

Continuous detection of concept drift in industrial cyber-physical systems using closed loop incremental machine learning

Dinithi Jayaratne¹ · Daswin De Silva¹ · Damminda Alahakoon¹ · Xinghuo Yu²

Received: 24 May 2021 / Accepted: 9 August 2021

Published online: 22 September 2021

© The Author(s) 2021 [OPEN](#)

Abstract

The embedded, computational and cloud elements of industrial cyber physical systems (CPS) generate large volumes of data at high velocity to support the operations and functions of corresponding time-critical and mission-critical physical entities. Given the non-deterministic nature of these entities, the generated data streams are susceptible to dynamic and abrupt changes. Such changes, which are formally defined as concept drifts, leads to a decline in the accuracy and robustness of predicted CPS behaviors. Most existing work in concept drift detection are classifier dependent and require labeled data. However, CPS data streams are unlabeled, unstructured and change over time. In this paper, we propose an unsupervised machine learning algorithm for continuous concept drift detection in industrial CPS. This algorithm demonstrates three types of unsupervised learning, online, incremental and decremental. Furthermore, it distinguishes between abrupt and reoccurring drifts. We conducted experiments on SEA, a widely cited synthetic dataset of concept drift detection, and two industrial applications of CPS, task tracking in factory settings and smart energy consumption. The results of these experiments successfully validate the key features of the proposed algorithm and its utility of detecting change in non-deterministic CPS environments.

Keywords Artificial Intelligence · Concept drift · Industrial data stream · Incremental learning · Unsupervised machine learning

1 Introduction

Recent advances in cyber-physical systems (CPS) have necessitated machine learning algorithms in embedded applications to operate in nonstationary, time variant environments [1]. In CPS, learning in nonstationary environments, commonly known as concept drift learning, focuses on event driven changes in the environment. The underlying models generated by learning algorithms are influenced by changes in feature information (x) and target variables (y) due to such evolving concepts [2]. Concept drift occurs when this feature information (x) and target variables (y) change over time. Concept drift can be formalized as a change in the joint probability $P(x, y)$, which is defined as:

$$P(x, y) = P(y/x) \times P(x)$$

In a smart factory setting, a large number of Industrial Internet of Things (IIoT) devices and sensors will be collecting data on machine status and factory operations [3]. These data are transmitted to CPS which will then use a variety

✉ Daswin De Silva, d.desilva@latrobe.edu.au | ¹Centre for Data Analytics and Cognition, La Trobe University, Melbourne, Australia. ²School of Engineering, RMIT University, Melbourne, Australia.



of methods to predict when a machine is malfunctioning, or a process is suboptimal. Such anomalous behaviors are detected as concept drifts [3]. Detection of concept drifts in CPS decreases the negative impact of a compounding error and enables cost-effective predictive maintenance. However, data streams in industrial CPS are composed of unlabeled target variables that do not fit into predefined classes [4, 5]. Ensemble learning algorithms that integrate multiple supervised algorithms find it infeasible and impractical to detect concept drifts in this environment. To address these challenges as well as to manage complex data patterns and distributional assumption violations embedded in the industrial applications of CPS data streams, a novel unsupervised machine learning technique is needed.

Current research literature defines two distinct types of concept drift, real and virtual [6]. In real concept drift, the conditional distribution $P(y/x)$ of the target variable $p(y)$ given the input features $P(x)$ changes while the distribution of the input remains unchanged. In virtual concept drift, input data $p(x)$ changes without affecting the conditional distribution $P(y/x)$. In both types the joint distribution $P(x, y)$ changes [6]. A large body of existing work assumes the immediate availability of labels and thereby focuses on supervised machine learning algorithms for concept drift detection and adaption [6]. This assumption is not valid for CPS data streams that generate virtual concept drifts where the target label is only available following an unknown/undefined delay. A closed loop framework that has been proposed for real concept drift detection [6], operates even in cases where the target variables are delayed. However, the framework does not support concept drift detection from unlabeled data in evolving data streams.

The key challenges of continuous concept drift detection from CPS data streams are: (1) learn from large volumes of unlabelled data arriving in a short time span as data storage is impractical and infeasible (2) incorporate detected concept drift information into new data (3) unlearn (or forget) data which corresponds to concepts that are irrelevant and (4) integrate with the proposed closed loop framework for updating predictive models based on drift detection. The proposed unsupervised machine learning algorithm overcomes the aforementioned challenges with the following research contributions:

1. A novel unsupervised learning algorithm for continuous detection and adaption to concept drifts that is also able to distinguish between reoccurring and abrupt concept drifts.
2. Extending an existing closed loop framework for concept drift detection to include unlabeled data from evolving data streams.
3. Demonstration of the proposed algorithm and extended framework on SEA dataset and CPS data streams; physical activity monitoring, and energy consumption.

The rest of the paper is organized as follows. Section 2 reports related work in CPS and concept drift, followed by Sect. 3 which delineates the algorithm development; an extension to the generic framework for concept drift followed by an explication of the adaptive learning paradigms (incremental, decremental and online) used in the proposed algorithm. Section 4 presents the proposed unsupervised, adaptive learning algorithm and demonstrates its features using the SEA dataset. Section 5 presents experiments conducted on two industrial datasets demonstrating distinctive features of the algorithm and Sect. 6 concludes the paper.

2 Related work

A CPS has been defined as a system that integrates its hardware function with a cyber-representation acting as a virtual representation for the physical part. It interlinks embedded systems, which are real-time and deterministic with cloud platforms, which are probabilistic and less-constrained [7]. Within this definition itself, the importance of unsupervised learning from unlabeled data is established as a key driver of the development of CPS through the integration and fusion of both cloud platforms and embedded systems. The introduction and integration of intelligent technologies has been discussed and advocated to address the challenges of flexibility, robustness, adaptation, and reconfigurability in CPS [7–9]. Furthermore, the key technological and operational characteristics required for the active use of cyber-physical systems in future smart factories is reported in [10]. Among these, the criticality of concept drift detection and the use of unsupervised machine learning have also been highlighted.

In concept drift literature, two distinct types of concept drift are defined, real and virtual [6]. A majority of this literature assumes the immediate availability of labels and thereby focuses on supervised machine learning algorithms for concept drift detection and adaption. This assumption is not valid for real-world data streams that generate virtual concept drifts where the target label is only available following an unknown/undefined delay. Adaptive machine learning algorithms

have been proposed for such unlabeled data streams, and these can be categorized into active drift detection techniques, ensemble techniques and hybrid techniques. Active drift detection learns from a partially labeled set of sample data [11], such as 'Just-in-time' classifier and 'Intersection of Confidence' which use Cumulative Sum based active drift detection [12, 13]. Sliding window mechanisms such as 'Concept Adapting Very Fast Decision Tree' [14] and 'Incremental Online Information Network' [15] algorithms have also been proposed for active drift detection. The 'Early Drift Detection Method' [16] identifies gradual drifts by monitoring the distance between errors of a classifier and comparing the mean to a threshold. On the other hand, ensemble techniques such as multi-classifiers attempt passive drift detection using techniques such as 'Streaming Ensemble Algorithm (SEA)' [17] and 'Adaptive Hoeffding Tree Bagging' [18] where the oldest concept is replaced with the newest concept. The 'Dynamic Integration' [19] and 'Dynamic Weighted Majority (DWM)' [20] replace the least contributing member. Combining sliding windows from active detection and classifier ensembles, hybrid approaches such as 'Random Forests with Entropy' [21] and 'ADWIN' have been proposed. 'Massive Online Analysis' (MOA) implements ADWIN as a hybrid approach [22]. A semi-supervised learning method for virtual concept drift detection proposed by is based on adaptable clustering, which analyzes the distribution of clusters and updates cluster centroids according to concept drifts in data streams. More recently, unsupervised learning methods have also been proposed for concept drift detection, such as the Plover algorithm that uses varied measure functions [23], online sequential extreme learning machines [24], and a discriminative classifier with a sliding window [25]. In industrial settings, concept drift detection approaches have been proposed for predictive maintenance [26], sensor networks [27], and smart city applications [28].

