Concept Drift Evolution In Machine Learning Approaches: A Systematic Literature Review

Manzoor Ahmed Hashmani[1],
Syed Muslim Jameel[2,*],
Mobashar Rehman[3] and
Atsushi Inoue[4]

[1]Center for Research in Data Science (CERDAS), UTP High-Performance Cloud Computing Centre (HPC3), UTP, Senior Member of Institute of Electrical and Electronics Engineers (IEEE), Department of Computer and Information Sciences, Universiti Teknologi PETRONAS (UTP), Perak, Malaysia.

[2]Department of Computer and Information Sciences Universiti Teknologi PETRONAS (UTP), Perak, Malaysia.

[3]Faculty of Information and Communication Technology (FICT), Universiti Tunku Abdul Rahman Perak, Malaysia.

[4]Department of Computer Science, Eastern Washington University, Cheney, WA 99004, United States.

*E-mail: muslim_16000370@utp.edu.my

Abstract

Concept Drift's issue is a decisive problem of online machine learning, which causes massive performance degradation in the analysis. The Concept Drift is observed when data's statistical properties vary at a different time step and deteriorate the trained model's accuracy and make them ineffective. However, online machine learning has significant importance to fulfill the demands of the current computing revolution. Moreover, it is essential to understand the existing Concept Drift handling techniques to determine their associated pitfalls and propose robust solutions. This study attempts to summarize and clarify the empirical pieces of evidence of the Concept Drift issue and assess its applicability to meet the current computing revolution. Also, this study provides a few possible research directions and practical implications of Concept Drift handling.

Keywords

Big data analysis, Concept drift, Nonstationary environment, Adaptive machine learning, Online learning.

Big Data (BD) is participating in the current computing revolution immensely. Industries and organizations are utilizing their insights for Business Intelligence using Machine Learning (ML) models. However, BD's dynamic characteristics introduce many critical issues for ML models, such as the Concept Drift (CD) issue. The issue of CD is observed when the statistical properties of data vary at a different time step. For example, a set of class examples has legitimate class labels at one time step and various labels at another time step, which substantially decreases the performance in terms of accuracy in image classification models (ICM) (Jameel et al., 2018).

CD issue frequently appears in Online Learning scenarios in which data trends change over time. The problem may even worsen in the BD environment due to veracity and variability factors. Due to the CD issue, the accuracy of classification results degrades in ML models, making ML models not applicable for further use. Therefore, ML models need to adapt quickly to

changes to maintain the accuracy level of the results. Since the last decade, the issue of CD has gained significant attention from the research community. Initially, many studies discuss the issue of CD in a stable environment. However, after BD analysis of the nonstationary environment, CD's meaning and taxonomy have changed, and researchers proposed different adaptation strategies for this newly emerging research area (Mehta, 2017). In the existing literature, most of the proposed solutions utilize the Extreme Learning Machine (ELM), Support Vector Machine (SVM), and Convolutional Neural Network (CNN) as base classifiers. These solutions' configurations are mostly a single classifier or an ensemble classifier (Zliobaite, 2010; Jameel et al., 2020a, b, c; Uddin et al., 2019). However, the ensemble classifier considers an appropriate solution than a single classifier to improve the classification performance after a CD.

Nevertheless, the ensemble approach does not adapt to numerous drift cases (Liu and Wang, 2010; van Schaik and Tapson, 2015). The adaptive classifiers can handle this issue in a better way. Few recent studies concentrated on adaptive learning techniques using ELM based single classifiers (van Schaik and Tapson, 2015; Budiman et al., 2016; Huang et al., 2012) and ensemble classifiers for CD adaptation (Zhai et al., 2014; Xu and Wang, 2016, 2017). For example, Incremental Data Stream ELM used an incremental approach to train the classifier. In this approach, the number of neurons in hidden layers and the selection of the activation layer are dynamic, enhancing the performance of the model. In contrast, this approach handles stream data for gradual drift scenario only (Xu and Wang, 2016). A substantial improvement in accuracy and adaptability is needed to make ML models robust in a nonstationary environment in current solutions.

Since the Concept Drift has significance in various critical applications and gained the researcher's attention from the last decade. Besides the fact, several foster studies well discuss many Concept Drift detection and adaptation techniques, but the consolidated information on this issue is not available in the existing literature. In a recent survey, Iwashita et al. (2019) present an overview of Concept Drift Learning in which authors mainly discussed adaptation and detection techniques and CD datasets used by past studies. However, in this study, the authors do not provide a comparative analysis of available adaptation and detection techniques and protective research directions of CD issues.

In literature, despite the considerable numbers of empirical studies on Concept Drift detection and adaptation in ML models, few inconsistent results have been reported regarding the performance accuracy of ML models that depict the provided solutions are not generic and are most feasible for the particular type of data set. Interestingly, it is also impossible to develop a generalized approach to detect and handle all kinds of Concept Drift. Moreover, CD's theory is not much clear for more complex types of data streams, for example, Imagery Streams. It is also crucial to summarize the empirical evidence on the practical implications and highlight the upfront potential challenges. The main contribution of this Systematic Literature Review are;

1. To investigate the CD fundamentals and current state of the art for CD handling techniques.
2. To identify the shortcomings (of existing CD handling approaches) and future research.

More precisely, this study summarizes existing literature related to the Concept Drift issue and provides the researchers with a road map to better contribute to this knowledge area.

The remainder of this paper is organized as follows. Section 2 presents the methodology of this paper. Section 3 offers the core contribution of this paper and states the research outcomes in more detail; this section gives the feature answers of designed research questions through rational justification. Section 4 presents the conclusion.

