing Machine Learning Approaches to Address IoT Sensor Drift," *arXiv preprint arXiv:2109.04356*, 2021.
[29] B. Schölkopf, "Causality for Machine Learning," *Probabilistic and Causal Inference*, pp. 765–804, Dec. 2019, doi: 10.1145/3501714.3501755.
[30] 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, Mar. 2025, doi: 10.3390/app15063056.
[31] F. Hinder, V. Vaquet, and B. Hammer, "Feature-based analyses of concept drift," *Neurocomputing*, vol. 600, p. 127968, Oct. 2024, doi: 10.1016/j.neucom.2024.127968.
[32] L. Yang, J. Cheng, Y. Luo, T. Zhou, and X. Zhang, "Detecting and rationalizing concept drift: A feature-level approach for understanding cause–effect relationships in dynamic environments," *Expert Syst Appl*, vol. 260, p. 125365, Jan. 2025, doi: 10.1016/j.eswa.2024.125365.
[33] A. Tziouvaras *et al.*, "Towards Reliable AI in 6G: Detecting Concept Drift in Wireless Network," *arXiv preprint arXiv:2508.00042*, 2025.
[34] J. Fonollosa, I. Rodríguez-Luján, and R. Huerta, "Chemical gas sensor array dataset," *Data Brief*, vol. 3, pp. 85–89, Jun. 2015, doi: 10.1016/j.dib.2015.01.003.
[35] S. Seifi, S. A. Schober, C. Carbonelli, L. Servadei, and R. Wille, "Detection of Sensor-to-Sensor Variations Using Explainable AI," *2023 Smart Systems Integration Conference and Exhibition (SSI)*, pp. 1-6, 2023, doi: 10.1109/SSI58917.2023.10387955.
[36] N. Dennler, S. Rastogi, J. Fonollosa, A. van Schaik, and M. Schmuker, "Drift in a popular metal oxide sensor dataset reveals limitations for gas classification benchmarks," *Sens Actuators B Chem*, vol. 361, p. 131668, Jun. 2022, doi: 10.1016/J.SNB.2022.131668.
[37] Z. Jiang, P. Xu, Y. Du, F. Yuan, and K. Song, "Balanced distribution adaptation for metal oxide semiconductor gas sensor array drift compensation," *Sensors*, vol. 21, no. 10, 2021, doi: 10.3390/S21103403,.
[38] X. Dong, S. Han, A. Wang, and K. Shang, "Online Inertial Machine Learning for Sensor Array Long-Term Drift Compensation," *Chemosensors 2021, Vol. 9, Page 353*, vol. 9, no. 12, p. 353, Dec. 2021, doi: 10.3390/CHEMOSENSORS9120353.
[39] H. Zhang, X. Chen, M. Hu, and V. Sugumaran, "Multilayer Concept Drift Detection Method Based on Model Explainability," in *IEEE Access*, vol. 12, pp. 190791-190808, 2024, doi: 10.1109/ACCESS.2024.3517697.
[40] L. Xu, Z. Han, D. Zhao, X. Li, F. Yu, and C. Chen, "Addressing Concept Drift in IoT Anomaly Detection: Drift Detection, Interpretation, and Adaptation," *IEEE Transactions on Sustainable Computing*, vol. 9, no. 6, pp. 913–924, Nov. 2024, doi: 10.1109/TSUSC.2024.3386667.
[41] C. Duckworth *et al.*, "Using explainable machine learning to characterise data drift and detect emergent health risks for emergency department admissions during COVID-19," *Sci Rep*, vol. 11, no. 1, p. 23017, Nov. 2021, doi: 10.1038/s41598-021-02481-y.