In terms of machine learning capabilities in CPS, clustering data streams from high throughput machining cycle conditions [29], real-time reliability evaluation of CPS system [30], an IoT-based wearable system for fetal movement monitoring [31], detecting time synchronization attacks on CPS [32], and behaviour-based attack detection and classification [33] are some of the leading instances of direct value generation from machine learning. In contrast, the number of studies focusing on concept drift detection in CPS is limited. The primary work is in the detection and adaption to imbalanced industrial data streams using an ensemble of offline classifiers [3]. This paper highlights the limitations of condition-based maintenance in addressing or even detecting concept drift and they propose an ensemble approach to offline classification to address the three-stages of condition-based maintenance with concept drifts and imbalance data. It is also pertinent to note that a primary recommendation for future work in concept drift is the detection and validation of change detection and adaptation in the absence, delay and on-demand labeling of CPS data streams. Drawing on this context of technological and operational characteristics required of industrial CPS, as well as the limited application of machine learning in the development of such features leads up towards the contribution of this paper, where we propose a novel machine learning algorithm for continuous detection and adaption to concept drifts from CPS data streams and the integration of this capability into an established closed loop framework for concept drift detection.

3 Algorithm development

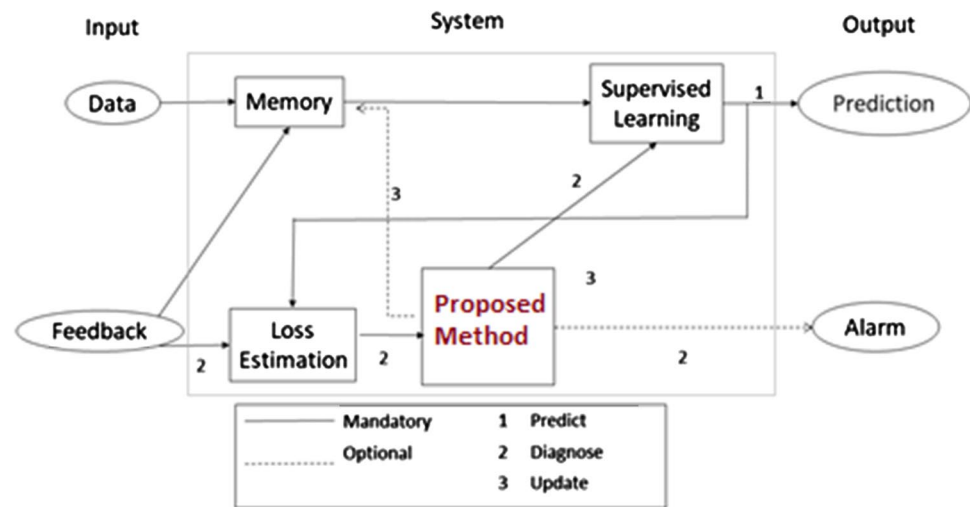
This section begins by extending the aforementioned closed loop framework [6] to include the proposed unsupervised machine learning algorithm, followed by a subsection on the novel learning features of the proposed algorithm, incremental, decremental and online learning.

3.1 Extension to the closed loop framework

The closed loop framework proposed for real concept drift detection updates a predictive model based on drift detection [6]. The framework is composed of four modules; memory, machine learning, loss estimation, and change detection. The data stream is initially received by the memory module and then presented to the machine learning module. The loss estimation module tracks the performance of the machine learning algorithm and sends information to the change detection module to update the model and machine learning algorithm.

In the proposed extension (Fig. 1), the memory module defines what data is presented and how the data flow is managed. The unsupervised learning module (the proposed algorithm) determines how online, incremental and decremental learning are used for detection and adaption to concept drift. Concept drift detection module defines the measure that can be used for detection of various types of concept changes that occur in the data stream and generate alerts for decision-making. The supervised learning module will be notified as the concept drift are detected and trigger the loss estimation module to verify accuracy in the predictive module using late feedback.

Fig. 1 Extended closed loop framework for concept drift detection from unlabeled data streams, based on [5]



3.2 Adaptive learning properties

The proposed algorithm is based on three adaptive learning features that are required for concept drift detection from unlabeled data streams. They are, (1) incremental, (2) decremental and (3) online learning.

Incremental learning: is necessary for learning from data streams as it effectively addresses both time and memory constraints [34–36]. Since incremental learning algorithms learn from continuous incoming data streams, they do not need an initially labeled dataset for training. They assume that the concepts learned before are similar to the concept of new incoming data [37].

Decremental learning: is used to unlearn (to forget) representations of the data stream which are no longer relevant. Learning from data streams should be continuous while preserving the previously known useful knowledge. Natural cognitive systems gradually forget previously learned information [36]. Decremental learning is used for forgetting old concepts and adapt to new concepts since concepts learn at one time is not relevant at another and dilutes the new concept with the old concept.

Online learning: Data streams generate data at high speed and in large volumes. Online learning is introduced to address this limitation of high frequency and high-velocity data streams that influence the iterative nature of a machine learning algorithm.

The incremental learning features of the proposed algorithm are based on the Incremental Knowledge Acquisition and Self Learning (IKASL) algorithm [38]. The IKASL algorithm is an unsupervised, incremental learning algorithm that continues to learn new data based on generalized layers of past learning outcomes. It has been successfully demonstrated on social media text mining [39] and smart electricity meter data for pattern classification and demand forecasting [40–42]. Incremental learning in IKASL is initiated by aggregation of unsupervised machine learning outcomes with the formation of generalization layers. Each generalized node expands into its own feature map to generate a topological representation of subsequent input vectors. The proposed algorithm addresses the main limitation of existing concept drift detection in CPS through the above-mentioned unsupervised adaptive learning features. These features allow the proposed algorithm to detect concept drifts with increased accuracy and efficiency compared to the algorithms currently found in literature.

4 The proposed algorithm

Based on the IKASL learning approach, this algorithm advances into decremental learning and online learning for continuous detection and adaption to concept drift from an unlabeled data stream. A variation of this technique was applied to explore the importance of context awareness to estimate road traffic [43], investigate the impact of driver behavior change on the coordination between self-driven and human-driven vehicles [44], and as the core machine learning function of an expansive intelligent traffic data integration and analysis platform [45]. The proposed algorithm consists

of three primary functions, (1) online learning, (2) incremental and decremental learning and (3) concept drift detection (Fig. 2), each function is discussed below.