Methodology

This Systematic Literature Review (SLR) follows the review protocol to investigate the required outcomes and research objectives, as mentioned in (Kitchenham and Charters, 2007). Review protocol contains six (6) different phases, and each phase is a step towards the authentic pieces of evidence and quality assurance measures. Furthermore, to extract most of the relevant information for this topic, this study followed the PRISMA guidelines for systematic selection of relevant articles, as depicted in Figure 1. These six (6) phases of the review protocol illustrated in Figure 2. Phase 1 formulates the two (2) research questions to fulfill the objectives of this SLR. These research questions act as the pivot of this SLR. Therefore, the research questions' outcomes comprehend all relevant details related to the Concept Drift issue in Machine Learning models. Phase 2 defines the search strategy; this phase determines the proper search term, optimal literature sources, and adequate literature process to search from electronic databases systematically. The search strategy is performed by the first author (Syed Muslim Jameel) and third author (Dr. Mobashar
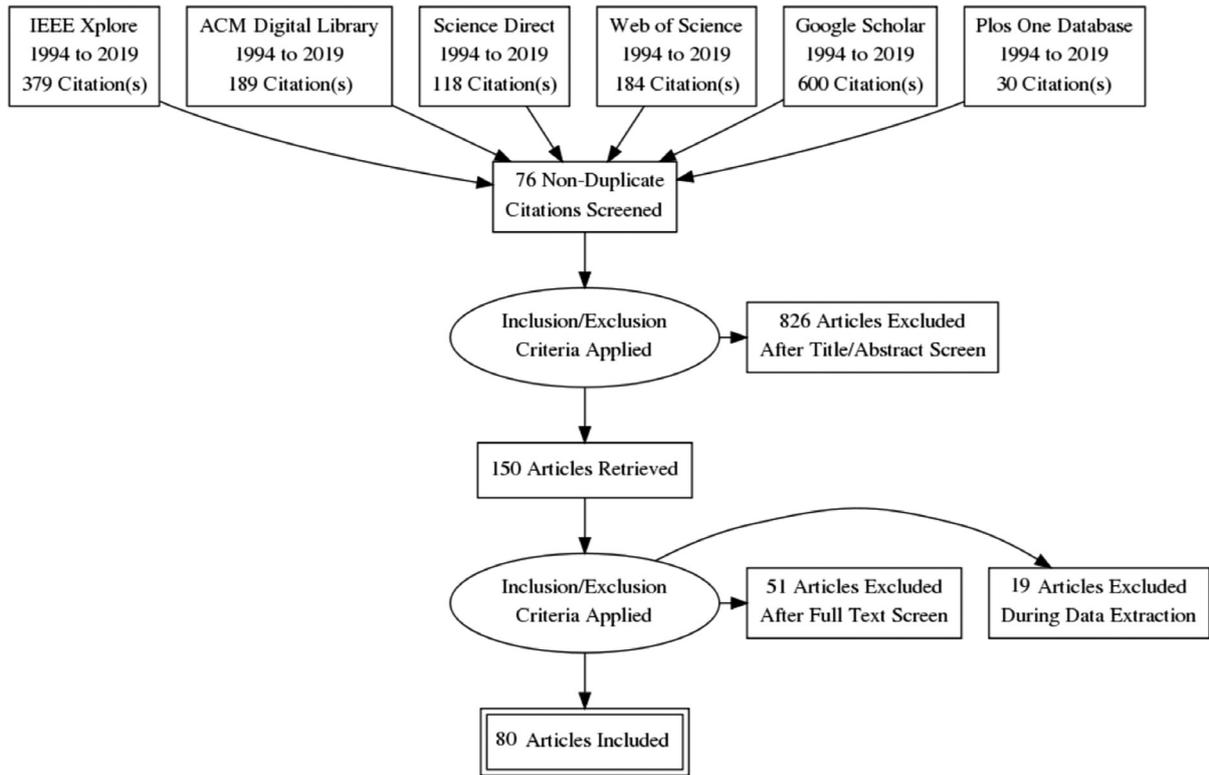
Figure 1: Flow chart for systematic selection of relevant articles using PRISMA guideline.

Figure 2: Six (6) phases of review protocol.

Rehman). Phase 3 discusses the appropriate selection criteria to segregate the studies which address the research questions. In Phase 4, the selected studies undergo quality assessment according to the established quality criteria.

Phase 5 is the data analysis phase, which defines the systematic process to figure out pertinent details addressing the research questions of this SLR. Phase 6 assembles all the obtained empirical evidence to justify the answers to research questions; this phase is essential because, in some cases, a few weak pieces of evidence compositely establish a strong justification of inquired research questions in this SLR. During each phase, the second author (Prof. Dr. Manzoor Ahmed Hashmani) performed his duty as the referee to resolve the possible conflict between the first and second authors.

Research questions

This SLR aims to summarize and clarify the empirical evidence towards understanding the Concept Drift issue and Machine Learning. To achieve the objective

of this study, the following four research questions were formulated.

Search strategy

The search strategy comprises search terms, literature resources, and search processes, which are detailed as follows;

Primary search terms and derived search terms

The keywords are used to broaden the search criteria. These keywords are searched from title, abstract, to the full text of the paper. Five (05) basic subject terms were used in the primary search to focus the base research papers. Later, a further twenty-four (24) different search terms were identified, called derived search terms. These twenty-four (24) search terms are the synonyms of the five (05) base terms. These derived search terms were formulated from the keywords used in search research papers using the primary search terms. The primary search term Concept Drift, Online Learning and Machine Learning, Adaptive Model, and Big Data are having five derived terms, each shown in Table 2.

The relevant literature search is a critical phase to dig out all the relevant literature of the area of the Concept Drift issue. This study took some significant steps. Initially, the search terms were derived from the research questions, and their synonyms were identified. Moreover, Boolean OR and AND were used to link the critical terms, as defined below. For example, Machine Learning AND (stream OR online OR real-time) AND (classification OR clustering) AND (concept drift OR Concept change OR "dynamic changes," OR "adaptivity."

Article resources

The majority of the research papers were acquired from well-reputed high-quality journal papers from the electronic databases, including IEEE Digital Library, ACM Digital Library, Science Direct, Web of Science, Google Scholar, and PLOS One database. The literature examined was from the duration of 1994 to 2019. Also, few research papers for supporting citation be taken from before 1994. Nevertheless, most of the relevant research papers are found from 2007 to 2019.

Article exploration process

Article exploration or search process is dependent on four phases. Phase 1 was dedicated to searching Fifteen-Hundred (1500) research papers from the most reputed electronic libraries based on search terminologies. In this phase, all the reputed digital libraries, including the IEEE Digital Library, ACM Digital Library, Science Direct, Web of Science, Google Scholar, and PLOS One database, were utilized. Twenty-nine (29) search terms were used to collect all research papers. Phase 2 contained a systematic process. This process segregated the relevant and non-relevant papers among Fifteen Hundred (1500) acquired research papers during phase 1.

In this process, the title, abstract, and conclusion parts were investigated and filtered. Further, one-hundred-fifty (150) papers were found relevant to the subject matter. In this phase, the articles' abstract and conclusion sections were the main drivers for further scrutiny. However, phase 3 details the selected papers analyzed through the Quality Assessment Criteria (QAC), and eighty (80) relevant articles were filtered. In this phase, the diagonal reading strategy was used to examine the chosen articles deeply. Phase 4 classified the candidate papers following research questions. Like against, twenty-nine (29) research papers were identified for RQ1 and forty-six (46) for RQ2. Also, few research papers were found to lie in the multiple research questions category, as described in Table 1. Phase 4 critically examined the selected articles and formulized the required outcomes of this study Table 2.

Study identification and selection

This study follows the recommendations presented in (Kitchenham and Charters, 2007; Kitchenham, 2004; Petersen et al., 2008). These studies are the standard reference for study identification and selection process in the Computer Science domain. However, this study also used some defined inclusion and exclusion criteria. The inclusion and exclusion criteria draw the boundary line towards the study selection, which is essential to ensure an unbiased and quality search.

Inclusion criteria

1. The title or abstract must clearly express that the research papers are pertinent to the study domain.
2. The research paper is explicitly related to the Concept Drift issue in the Machine Learning domain.
3. The research paper addresses the research questions of this study or provides any

Table 1. Research questions and their subsequent research objectives.

S.No.	Research questions	Research objectives	References
1	What are the fundamentals of the Concept Drift (CD) issue?	To provide an overview of the basics of CD and determine how CD fundamentals changed over time. Also, highlight CD measuring, quantification techniques, and possible ways to overcome CD.	(Budiman et al., 2016; Iwashita et al., 2019; Jameel et al., 2020a, b, c; Budiman et al., 2017; Zang et al., 2014; Zliobaite et al., 2014; Kuncheva, 2004; Ghorbani et al., 2017; Gupta and Dhawan, 2019; Jensen et al., 2019; Nishida et al., 2008; Harel et al., 2014; Dyer and Polikar, 2012; Khamassi et al., 2019; Saurav et al., 2018; Dongre and Malik, 2014; Dariusz, 2010; Sayed et al., 2018; Wadewale and Desai, 2015; Brzezinski and Stefanowski, 2014a; Hoens et al., 2012; Jagadeesh Chandra Bose et al., 2011; Huang et al., 2013; Minku et al., 2010; Tsymbal, 2004; Gomes et al., 2011; Hoens et al., 2011; Webb et al., 2016, 2018) Total of 29 papers
2	Do the state of art approaches (for CD handling) are adequate for current and future computing trends?	To investigate the existing CD handling approaches and determine their effectiveness and shortcomings for current and future trends.	(Jameel et al., 2020, 2020; Uddin et al., 2019, September; Liu and Wang, 2010; van Schaik and Tapson, 2015; Budiman et al., 2016; Huang et al., 2012; Zhai et al., 2014; Xu and Wang, 2016; 2017; Khamassi et al., 2019; PAC learning model Kearns and Vazirani, 1994; Gama et al., 2004; Baena-Garcıa et al., 2006; Lavaire et al., 2015; Friedman and Rafsky, 1979; Spinosa et al., 2007; Zeira et al., 2004; Kifer et al., 2004; Demšar and Bosnić, 2018; Page, 1954; Mouss et al., 2004; Yasumura et al., 2007; Freund and Schapire, 1997; Bach and Maloof, 2008; Bifet, 2009; Bifet and Gavalda, 2007; Nishida, 2008; Ross et al., 2012; Raza et al., 2014; Rouse, 2009; Ditzler and Polikar, 2013; Zliobaite et al., 2012; Huang, 2006; Liang et al., 2006; Lan et al., 2009; Liu et al., 2019; Krawczyk, 2015; Cao et al., 2015; Khamassi et al., 2015; Brzezinski and Stefanowski, 2012; Sidhu and Bhatia, 2018; Bifet et al., 2009; Wang et al., 2003; Brzeziński and Stefanowski, 2011; Brzezinski and Stefanowski, 2014b; Street and Kim, 2001) Total of 46 papers

Table 2. Primary and derived search terms for relevant research paper elicitation.

Search Terms				
Primary	Concept Drift	Online learning	Machine Learning	Adaptive model	Big Data
Derived	Nonstationary features	Fast Learning	Supervised	Self-regulatory	Continuous data
	Variability and Veracity	Real-Time Learning	Unsupervised	Dynamic model	Stream data
	Conceptual change	Adaptive Learning	Clustering	Meta-Cognitive model	Unbalanced data
	Concept Shift	Dynamic Learning	Classification	Robust model	Complex data
	Feature Variability	Continuous Learning	Regression		Evolving stream

empirical evidence for the investigated query's support.

4. The research paper must belong to a conference paper, journal paper, book chapter, or thesis report.
5. The research paper must be between 1994 to 2019.

Exclusion criteria

1. The research paper must not be in any other language except the English language.
2. The research paper must not belong to an editorial, white papers, introduction to proceedings, poster presentation, or symposium reports.
3. Any research paper which does contain personal biased of the author.
4. The research paper is not relevant to Concept Drift or Machine Learning.

Study quality assessment

The study selection and search criteria do not guarantee the quality of the article. Therefore, this study defines seven Quality Assessment Criteria (QAC) questions to ensure the selected research paper's credibility and quality. Does the research paper clearly define the aim, objectives, methodology? Does the research paper adequately refer to the reputed literature to prove its assumptions or hypothesis (if any)? Do experimental results convey the claimed contribution by the article? Does the research paper use the appropriate experimental environment? Do the selected datasets illustrate the Concept Drift issue? Does the research paper conclude the study?

Data extraction

This Systematic Literature Review (SLR) exploits the relevant articles that address this study's research questions. However, few articles do not directly act as evidence for the problem area but are essential for supporting evidence. Besides, these papers were not relevant to the subject matter.

Data synthesis

The goal of data synthesis is to aggregate evidence from the selected studies for answering the research questions. A single piece of evidence might have small evidence force, but the aggregation of many of them can make a point stronger (Pfleeger, 2005). The data extracted in this review include quantitative data

(e.g., values of estimation accuracy) and qualitative data (e.g., strengths and weaknesses of Concept Drift adaptation techniques).

Results and Discussions

RQ1: What are the fundamentals of the Concept Drift (CD) issue?

The taxonomy and types of CD issues are well defined in numerous studies (Jameel et al., 2020a, b, c). However, these studies do not discuss their quantification and measuring methods (essential to handling CD issues). Therefore, this research question investigates causes, quantification, and measurement techniques. Many assumptions in ML are being used in static data (Budiman et al., 2017). However, the current trends demand the analysis (using ML) in the non-static assumption or online machine learning where dynamic conditions of data changes are often. Therefore, due to the addition of new data features, ML models degrade their performance accuracy or could fail to classify or predict the correct output. Notably, in Supervised Online ML, the model is learned through the input and output features from data of one-time span and will likely predict or classify the output (class category) from another time. The change in features (among both periods) are due to various conditions. It could be due to the data format (variety), distribution (variability), or sources (complexity), which change over time. Another term for Concept Drift refers to the classification boundary or clustering centers that continuously change with time elapsing (Zang et al., 2014). These conditions will adversely affect the classification performance of the model. In studies, the CD is modeled based on Bayesian decision theory for class output 'c' and input data X, as shown below. Zliobaite et al. (2014)

$$P(c/X) = P(c)P(X/C)/P(X), \qquad (1)$$

Where P(c/X), P(c), P(X/c), and P(X) are posterior, prior, conditional, and feature-based probabilities, respectively (Budiman et al., 2017). One of the possible conditions is Real Concept Drift. Real Concept Drift arises when P(c/X) undergoes changes and causes a shift in the class boundary (conditional probabilities). In this condition, the number of output classes may change (Zliobaite et al., 2014). Furthermore, suppose the P (X) (feature-wise distribution of data changes) is due to insufficient or partial feature representation of existing data distribution (new additional feature adds or some feature updates). In that case, it is called Virtual Drift (Zliobaite et al., 2014). Also, a study

introduces Hybrid Drift as a condition P(c/X), and P(X) occurred consequently (Budiman et al., 2016). However, few studies discuss possible configuration patterns based on the Frequency of drift, gradual drift (when the variety of concepts changes progressively), consecutive drift (when previous concepts reoccur), and sudden drift patterns (when a concept changes/ substitutes abruptly) (Kuncheva, 2004). ML models are trained to classify according to input and output features with a predefined number of classes. Suppose a feature or class-wise distribution changes over time. In that case, ML models will face a substantial degradation in their performance (because ML models do not have prior knowledge of these changes). However, if these ML models retrain according to newly-arrived data, they cannot keep understanding of the Recurrent Context (previous training knowledge).