Online learning: Online k-means clustering is used for one pass online learning for efficient one pass processing of a stream of data rather than storing and processing in batches [46]. In the first iteration k and t_a are user-defined for online k-means and the generated cluster feature vectors (CFV_{OC}) are input to the offline IKASL function. In subsequent iterations, k is the number of cluster feature vectors ($\#GFV_{IKASL}$), t_x (e.g. t_b-t_a) is the time taken by IKASL for the learning process, and cluster feature vectors for online k-means are the generalized nodes received from the IKASL function. These automated k and t_x implements the nonparametric nature of the algorithm.

Incremental and decremental learning: IKASL learning occurs as per the original algorithm for incremental learning. Inputs are batches of CFVoc received periodically from the online learning function (Fig. 2). We extended the IKASL function to facilitate decremental learning by forgetting the generalized node that is not the winner of any of the inputs in the data set of the subsequent learning phase. In this case, the generalized node is forgotten indicating the concept has changed or evolved. Associations between nodes in the generalization layers will be persistent, leading to the creation of a memory-like structure based on the aggregated outcomes of the learning stages. Adaptation to a new concept is formalized with the incremental and decremental learning.

Concept drift detection: Concept drift detection is carried out by calculating the distance between generalized nodes (CFV_{IKASL}) of consecutive iterations. The algorithm is sufficiently generic for any distance measure to be used, such as Euclidean distance, heterogeneous Euclidean overlap distance, Mahalanobis distance, Hellinger distance [47]. As a concept drift occurs, there would be a significant distance change, followed by a reduced distance change in the following iteration. Concept drifts detected are further identified by the algorithm as abrupt concept drift and reoccurring concept drift.

4.1 Demonstration

The SEA dataset [48], a synthetic dataset widely used in supervised concept drift detection, was used to demonstrate features of the proposed algorithm. SEA concept generator models real, abrupt concept drifts which have three independent real-valued attributes in $[0, 10]$. The data set consists of 60,000 examples in four concepts, 15,000 examples for each having different threshold values for the concept function.

Figure 3 illustrates concept drifts detected from the SEA dataset. The x-axis denotes timestamps of incremental learning, and distance measure (in this case Euclidean distance, ED_n) calculations from step 5 of the algorithm are denoted on the y-axis. Abrupt concept drifts were detected at timestamps 2, 9, 16 and 26 with ED_n 0.43, 0.39, 0.31 and 0.39 respectively. Results were validated with concept drifts detected in the same dataset by [48, 49].

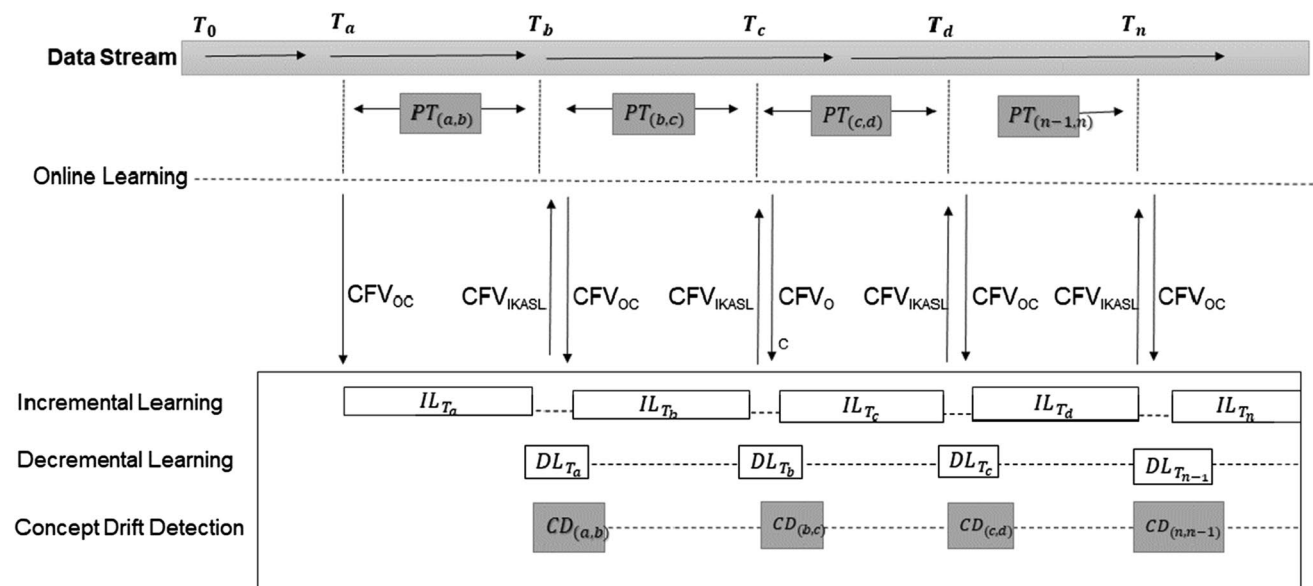
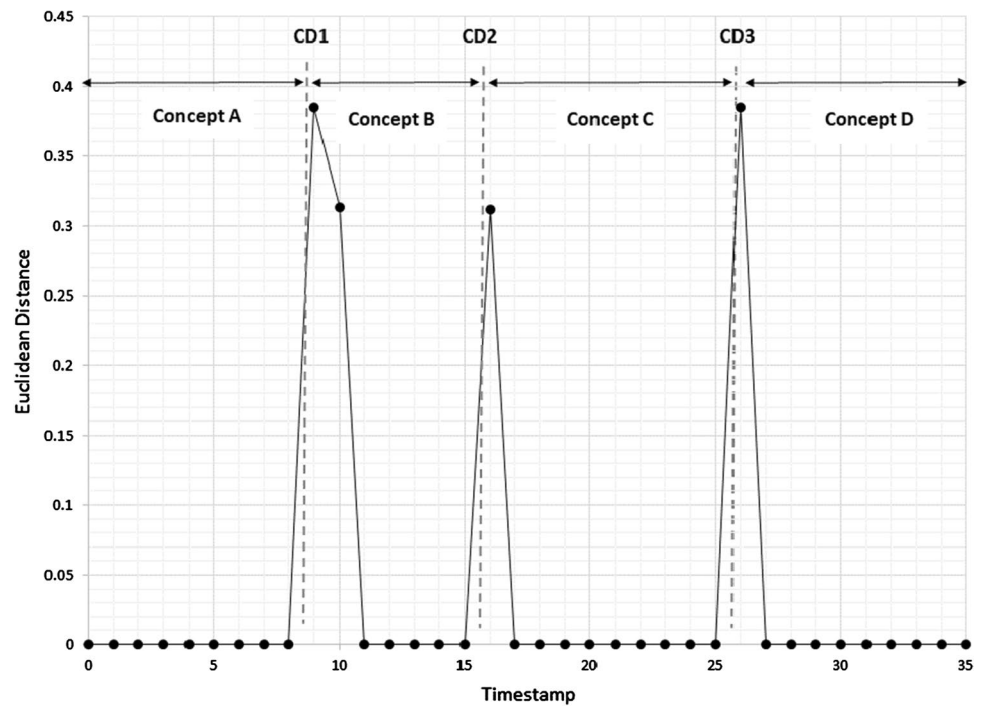


Fig. 2 Proposed unsupervised, adaptive learning algorithm for concept drift detection

Fig. 3 SEA dataset—concept drift detection



To demonstrate the importance of real-time concept drift detection, accuracies of a supervised predictive algorithm with and without concept drift detection were compared (Fig. 4). For the latter case, the algorithm was trained with first 1000 records, and the trained model was used to test the data in each subsequent batch of 1000 records. The accuracy of the algorithm reduces as the concepts evolve over time (Fig. 4). For the former case, the algorithm was trained and retrained at each concept drift detection with the most recent 1000 records. The accuracy of the algorithm improves as the algorithm was re-trained with the evolved concepts (Fig. 4).