Today, the role of Machine Learning has found more applications in everyday life. However, the demand for online analysis has exponentially increased in several critical application domains, such as seismic analysis (Ghorbani et al., 2017), sustainability (Gupta and Dhawan, 2019), and others (Jensen et al., 2019).

Transition frequencies and concept of recurrence Concept Drift (CD)

Fundamentally, the types of CD are classified by their feature and class boundaries. However, several studies also ranked CD types with the nature of occurrence or the concept transition frequency from one concept to another. Some introductory details regarding the types of CD concerning feature-wise or class wise change distribution. The abrupt change from concept one to concept two is known as Sudden Drift. The recurrent concept is not frequently in this type of drift (Nishida et al., 2008). Gradual drift involves the progressive change from one concept to another concept. These can be small or significant changes. Gradual drift is hard to detect because the shift in the boundaries needs a more sophisticated method because of drastic differences between the nature of these changes' individualities. Sometimes, these changes could be an unvarying progression, and sometimes it could be varying and non-steady progress (Harel et al., 2014; Dyer and Polikar, 2012). Continuous drift follows a systematic pattern; these patterns repeatedly occur after a specific time interval (Khamassi et al., 2019; Saurav et al., 2018). Unlike the sudden drift, the blip drift is a spike of a new concept from the previous concept and recalls the previous concept back abruptly. The blip concept is coupled

with minimal duration (Dongre and Malik, 2014). A typical example of blip drift could be a sale promotion offer for a limited time, for example, low fare rates by airlines for their customers on the day of its completion year cycle. Notably, these changes are not frequent or continuous; this drift in customer behaviors' policies is relevant to a specific event. Therefore, due to its minor contribution towards understanding the behavior of the system or customer, some studies argued to not consider blip drift as a type of CD (Dariusz, 2010).

Several studies like (Iwashita et al., 2019; Sayed et al., 2018; Wadewale and Desai, 2015) emphasize to add all the possibilities of CD occurrence. These studies argue to consider blip concept drift as a type of Concept Drift. To consider blip drift as one of the types of CD is an entirely legitimate argument.

Moreover, the incremental drift follows a specific pattern; this pattern is based on steady progression from one concept to another by incrementing 1, each time internal the concept x steps up to x + 1. A typical example of a variation of fraud pattern is used to illustrate incremental CD (Iwashita et al., 2019). In short, CD types are based on transition frequencies, which define the manners between two concepts. Moreover, these concepts certainly also belong to either real, virtual, or hybrid type of CD. Several studies, like (Brzezinski and Stefanowski, 2014a; Hoens et al., 2012; Jagadeesh Chandra Bose et al., 2011; Huang et al., 2013; Minku et al., 2010; Tsymbal, 2004), discussed these changes in detail.

Concept drift recurrence

There are two possibilities in the concept of drift. Either the drift is a new concept or an old concept. If the drift has previously appeared, then it is called drift recurrence. Drift recurrence is more complex to adapt to the new concept because keeping the knowledge of the previous concept is enough to challenge. (Hoens et al., 2012; Jagadeesh Chandra Bose et al., 2011; Minku et al., 2010; Gomes et al., 2011). A typical example of recurrent drift can be the purchasing behavior of the customer to buy the garments. The concept of garments purchases reoccurs every winter. (Tsymbal, 2004) and (Hoens et al., 2011) discuss the concept of cyclic drift and cyclic duration. These situations arise when the drift recurrence follows a specific cycle of certain concepts and causes the drift recurrence periodicity. After the formal analysis of the current study (SLR) reveals that the Concept Drift Recurrence encompasses multiple dimensions. In (Webb et al., 2016, 2018), it is explained that the

various conditions of cyclic duration are based on the stability of certain constant parameters, such as;

1. Fixed Frequency, when the recurrent Frequency is constant.
2. Fixed Concept Duration, stability duration of a concept is constant.
3. Fixed Drift Duration, drift occurrence time is constant.
4. Fixed Drift Onset, every new concept activates at a specific time in each cycle.

Potential ways to address the Concept Drift

Static model

One of the primary ways to address the concept drift issue is using a static model. In this approach, the model is trained on a particular dataset and available for all input stream possibilities. If the input stream realizes a concept drift, then the model will decrease its performance accuracy. This approach is commonly used to validate the problem formulation of concept drift and analyze a static model's performance.

Continuous refit approach

This approach continuously updates the static model. The continuous refit approach uses the back-testing technique for required periodical updates in the model. In this approach, the model completely retrains from the new historical data.

Continues updating approach

Unlike the refit approach, this approach updates the existing static model from the current state and only learn the newly arrived changes in the input stream.

Weight data approach

It is a technique that distributes the historical data with the period. Weight each data block by the age of data — for example, more weight is assigned to the most recent data block.

Learn the change using the ensemble

After detecting the new drift in an ensemble approach, a new instance of the ensemble is added. The model does not update or refit through this approach, but another new instance learns the latest changes (Concept Drift) and becomes the part of the ensemble classifier.

Dynamic Selection (DS) approach

D.S. of the appropriate model is an approach in which several classifiers are present to handle different concept drift concepts. In this approach, after the drift is detected, it recognizes its respective classifier and uses it for prediction. However, this approach is not appropriate to adapt to the sudden concept drift.

Quantification of Concept Drift (CD)

CD significantly degrades the performance of various online models. These online models participate in several real-world applications. It is essential to figure out the quantitative measurements of CD before mitigating it. In the literature, Concept Drift detection approaches are qualitative. However, few studies discuss the quantitative measure for the characterizing of CD. These quantitative measures act as an essential prerequisite to adapt to a CD. Geoffrey I. et al. (Webb et al., 2018) proposes the novel framework, measuring the CD on a quantitative basis, and suggests the first formal quantification of concept drift, which is a solid foundation to address the problem of this nonstationary environment. Any measure of the distance between distributions could be employed. Geoffrey I. et al. used Hellinger Distance (Hoens et al., 2012) to measure the CD through the drift magnitude and the degree of difference between two-time intervals. It also highlighted the necessity of quantitative description, presented the quantitative drift mapping techniques and CD visualization methods, and used maximum likelihood estimates of the probability distribution to illustrate the concept drift. Describing drift in different attribute subspaces is to measure the drift in the marginal distributions defined over various combinations of attributes. The proposed technique approximates the drift among a two-time step, initially by approximation distributions every time step and later computing the magnitude among those drifts using the maximum Likely Hood approach. Web b and Geoffrey et al. Webb et al. (2018) claimed that their proposed measures are more practical in real-world applications. The study proposes the measuring marginal concept drift and its different variants. The quantitative measurements of drift magnitude use the total drift magnitude between any two concepts; this approach uses Hellinger Distance (Hoens et al., 2011) and Total Variation Distance for this purpose. Marginal Drift

Magnitude measures the drift by approximate Probability Distribution. Geoffrey I. Webb et al. (2018) uses the weighted averaging approach to deal with Conditional Distribution to tackle multiple distribution problems.

Analysis and deduction

CD is a phenomenon that mostly occurs in Online Learning. The various conditions and types of CD make the existing ML models inappropriate for an online scenario. More specifically, the condition even goes worsens when dealing with multiple CDs at the same time. Furthermore, the available handling approaches are not generic and require separate handling mechanisms for each CD type. The adaptability in the learner is a primary way to overcome performance degradation due to a CD. However, the handling of drift recurrence scenarios is still challenging. Even though CD's quantification is one of the primary factors to detect CD, only a few studies investigated this area. Mostly, the classifier/learner's performance accuracy is considered the most appropriate way to observe the CD.

RQ2: Do the state of art approaches (for CD handling) are adequate for current and future computing trends?

The term handling (of CD) refers to the detection and adaptation process.

Concept Drift (CD) detection

CD detection identifies the real changes in the input stream during online classification. CD detection is a prerequisite for adaptation. However, the process of determining the changes is merely dependent on the type of data stream or the nature of Concept Drift. Hence the provided solution is not generalizing. An early study, Probably Approximately Correct learning model (PAC learning model Kearns and Vazirani, 1994), states that the error rate will always be minimum after expansion in sample size in a static learning environment.

Contrary to this, the error rate significantly escalated after observing the change in class distribution. The CD detection techniques are classified into a type of Machine Learning problem. For example, most of the proposed techniques which are applicable for supervised learning, such as DDM by Gama et al. (2004) and Early Drift Detection Method (EDDM), (Baena-Garcıa et al., 2006) are not applicable for the unsupervised problem (Lavaire et al., 2015).

Unlike the Supervised Drift Detection, possible change is observed in the Unsupervised Drift Detection after statistical hypothesis tests. Also, the classifier does not participate in detecting drift. For example, in a study (Friedman and Rafsky, 1979), Friedman and Rafsky's propose an Unsupervised Drift Detection algorithm. This empirical study focuses on the growth of execution time on datasets with increasing dimensions and comparative accuracy of algorithms concerning their drift detection ability. Few detection models work well for both supervised and unsupervised learning scenarios. For example, ONLINDDA, proposed by Spinosa et al. (2007), uses the integrated set of clusters to identify the newly emerging concept drift scenarios. These clusters are capable of detecting possible changes in both scenarios.

The majority of the studies in the literature detects the Concept Drift using the performance monitoring algorithms (performance measures, properties of the data are monitored over time) (Zeira et al., 2004) and distribution comparing algorithms (Monitoring distributions on two different time-windows. A reference window, that usually summarize past information, and a window over the most recent examples) (Kifer et al., 2004). In a recent study, a concept drift detector is based on computing multiple model explanations over time and observing their changes' magnitudes. The model explanation is calculated using a methodology that yields attribute-value contributions for prediction outcomes, provides insight into the model's decision-making process, and enables transparency. The evaluation has revealed that the methods surpass the baseline methods in terms of concept drift detection, accuracy, robustness, and sensitivity (Demšar and Bosnić, 2018). Many studies discuss the available Concept Drift Detection techniques (Zeira et al., 2004; Kifer et al., 2004; Demšar and Bosnić, 2018). Most of the methods use the Weight Determination Approach, Window Approach, or Statistical Analysis Approach. In 2004, Gama et al. (2004) proposed a novel Drift Detection Method (DDM) framework. DDM is one of the preliminary frameworks which identifies the expected drift employed by a probability distribution and online error-rate. In this approach, Gama et al. mention a specific error-rate threshold for warning and drift levels. The drift observes after the error-rate crosses the warning threshold level (when it observes more than 30 errors). Later, the model starts its training mechanism to tune with detected changes. In 2006, an extension of DDM, the Early Drift Detection Method (EDDM), was proposed by Baena-Garcıa et al. (2006). EDDM technique uses two parameters to detect the drift, 1) many error rate and 2) the difference between two successive errors. The proposed model follows the

inverse relation of concept drift and distance between errors. The authors state that when the new concept arrives, the distance among the errors significantly decreases. EDDM is proven to be a better drift detection approach (especially for gradual drift) than DDM, and it is more delegate to noise than DDM.

A sequential analysis-based technique is Page-Hinkley Test (PHT) (Page, 1954; Mouss et al., 2004). PHT calculates the classifier accuracy to determine the occurrence of Concept Drift. For example, if the classifier degrades its performance accuracy to a specific threshold value, it is considered a drift situation. This approach fundamentally computes the Actual Accuracy and Average Accuracy (up to the current moment). The cumulative difference between Actual Accuracy and Average Accuracy represents "U.T.," and the minimum difference between Actual Accuracy and Average Accuracy is described as "mT." Both "U.T." and "mT" values are computed to determine the drift occurrence. For example, higher "U.T." values indicate that the observed values differ considerably from their previous values. More specifically, the drift observes when the difference between U.T. and mT is above a specified threshold corresponding to the magnitude of allowed (k) changes. An ensemble-based drift detection method was proposed by Kitani et al. Yasumura et al. (2007). This approach determines the drift by comparing the classification accuracy of two ensemble classifiers. In this approach, the AdaBoost (Freund and Schapire, 1997) algorithm is used to determine each ensemble classifier's inverse weight to distinguish between the actual drift and noise. In 2008, Maloof et al. proposed a novel Paired Learner (PL) drift detection approach (Bach and Maloof, 2008). P.L. typically presents two different algorithms to detect the change from the input stream, such as Stable Learner (SL) and Reactive Learner (R.L.). The SL utilizes its historical knowledge for prediction contrary, the R.L. predicts based on a window of recent examples. However, the drift identifies by the computational contribution of both learners and their accuracy.

Bifet et al. propose a dynamic sliding window approach ADWIN (Bifet, 2009). ADWIN typically handles one-dimension data using a single window. However, multiple sliding windows can detect multi-dimension data (each window for each dimension). The window size narrows down when a rate of change perceives from the data in these windows, and an apparent change has been established. This approach dynamically regulates window size to the most appropriate point between reaction time and small variation. An extension of ADWIN, known as ADWIN2, was proposed to overcome the deficiencies in time and memory in ADWIN. The experimental results are better by ADWIN2 (Bifet and Gavalda, 2007); maintaining the same accuracy performance while utilizing less memory and consuming shortens the time. Nishida et al. proposed a statistical-based approach, namely the Statistical Test of Equal Proportions (STEPD) (Nishida, 2008). Like other various studies (Gama et al., 2004; Baena-Garcıa et al., 2006), warning and drift threshold are specified in this approach. This approach classifies the drift and non-drift scenarios based on recent accuracy and overall accuracy (from the beginning) of the classifier. It suggests a concept drift scenario; the current accuracy will always have a significant difference compared to the overall accuracy of the model. In this approach, Nishida et al. performed chi-square tests and computed the acquired value from the standard normal distribution to determine the significance level; a secondary significance level portrays the drift occur acne. Rose et al. proposed a novel approach EWMA (Ross et al., 2012). This approach used an exponentially weighted average moving mechanism and meant a sequence of random variables to detect the changes in the underlying distribution of the input stream. Typically, this mechanism constructs the EWMA chart (to monitor a streaming classifier) to observe the possible drift. EWMA is a modular technique and added an extra layer for observing a drift; this addition of drift detection layer contributes to detecting drift in a parallel execution manner with any underlying classifier. In contrast with other available drift detection techniques, in EWMA, the rate of false-positive detections is controlled and constant over time. Adaptive Learning with Covariant Shift-Detection (ALCSD) is an extension of EWMA. This approach detects the possible shifts using the EWMA shift detection text and covariant shift analysis (Raza et al., 2014). Sobhani et al. also proposed a novel approach to detect the drift using the nearest neighbor approach. The algorithm handles input data chunks by chunks of several batches. All the previous batches with the most immediate neighbor values are computed for every instance in the current batch and compared with their respective labels. A distance map is generated to classify the drift and non-drift batches and instances. More specifically, a drift is observed when the average and standard deviation of all degrees of drift evaluate, and contemporary value (current) are distinct to the average or above than the standard deviation parameter "S."