4.2 Demonstration on modified SEA dataset

An advantage of the SEA dataset generator is that it can be configured to generate data with the repetition of the same four concepts to evaluate the identification of reoccurring concepts. For this demonstration, we generated a SEA dataset with four concepts repeated three times. The proposed unsupervised algorithm was analyzed against the corresponding concept drifts shown by MOA (Fig. 5). A total of twelve (four concepts repeated three times) concept drifts were identified by the proposed algorithm (Fig. 6) and directly corresponded to the MOA output. Concept drifts were identified at execution timestamps; [t2], [t4], [t7], [t9], [t12], [t14], [t16], [t18], [t20], [t23], [t5], [t28]. As

Fig. 4 Accuracy with and without concept drift detection

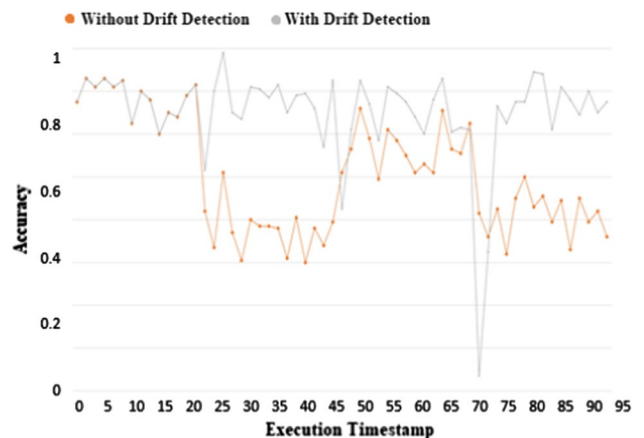
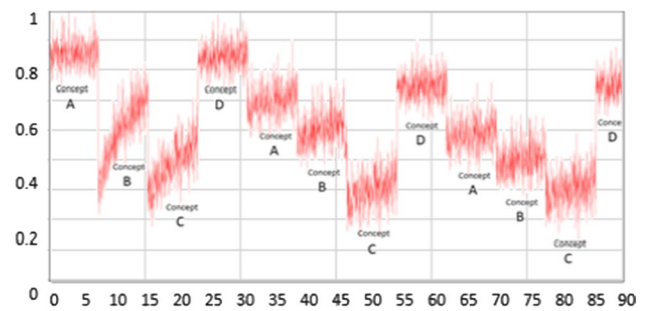


Fig. 5 MOA [26] output for modified SEA data set



shown in Table 1, time taken to detect a reoccurring concept drift reduces overtime demonstrating the incremental nature of the learning.

5 Experiments

This section presents experiments conducted on two industrial applications of CPS data streams, activity monitoring and energy consumption. Both experiments are based on real-world settings, where cyber-physical systems have to address the technical challenges of volume of data, frequency of data generation as well as the variety of data, in terms of recurring patterns, outliers and noise.

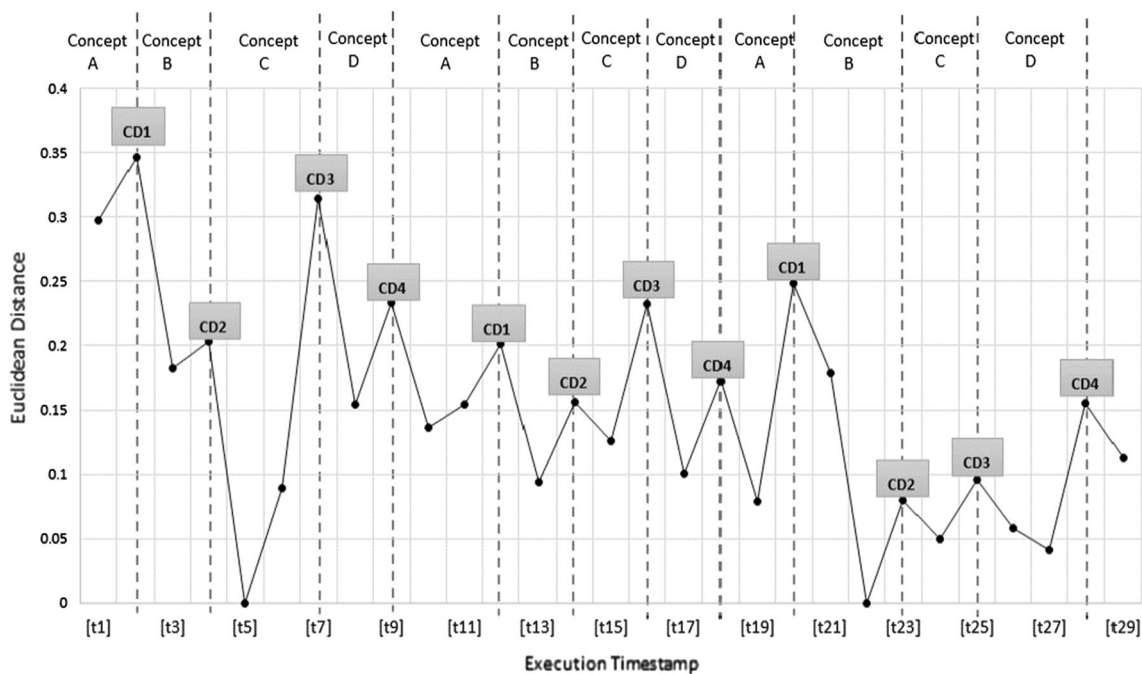


Fig. 6 Concept drift detection for modified SEA data set

Table 1 Automated time window analysis

Concept	SEA dataset 1		SEA dataset 2		SEA dataset 3	
	Concept drift	Execution time (ms)	Concept drift	Execution time (ms)	Concept drift	Execution time (ms)
A	[t2]	295	[t12]	285	[t20]	279
B	[t4]	271	[t14]	267	[t23]	258
C	[t7]	247	[t16]	239	[t25]	234
D	[t9]	244	[t18]	234	[t28]	230

5.1 Wearable sensors in industry CPS

Activity monitoring aims at providing accurate information on human activities by leveraging wearable devices available in today’s sensory rich industrial data environment. Numerous applications in industrial settings propose use of activity monitoring. Activity recognition is proposed in proactive instruction systems where instructions for the next activity are displayed at end of a tracked activity [50]. Further, task tracking by activity monitoring is used in training car assemble line workers [51]. Another major use case is quality control which verifies task performance and completion. In industrial health and safety monitoring systems, activity is monitored for unusual movements such as vibration or acceleration to generate alerts [52].