Concept Drift (CD) adaptation

To find new means to handle CD in the context of BD and OML is an essential task for the future of ML

(Rouse, 2009). Nevertheless, several studies urged to adopt these dynamic changes (in classifier) through self-regulatory mechanisms (Ditzler and Polikar, 2013; Zliobaite et al., 2012). Existing adaptation approaches for CD handling are Shallow Learning and Deep Learning Classifiers or Hybrid using Single and Ensemble approaches.

Shallow Learning, Deep Learning, and Hybrid CD adaptation approaches

Shallow Learning classifiers (for example, Extreme Learning Machine (ELM), Support Vector Machine (SMV), Multi-Layer Perception Neural Network (MLP NN), Hidden Markov Model) handle classification and regression problems efficiently in structured data (Huang, 2006). These approaches do not perform well for complex unstructured data (Big Data) (Jameel et al., 2020a, b, c). However, Deep Learning algorithms such as CNN, Autoencoder perform well in complex and unstructured data streams. DL classifiers extract more detailed value (from Big Data) and yield more accuracy over conventional approaches (Budiman et al., 2017), whereas S.L. approaches are simple and acquire less computation. In the literature, many studies propose Hybrid approaches to handle the CD issue. These Hybrid approaches combine the valuable characteristics of both S.L. and DL approaches. For example, to benefit from the simplicity and fast processing of S.L., and to utilize the more accurate feature extraction mechanism of DL.

Single classifier-based CD adaptation approaches

Single classifier-based approaches make the necessary parameter tuning (within classifiers) for CD adaptation. However, a single classifier approaches face complications to incorporate the forgetting mechanism within the Online Learner. Notably, Extreme Learning Machine (ELM), Support Vector Machine (SVM), and Convolutional Neural Network (CNN) are the famous classifiers used for handling CD. Whereas, ELM is found better due to its simplicity and uncomplicated parameter tuning for new concept adjustment. In literature, many studies proposed different variations in the simple ELM to cope with the CD issue, for example, Online Sequential ELM and Adaptive Online Sequential ELM. In literature, the ELM based models adapt new changes with high accuracy rates. These classifiers used two types of approaches to handle CD: Single classifier and Ensemble classifier (Jameel et al., 2020a, b, c;

Uddin et al., 2019, September). In contrast with the Single classifier, the Ensemble classifier is an effective solution and mostly reported a significant improvement in performance accuracy (after CD) than a single classifier. Nevertheless, the ensemble approach does not adapt to the numerous drift cases (Liu and Wang, 2010; van Schaik and Tapson, 2015); such drift can be handled through classifiers' adaptive nature.

Few recent studies concentrated on adaptive learning techniques using ELM based Single classifiers (van Schaik and Tapson, 2015; Huang et al., 2012; Huang, 2006) and Ensemble classifier for CD adaptation (Zhai et al., 2014; Xu and Wang, 2016; 2017). However, all these solutions lie in the semi-adaptive category (does not implement fully autonomous learning behavior). For example, Incremental Data Stream ELM is used as an incremental approach to train the classifier. In this approach, the number of neurons in hidden layers and the selection of the activation layer are dynamic, enhancing the performance of the model. At the same time, this approach handles stream data for gradual drift scenario only (Zhai et al., 2014).

A Dynamic-ELM model uses ELM as a first classifier, whereas the online learning approach was adopted to train the double hidden layer structure of ELM. The improvement in the generalized characteristics of the classifier was incorporated by adding more hidden layers. This approach is capable of mitigating the CD in a short time. However, the performance of this model suffers due to the fast processing speed (Xu and Wang, 2017). Meta-Cognition Online Sequential Extreme Learning Model (MOSELM) proposed improving class imbalance (binary and multiclass) and Concept Drift for online data classification. This model uses Meta-Cognition principles and Online Sequential Extreme Learning Machine (OSELM) but only handles Real Drift (Liang et al., 2006). A new adaptive windowing approach is proposed to improve adaptability in Real Drift only (Huang et al., 2012). Online Pseudo Inverse Method (OPIUM) is based on Graviel methods, the incremental solutions to computing pseudoinverse of a matrix. OPIUM tackles the real Concept Drift with the discriminant function boundary shift in streaming data only (van Schaik and Tapson, 2015). A recent study proposed an adaptive ML model (AOSELM) (Budiman et al., 2016) using a single classifier approach based on Online Sequential Extreme Learning Machine (OSELM) (Liang et al., 2006) and Constructive Sequential Extreme Learning Machine (COSELM) (Lan et al., 2009) to handle the Concept Drift issue for classification and regression problem.

AOSELM is a simple solution, which used the matrix adjustment technique CD adaptation. Results were satisfactory for handling Real Drift but not satisfactory to handle virtual and Hybrid Drift and did not yield better output on real data. In 2019, a study (Liu et al., 2019) on a novel approach Meta-cognitive Recurrent Recursive Kernel Online Sequential Extreme Learning Machine with Drift Detector Mechanism (meta-RRKOS-ELM-DDM) was proposed. This proposed approach utilizes the Recurrent Kernel Online Sequential Extreme Learning Machine with an enhanced Drift Detector Mechanism (DDM) and Approximate Linear Dependency Kernel Filter (ALD). This approach found it better to handle Concept Drift with less complex computation. In 2015, Cao et al. Krawczyk (2015) proposed an adaptive model of the Weighted One-Class Support Vector Machine. Due to incremental learning and forgetting strategy. This model smoothly adapts the new changes with the intervention of the drift detection module.

Ensembles classifiers based CD adaptation approaches

In literature, most of the ensemble classifier approaches are found better than the single classifier approaches. In an ensemble approach, several individual instances (classifiers) participate in making a final decision. The decisions of each instance are aggregated by several approaches to predict a final decision. These approaches include Max Voting, Averaging, Stacking, Blending, Bagging, and Boosting. Whereas, modularity feature of ensembling makes it more feasible to adapt to any new concept during online learning. For example, a study (Cao et al., 2015; Khamassi et al., 2015) proposed the ELM based Weighted Ensemble Classifier to adjust the classifier after observing the concept drift issue dynamically. Block-Based Ensemble Approach (Brzezinski and Stefanowski, 2012) and Weighting Data Ensemble Approach (Sidhu and Bhatia, 2018) are the two most effective available approaches. These techniques are more appropriate to handle Simple drift in a better way. However, the Complex drifts may present a mixture of several critical characteristics such as speed, severity, influence zones in the feature space, which may vary over time (Khamassi et al., 2019). Furthermore, in the literature, the adaptivity through ensemble is achieved by;

Horse racing approach: The forgetting approach is used to train the component classifiers; mainly, this forgetting approach trains on a different combination. For example, two bagging approaches, ADWIN, and ASHT are evidence of the better performance of ensemble approaches for CD adaptation (Bifet et al., 2009). These methods utilize the tree of various sizes and dynamically expand to new concept adjustment using the forgetting mechanism. However, due to the different ensemble-based tree structures, both techniques are memory and time expensive.

Training update approach: Instances of classifier increase incrementally to train on the newly arrived concept. For example, Accuracy Weigh Ensemble (AWE) (Wang et al., 2003) having a notable contribution to adjusting the recurring drift. However, its performance compared to other online learners is not satisfactory, such as Accuracy Update Ensemble AUE. In the AWE approach, a new instance of the ensemble adds and trains after each input data block's arrival. Then, this data block uses to evaluate the performance of some other cases in the ensemble. The instance of the ensemble with a higher accuracy rate is selected for classification. The size of the ensemble is also crucial to handle. The instances of the ensemble with less accuracy are removed to manage the size of the ensemble.

Instance update approach: The existing classifier is retrained from the new concept. In the instance update approach, it is critical to handle recurrent concept adjustment. For example, Accuracy Update Ensemble (AUE) (Brzeziński and Stefanowski, 2011) is a better approach than AWE. Unlike the AWE, AUE updates conditionally update its ensemble instances using the weighted voting rule. Despite deleting the weak classifier and add a new classifier for a new data block, it updates the weak classifier with a new weight and current distribution. Another extension of AUE is OAUE, which combines block-based ensembles and online processing with improving time and memory (Brzezinski and Stefanowski, 2014).

Structure update approach: The less accurate and old classifier is retrained with the new concept. Streaming Ensemble Algorithm (Street and Kim, 2001) dynamically changes its structure as per the new concept change. It is a heuristic replacement strategy of the weakest base classifier based on accuracy and diversity. The combined decision is based on simple majority voting and classifiers (base) unpruned. This algorithm works best for at most 25 components of the ensemble.

Feature update approach: This approach identifies the most appropriate features for the classifier performance. These features are dynamically selected based on current features' significance, without redesigning the ensemble structure (Kuncheva, 2004).

Analysis and deduction

The comprehensive literature analysis deduces that the existing CD handling (detection and adaptation) approaches are classified based on their behavior and structure. The current concept drift detector can be classified into three major categories;

1. Detection Concept Drifts by Data Distribution.
2. Detecting Concept Drift by Learner Outputs.
3. Detecting Concept Drift by Parameters.

However, few studies utilized the weight, window approach, ensemble approach, and statistical methods to determine the possible change from the input stream among these categories. Furthermore, most of the studies define a threshold value for warning a drift and actual drift. However, all these solutions possess a common problem that they have overlooked noise as concept drift, and there is no particular way to distinguish the noise and potential concept drift. Also, there is not a single generalized adaptation approach, which applies to all types of CD. Moreover, the classification degradation does not reasonably retain after CD handling for complex datasets (such as Imagery Streams) and complex CD scenarios (recurrence scenarios). In literature, the ensemble way of handling CD issues is more appropriate. However, it requires further online training options to avoid any manual intervention for CD adaptation (which is desirable for future analysis trends). The ensemble classifier approach ensures the CD adaptation due to its diversity feature to adopt new changes. In comparison, single classifier results may not exceed the ensemble due to its shared weight changes.

Conclusion

This Systematic Literature Review (SLR) investigates two (2) basic research questions relevant to Concept Drift (CD) phenomena. Initially, the first research question discusses the three primary types of CD, such as virtual concept drift (VCD), real concept drift (RCD), and hybrid concept drift (HCD). Handling VCD is less complicated than RCD, and HCD is still a challenge to be resolved. These types are also different categories due to their transition frequency patterns, such as sudden, gradual drift, continuous, incremental, and blip patterns. However, several studies do not consider blip pattern as CD due to its less significance to the overall model performance. Whereas, some studies argue not to overlook any precision of change during analysis. Besides, the problem of CD recurrence requires a more sophisticated mechanism to adapt to the new changes. CD's issue is addressed in the existing literature through several approaches, such as the static model, continuous refit approach, continuous updating approach, weight data approach, ensemble approach, and dynamic selection approach. The majority of provided approaches are based on the ensemble method. Measuring the CD using the quantitative approach is desirable; however, it is mostly detected through qualitative measurements. A few studies figured out the quantitative measurements of CD using distance and magnitude measures, which are not applicable for problematic concept drift. The second (2) research question investigated the existing CD handling techniques and determined their current and future computing applicability. This question concludes that existing CD handling approaches yet to be matured to handle online learning in the present scenario and required more robust dynamic adaptive approaches. Currently, most CD detection methods observe the CD using data distribution, classifier output, or weight parameter and cannot correctly differentiate between the noise and the original CD. Similarly, provided solutions either apply for a specific CD or do not effectively work for complicated CD types, or limited to handle particular data. Since CD adaptation cannot be generalized due to the change of the data stream's nature, a uniform approach is not applicable to handle all types of CD.

Acknowledgments

Literature Cited

Bach, S. H. and Maloof, M. A. 2008. "Paired learners for concept drift." Eighth IEEE International Conference on Data Mining. IEEE.

Baena-García, M., del Campo-Ávila, J., Fidalgo, R., Bifet, A., Gavalda, R. and Morales-Bueno, R. 2006. "Early drift detection method". *Fourth International Workshop on Knowledge Discovery from Data Streams* 6: 77–86.

Bifet, A. 2009. "Adaptive Learning and Mining for Data Streams and Frequent Patterns", Doctoral Thesis.

Bifet, A. and Gavalda, R. 2007. "Learning from time-changing data with adaptive windowing." Proceedings of the 2007 SIAM international conference on data mining. Society for Industrial and Applied Mathematics.

Bifet, A., et al. 2009. "New ensemble methods for evolving data streams." Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM.

Brzeziński, D. and Stefanowski, J. 2011. *"Accuracy Updated Ensemble for Data Streams with Concept Drift." International Conference on Hybrid Artificial Intelligence Systems* Springer, Berlin and Heidelberg.

Brzezinski, D. and Stefanowski, J. 2012. "From block-based ensembles to online learners in changing data streams: If-and how-to." Proceedings of the 2012 ECML PKDD Workshop on Instant Interactive Data Mining, Available at: http://adrem.ua.ac.be/iid2012.