The PAMAP2 dataset [53] comprises sensor data from three inertial measurement units and a heart-rate monitor. The data are recorded while nine subjects’ complete different physical activities such as lying, standing, walking, running, cycling and rope jumping. This multivariate, time-series dataset includes 52 attributes and more than 3.8 million data records. With the use of this labeled dataset, we aim to evaluate the detection of concept drifts. Activity data from one subject, processed in a single data stream is used for the demonstration.

Figure 7 illustrates the concept drifts detected from the activity dataset. Each unsupervised concept drift was mapped to the labelled activity as shown in Table 2. CD6 and CD8 were identified as reoccurring concept drifts, which was confirmed by the labels ‘Ascending stairs → Descending stairs’. CD5 resembling vacuum cleaning is a gradual concept drift [54] where the drift happens during a period of time. Further experiments on the data showed that the subject’s heart rate gradually increased during this period due to the activity. The algorithm proposed in the paper cannot detect gradual concept drifts. This has also been noted in sect. 6 as a future work.

Further, the multi-dimensional generalization nodes (explained in Sect. 3) are visualized using Sammon’s mapping [55], a nonlinear projection technique that preserve correlations among nodes, to understand the concept drift detection (Fig. 8). Each activity is learnt in several execution iterations and is denoted by several generalization nodes. Generalization nodes mapped to an activity are clustered together, and low-intensity activities and high-intensity activities are separated in the feature space. Hence, Sammon’s mapping results confirm the learning of the concept drift detection and adaptation are accurate.

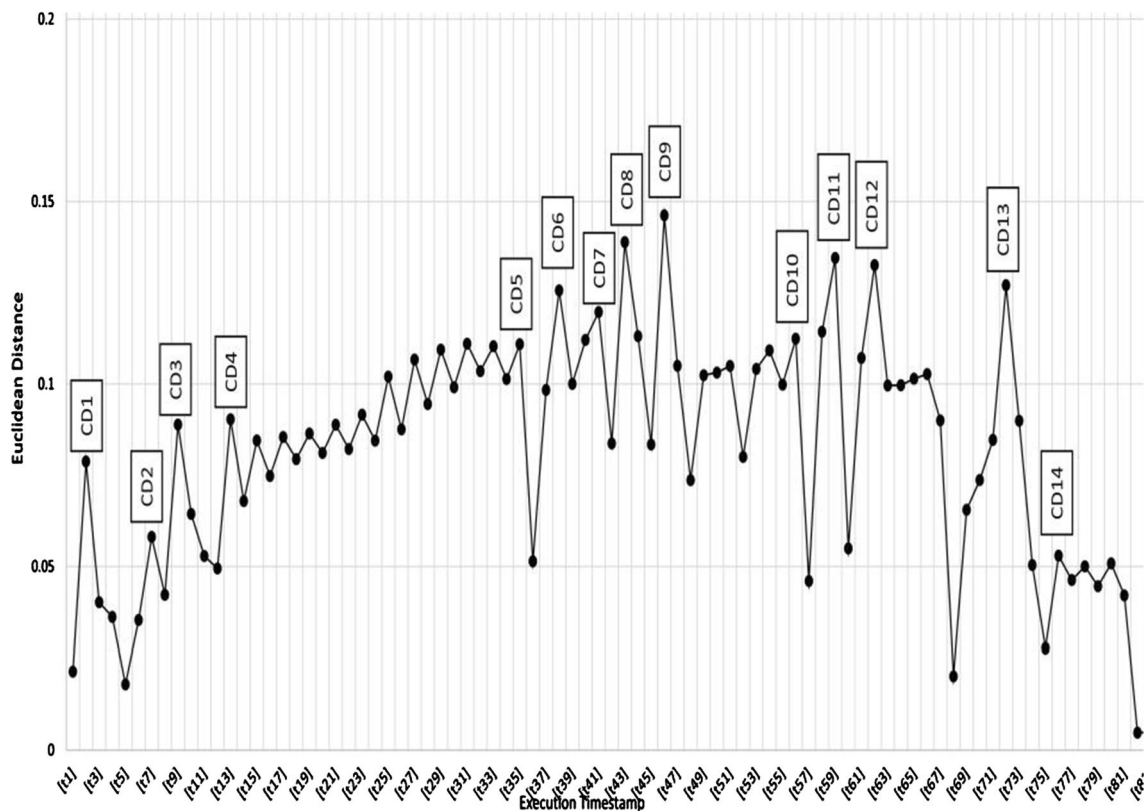


Fig. 7 Activity dataset—concept drift detection

Table 2 Activity mapping for concept drift detection

Concept drift	Activity change
CD1	Lying → sitting
CD2	Sitting → standing
CD3	Standing → ironing
CD4	Ironing → vacuum cleaning
CD5	Vacuum cleaning → ascending stairs
CD6	Ascending stairs → descending stairs
CD7	Descending stairs → ascending stairs
CD8	Ascending stairs → descending stairs
CD9	Descending stairs → walking
CD10	Walking → nordic walking
CD11	Nordic walking → cycling
CD12	Cycling → running
CD13	Running → rope jumping
CD14	Not a concept drift

In this labelled dataset, performance of concept drift detection is evaluated with the indicators defined respectively by:

$$\text{Precision} = \frac{\text{correct number of concept drift detection}}{\text{total number of concept drift detection}}$$

$$\text{Recall} = \frac{\text{correct number of concept drift detection}}{\text{number of TRUE concept drifts}}$$

$$F_Score = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$

These indicators provide an overview of the abrupt and reoccurring concept drift detection, where precision is the probability of a concept drift detection is a true positive; recall is the probability that a true positive concept drift is detected; F_score is a comprehensive indicator which is the harmonic mean between precision and recall. The accuracy of abrupt and reoccurring concept drift detection for all nine subjects are as shown in Table 3, accuracy has significantly improved above the baseline performance of 90% stated in CPS literature [3].

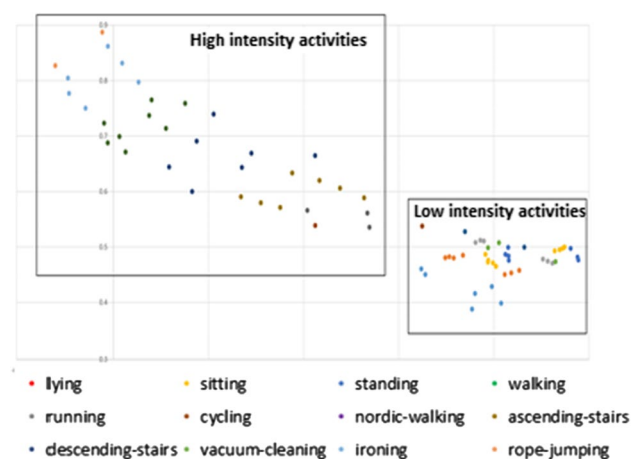
Fig. 8 Sammon's mapping [28] of generalization nodes

Table 3 Accuracy for concept drift detection in activity dataset

	Precision (%)	Recall (%)	F1
Abrupt concept drift detection	89.31	93.80	91.50
Reoccurring Concept Drift Detection	97.20	94.00	95.57

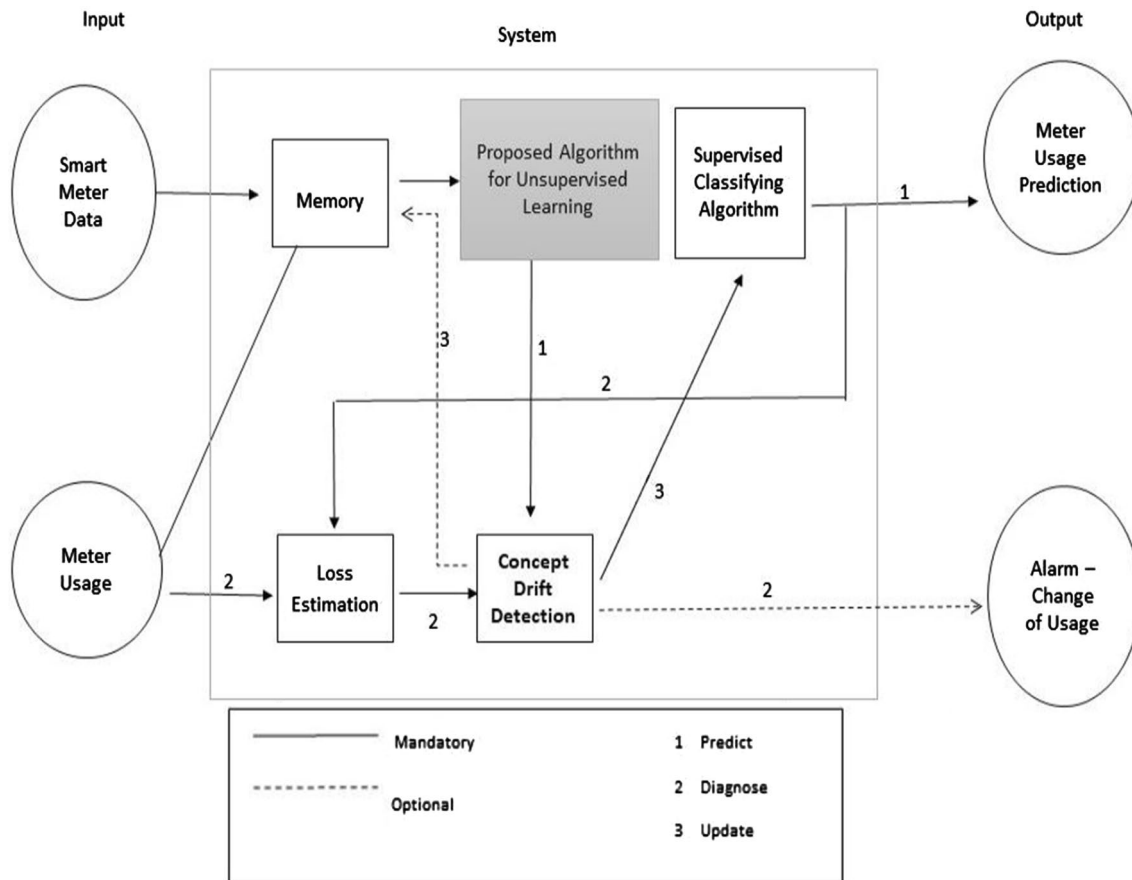


Fig. 9 Application of the extended framework for electricity consumption dataset

5.2 Industrial energy consumption

Smart meters are widely used for energy consumption recording in industrial settings and frequently linked to the CPS data streams for overall monitoring of a smart factory. This dataset contains measurements of electricity consumption at a one-minute sampling rate, for four years, between December 2006 and November 2010 [56]. The extended framework was tested with this dataset to identify daily and monthly patterns (Fig. 9).

Figure 10 demonstrates concept drift detection of daily pattern recognized through concept drift detection. Figure 10a denotes concept drifts (reoccurring and abrupt) detected through one week. The section highlighted in Fig. 10a illustrates the concept drifts detected on Sunday, 17th December 2006. The reason for the concept drift is demonstrated in Fig. 10b–d and outlined in Table 4. Usage of sub-meter-1 (kitchen appliances) at approximate timepoints; 10.30 a.m. and 2.30 p.m. have been detected as CD3 and CD4 respectively (Fig. 10b). Usage of sub-meter-2 (laundry room appliances) at approximate timepoints; 1 a.m. and 10.30 a.m. has been detected as CD1 and CD3 respectively (Fig. 10c). Usage of sub-meter-3 (water heater and air-conditioner) at approximate timepoints; 5 a.m. and between 10.30 a.m. and 10 p.m. has been detected as CD2 and CD3 respectively (Fig. 10d).

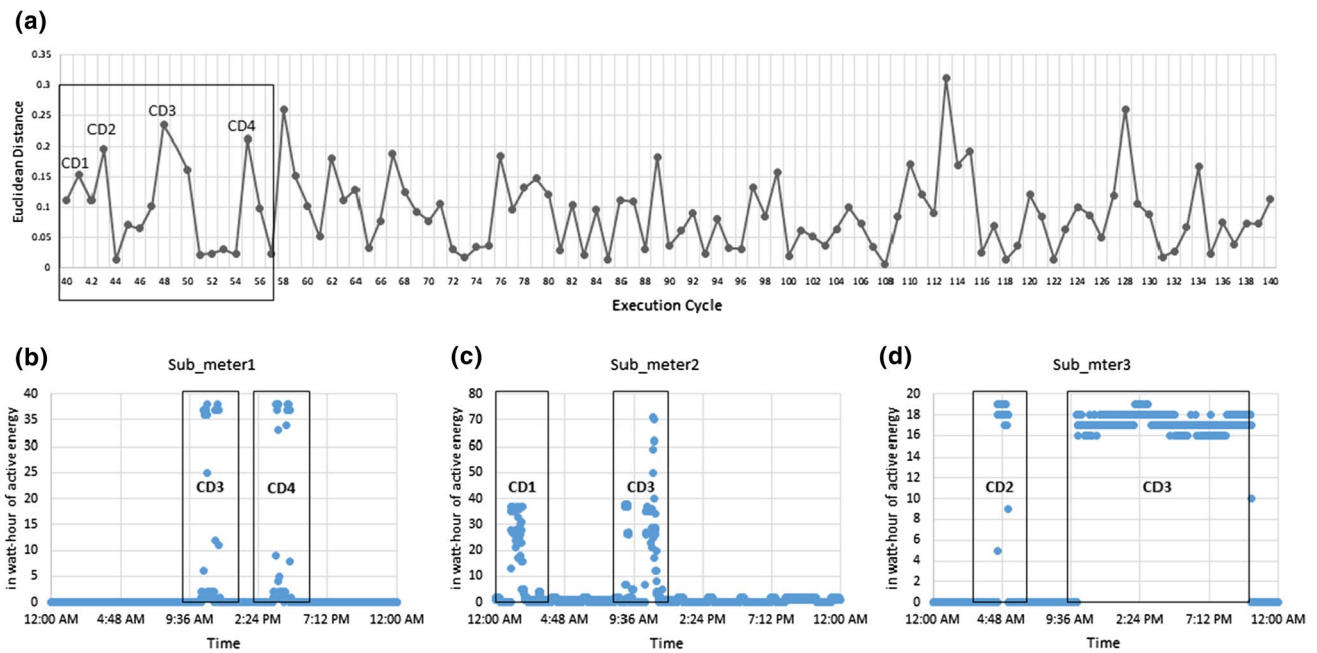


Fig. 10 **a** Concept drift detection for the entire energy consumption data stream, Hourly electricity usage (segment marked in concept drift detection diagram); **b** sub_meter_1—kitchen appliances (dishwasher, oven, and microwave), **c** Sub_meter_2—laundry room (washing machine, tumble dryer, refrigerator and, a light), **d** Sub_meter_3—electric water heater and air-conditioner

Table 4 Detected concept drifts and associated descriptions

Concept drift	Timestamp	Description
CD1	[41]	Increase in sub-meter_2 around midnight
CD2	[43]	Increase in sub_meter_3 around 5 a.m.
CD3	[48]	Increase in sub_meter_1, sub_meter_2 and sub_meter_3 around 9.30 a.m.
CD4	[55]	Increase in sub_meter_1 around 2.30 p.m.

6 Conclusion

CPS data streams of industrial applications generate large volumes of data at high velocity for real-time monitoring of the corresponding physical entities. The detection of dynamic and abrupt changes (formally defined as concept drifts) in these time-critical and mission-critical systems is a complex challenge. In this paper, we proposed a new unsupervised, incremental machine learning algorithm to detect and adapt to concept drifts and distinguish between abrupt and reoccurring drifts. We further extended a closed loop concept drift detection framework to incorporate drift detection from unlabeled data streams, such as industry CPS. The proposed algorithm exhibits three learning features; online, incremental and decremental. Experiments were conducted on a benchmark concept drift dataset, the SEA dataset, and CPS data streams from practical industrial application; activity monitoring and energy consumption. Results from all three experiments successfully demonstrate key features of the proposed algorithm in detection, adaption to concept drift and identification of abrupt and reoccurring concept drift. Extension to the concept drift detection framework was also demonstrated using the energy consumption dataset to provide classifier independent, near real-time analysis of drifts in energy usage. As future work, we intend to improve the algorithm to detect concept drifts of other types such as gradual and incremental. Furthermore, we intend to develop a methodology based on sequence analysis to determine causality of concept drift.

Authors' contributions DJ, DDS, DA, XY: Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work. DA, DDS, XY: Drafting the work or revising it critically for important intellectual content. DDS, DA, XY:

Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All authors read and approved the final manuscript.

Data availability The datasets used in this study are available from the UCI Machine Learning Repository.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Alippi C, Roveri M. The (Not) far-away path to smart cyber-physical systems: an information-centric framework. *Computer*. 2017;50(4):38–47.
2. Dong F, Zhang G, Lu J, Li K. Fuzzy competence model drift detection for data-driven decision support systems. *Knowl Based Syst*. 2018;143:284–94.
3. Lin CC, Deng DJ, Kuo CH, Chen L. Concept drift detection and adaption in big imbalance industrial IoT data using an ensemble learning method of offline classifiers. *IEEE Access*. 2019;7:56198–207.
4. Krawczyk B. Active and adaptive ensemble learning for online activity recognition from data streams. *Knowl Based Syst*. 2017;138:69–78.
5. Di Mauro M, Maggioni MF, Grasso M, Colosimo BM. Design performance analysis of a self-organizing map for statistical monitoring of distribution-free data streams. *Procedia CIRP*. 2016;41:448–53.
6. Gama J, Žliobaite IE, Bifet A, Pechenizkiy M, Bouchachia A. A survey on concept drift adaptation. *ACM Comput Surv*. 2014;46(4):1–37.
7. Leitao P, Karnouskos S, Ribeiro L, Lee J, Strasser T, Colombo AW. Smart agents in industrial cyber–physical systems. *Proc IEEE*. 2016;104(5):1086–101.
8. Cao K, Hu S, Shi Y, Colombo A, Karnouskos S, Li X. A survey on edge and edge-cloud computing assisted cyber-physical systems. *IEEE Trans Ind Inf*. 2021. <https://doi.org/10.1109/tii.2021.3073066>.
9. Hu S, Shi Y, Colombo A, Karnouskos S, Li X. Cloud-edge computing for cyber-physical systems and internet-of-things. *IEEE Trans Ind Inf*. 2021. <https://doi.org/10.1109/tii.2021.3064881>.
10. Napoleone A, Macchi M, Pozzetti A. A review on the characteristics of cyber-physical systems for the future smart factories. *J Manuf Syst*. 2020;54:305–35.
11. Kurliej B, Wozniak M. Active learning approach to concept drift problem. *Log J IGPL*. 2012;20(3):550–9.
12. Alippi C, Roveri M. Just-in-time adaptive classifiers—Part II: designing the classifier. *IEEE Trans Neural Netw*. 2008;19(12):2053–64.
13. Alippi C, Boracchi G, Roveri M. Change detection tests using the ICI rule. In: *The 2010 international joint conference on neural networks (IJCNN)*. New York: IEEE; 2010. p. 1–7.
14. Hulten G, Spencer L, Domingos P. Mining time-changing data streams. In: *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*; 2001. p. 97–106.
15. Cohen L, Avrahami-Bakish G, Last M, Kandel A, Kipersztok O. Real-time data mining of non-stationary data streams from sensor networks. *Inf Fus*. 2008;9(3):344–53.
16. Baena-Garcia M, del Campo-Ávila J, Fidalgo R, Bifet A, Gavalda R, Morales-Bueno R. Early drift detection method. In: *Fourth international workshop on knowledge discovery from data streams*, vol. 6; 2006. p. 77–86.
17. Street WN, Kim Y. A streaming ensemble algorithm (SEA) for large-scale classification. In: *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*; 2001. p. 377–82.
18. Bifet A, Holmes G, Pfahringer B, Kirkby R, Gavalda R. New ensemble methods for evolving data streams. In: *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*; 2009. p. 139–48.
19. Tsybmal A, Pechenizkiy M, Cunningham P, Puuronen S. Dynamic integration of classifiers for handling concept drift. *Inf Fus*. 2008;9(1):56–68.
20. Kolter JZ, Maloof MA. Dynamic weighted majority: an ensemble method for drifting concepts. *J Mach Learn Res*. 2007;8:2755–90.
21. Abdulsalam H, Skillicorn DB, Martin P. Classification using streaming random forests. *IEEE Trans Knowl Data Eng*. 2010;23(1):22–36.
22. Bifet A et al. Moa: massive online analysis, a framework for stream classification and clustering. In: *Proceedings of the first workshop on applications of pattern analysis*. PMLR; 2010. p. 44–50.
23. de Mello RF, Vaz Y, Grossi CH, Bifet A. On learning guarantees to unsupervised concept drift detection on data streams. *Expert Syst Appl*. 2019;117:90–102.
24. Yang Z, Al-Dahidi S, Baraldi P, Zio E, Montelatini L. A novel concept drift detection method for incremental learning in nonstationary environments. *IEEE Trans Neural Netw Learn Syst*. 2019;31(1):309–20.
25. Gözüaçık Ö, Büyükkacı A, Bonab H, Can F. Unsupervised concept drift detection with a discriminative classifier. In: *Proceedings of the 28th ACM international conference on information and knowledge management*; 2019. p. 2365–8.
26. Zenisek J, Holzinger F, Affenzeller M. Machine learning based concept drift detection for predictive maintenance. *Comput Ind Eng*. 2019;137:106031.

27. Liu S, Feng L, Wu J, Hou G, Han G. Concept drift detection for data stream learning based on angle optimized global embedding and principal component analysis in sensor networks. *Comput Electr Eng*. 2017;58:327–36.
28. Mehmood H, Kostakos P, Cortes M, Anagnostopoulos T, Pirttikangas S, Gilman E. Concept drift adaptation techniques in distributed environment for real-world data streams. *Smart Cities*. 2021;4(1):349–71.
29. Diaz-Rozo J, Bielza C, Larrañaga P. Machine learning-based CPS for clustering high throughput machining cycle conditions. *Procedia Manuf*. 2017;10:997–1008.
30. Wang H. Research on real-time reliability evaluation of CPS system based on machine learning. *Comput Commun*. 2020;157:336–42.
31. Zhao X, Zeng X, Koehl L, Tartare G, de Jonckheere J, Song K. An IoT-based wearable system using accelerometers and machine learning for fetal movement monitoring. In: 2019 IEEE international conference on industrial cyber physical systems (ICPS). New York: IEEE; 2019. p. 299–304.
32. Wang J, Tu W, Hui LC, Yiu SM, Wang EK. Detecting time synchronization attacks in cyber-physical systems with machine learning techniques. In: 2017 IEEE 37th international conference on distributed computing systems (ICDCS). New York: IEEE; 2017. p. 2246–51.
33. Junejo KN, Goh J. Behaviour-based attack detection and classification in cyber physical systems using machine learning. In: Proceedings of the 2nd ACM international workshop on cyber-physical system security; 2016. p. 34–43.
34. Furoo S, Hasegawa O. An incremental network for on-line unsupervised classification and topology learning. *Neural Netw*. 2006;19(1):90–106 (in Eng).
35. Mouchaweh MS, Devillez A, Lecolier GV, Billaudel P. Incremental learning in Fuzzy Pattern Matching. *Fuzzy Sets Syst*. 2002;132(1):49–62.
36. Navarro-Gonzalez JL, Lopez-Juarez I, Ordaz-Hernandez K, Rios-Cabrera R. On-line incremental learning for unknown conditions during assembly operations with industrial robots. *Evol Syst*. 2015;6(2):101–14 (in en).
37. Sayed-Mouchaweh M. Learning in dynamic environments. In: Learning from data streams in dynamic environments. SpringerBriefs in applied sciences and technology; 2016. p. 11–32.
38. De Silva D, Alahakoon D. Incremental knowledge acquisition and self learning from text. In: The 2010 International Joint Conference on Neural Networks (IJCNN); 2010. p. 1–8.
39. Bandaragoda T, De Silva D, Alahakoon D. Automatic event detection in microblogs using incremental machine learning. *J Assoc Inf Sci Technol JASIST*. 2017;68:2394–411.
40. De Silva D, Yu X, Alahakoon D, Holmes G. A data mining framework for electricity consumption analysis from meter data. *IEEE Trans Ind Inf*. 2011;7(3):399–407.
41. De Silva D, Yu X, Alahakoon D, Holmes G. Incremental pattern characterization learning and forecasting for electricity consumption using smart meters. In: 2011 IEEE international symposium on industrial electronics, 2011; 2011. p. 807–12.
42. De Silva D, Yu X, Alahakoon D, Holmes G. Semi-supervised classification of characterized patterns for demand forecasting using smart electricity meters. In: 2011 international conference on electrical machines and systems, 2011; 2011. p. 1–6.
43. Nallaperuma D, Silva DD, Alahakoon D, Yu X. A cognitive data stream mining technique for context-aware IoT systems. In: IECON 2017—43rd annual conference of the IEEE industrial electronics society, 2017; 2017. p. 4777–82.
44. Nallaperuma D, Silva DD, Alahakoon D, Yu X. Intelligent detection of driver behavior changes for effective coordination between autonomous and human driven vehicles. In: IECON 2018—44th annual conference of the IEEE industrial electronics society, 2018; 2018. p. 3120–5.
45. Nallaperuma D, et al. Online incremental machine learning platform for big data-driven smart traffic management. *IEEE Trans Intell Transp Syst*. 2019. <https://doi.org/10.1109/TITS.2019.2924883>.
46. Câmpian A, Șerban G. Adaptive clustering algorithms. *Adv Artif Intell*. 2006;2006:407–18.
47. Gonçalves PM Jr, de Carvalho Santos SGT, Barros RSM, Vieira DCL. A comparative study on concept drift detectors. *Expert Syst Appl*. 2014;41(18):8144–56.
48. Street WN, Kim Y. A streaming ensemble algorithm (SEA) for large-scale classification. In: Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining. New York, NY, USA, 2001; 2001, p. 377–82.
49. Bifet A, Holmes G, Kirkby R, Pfahringer B. MOA: Massive Online Analysis. *J Mach Learn Res*. 2010;11:1601–4.
50. Koskimaki H, Huikari V, Siirtola P, Laurinen P, Roning J. Activity recognition using a wrist-worn inertial measurement unit: a case study for industrial assembly lines. In: 2009 17th mediterranean conference on control and automation, 2009; 2009. p. 401–5.
51. Lukowicz P, Timm-Giel A, Lawo M, Herzog O. WearIT@work: toward real-world industrial wearable computing. *IEEE Pervas Comput*. 2007;6(4):8–13.
52. Kortuem G et al. Sensor networks or smart artifacts? An exploration of organizational issues of an industrial health and safety monitoring system. In: UbiComp 2007: ubiquitous computing, 2007; 2007. p. 465–82.
53. Reiss A, Stricker D. Creating and benchmarking a new dataset for physical activity monitoring. In: Proceedings of the 5th international conference on Pervasive technologies related to assistive environments, New York, NY, USA, 2012; 2012; 2012. p. 40:1–40:8.
54. Duan F, Dai L. Recognizing the gradual changes in sEMG characteristics based on incremental learning of wavelet neural network ensemble. *IEEE Trans Ind Electr*. 2017;64(5):4276–86.
55. Sammon JW. A nonlinear mapping for data structure analysis. *IEEE Trans Comput*. 1969;C-18(5):401–9.
56. M. Lichman, "{UCI} Machine Learning Repository," 2013 2013.