Brzezinski, D. and Stefanowski, J. 2014a. Reacting to different types of concept drift: The accuracy updated ensemble algorithm. *Neural Networks and Learning Systems, IEEE Transactions on* 25(1): 81–94, doi: 10.1109/TNNLS.2013.2251352.

Brzezinski, D. and Stefanowski, J. 2014b. "Combining block-based and online methods in learning ensembles from concept drifting data streams". *An International Journal: Information Sciences* 265: 50–67.

Budiman, A., Fanany, M. I. and Basaruddin, C. 2016. Adaptive Online Sequential ELM for Concept Drift Tackling. *Computational Intelligence and Neuroscience* 2016(20): 17, Available at: https://doi.org/10.1155/2016/8091267.

Budiman, A., Fanany, M. I. and Basaruddin, C. 2017. Adaptive Parallel ELM with Convolutional Features for Big Stream Data. Thesis Dissertation, Faculty of Computer Science, University of Indonesia, doi: 10.13140/RG.2.2.18500.22404.

Cao, K., Wang, G., Han, D., Ning, J. and Zhang, X. 2015. Classification of Uncertain Data Streams Based on Extreme Learning Machine. *Cognitive Computation* 7(1): 150–160.

Dariusz, B. 2010. Mining data streams with concept drift. Master's thesis, Poznan University of Technology.

Demšar, J. and Bosnić, Z. 2018. Detecting concept drift in data streams using model explanation. *Expert Systems with Applications* 92: 546–559.

Ditzler, G. and Polikar, R. 2013. Incremental learning of Concept Drift from Streaming Imbalanced Data. *IEEE Trans. Knowledge Data Engineering* 25(10): 2283–2301.

Dongre, P. B. and Malik, L. G. 2014. A review on real time data stream classification and adapting to various concept drift scenarios. In Advance Computing Conference (IACC), 2014 IEEE International, February, pp. 533–537, doi: 10.1109/IAdCC.2014.6779381.

Dyer, K. B. and Polikar, R. 2012. "Semi-supervised learning in initially labeled nonstationary environments with gradual drift." The International Joint Conference on Neural Networks (IJCNN). IEEE.

Freund, Y. and Schapire, R. E. 1997. A decision-theoretic generalization of online learning and an application to boosting. *Journal of Computer and System Sciences* 55(1): 119–139.

Friedman, J. H. and Rafsky, L. C. 1979. "Multivariate generalizations of the wald-wolfowitz and smirnov two-sample tests". *Institute of Mathematical Statistics*, 7(4): 697–717, doi: 10.1214/aos/1176344722.

Gama, J., Medas, P., Castillo, G. and Rodrigues, P. 2004. Learning with drift detection. *In Advances in Artificial Intelligence–SBIA*, Springer Berlin and Heidelberg, pp. 286–295.

Ghorbani, S., Barari, M. and Hosseini, M. 2017. "A modern method to improve of detecting and categorizing mechanism for micro seismic events data using boost learning system". *Civil Engineering Journal* 3(9): 715–726.

Gomes, J. B., Menasalvas, E. and Sousa, P. A. C. 2011. "Learning recurring concepts from data streams with a context-aware ensemble", *Proceedings of the 2011 ACM Symposium on Applied Computing, SAC '11* ACM, New York, NY, pp. 994–999, doi: 10.1145/1982185.1982403.

Gupta, B. M. and Dhawan, S. M. 2019. Deep Learning Research: Scientometric Assessment of Global Publications Output during 2004-17. *Emerging Science Journal* 3(1): 23–32.

Harel, M., et al. 2014. Concept drift detection through resampling. International Conference on Machine Learning.

Hoens, T. R., Chawla, N. V. and Polikar, R. 2011. "Heuristic updatable weighted random subspaces for nonstationary environments", In Cook, D. J., Pei, J. W., Wei, Z., Osmar, R. and Wu, X. (Eds), *IEEE International Conference on Data Mining, ICDM-11*, IEEE, pp. 241–250.

Hoens, T. R., Polikar, R. and Chawla, N. V. 2012. Learning from streaming data with concept drift and imbalance: an overview. *Progress in Artificial Intelligence* 1(1): 89–101, doi: 10.1007/s13748-011-0008-0.

Huang, D. T. J., Koh, Y. S., Dobbie, G. and Pears, R. 2013. "Tracking drift types in changing data streams", In Hiroshi, M., Wu, Z., Cao, L., Zaiane, O., Yao, M. and Wang, W. (Eds), *Advanced Data Mining and Applications, volume 8346 of Lecture Notes in Computer Science*, Springer, Berlin and Heidelberg, pp. 72–83, doi: 10.1007/978-3-642-53914-57.

Huang, G. B. 2006. Extreme Learning Machine. Theory and Applications. *Neuro Computing* 70(1–3): 489–501.

Huang, G. B., Zhou, H., Ding, X. and Zhang, R. 2012. Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Transactions on Systems, Man, and Cybernetics* 42(2): 513–529.

Iwashita, A., Sayuri and Papa, J. P. 2019. "An Overview on Concept Drift Learning". *IEEE Access* 7: 1532–1547.

Jagadeesh Chandra Bose, R. P., van der Aalst, W. M. P., Zliobaite, I. and Pechenizkiy, M. 2011. "Handling concept drift in process mining", In Haralambos, M. and Colette, R. (Eds), *Advanced Information Systems Engineering, volume 6741 of Lecture Notes in Computer Science*, Springer, Berlin and Heidelberg, pp. 391–405, doi: 10.1007/978-3-642-21640-430.

Jameel, S. M., et al. 2018. *"A Fully Adaptive Image Classification Approach for Industrial Revolution 4.0." International Conference of Reliable Information and Communication Technology* Springer, Cham.

Jameel, S. M., Hashmani, M. A., Rehman, M. and Budiman, A. 2020a. An Adaptive Deep Learning Framework for Dynamic Image Classification in the Internet of Things Environment. *Sensors* 20(20): 5811, doi: 10.3390/s20205811.

Jameel, S. M., Hashmani, M. A., Rehman, M. and Budiman, A. 2020b. Adaptive CNN Ensemble for Complex Multispectral Image Analysis. *Complexity* 2020: 21, Available at: https://doi.org/10.1155/2020/8361989.

Jameel, S. M., Hashmani, M. A., Alhussain, H., Rehman, M. and Budiman, A. 2020c. "A Critical Review on Adverse Effects of Concept Drift over Machine Learning Classification Models". *International Journal of Advanced Computer Science and Applications (IJACSA)* 11(1): 2020, Available at: http://dx.doi.org/10.14569/IJACSA.2020.0110127.

Jensen, C., et al. 2019. "Piloting a Methodology for Sustainability Education: Project Examples and Exploratory Action Research Highlights". *Emerging Science Journal* 3(5): 312–326.

Kearns and Vazirani. 1994. PAC learning model.

Khamassi, I., Sayed-Mouchaweh, M. and Hammami, M. 2015. Self-Adaptive Windowing Approach for Handling Complex Concept Drift. *Cognitive Computing* 7(6): 772–790.

Khamassi, I., et al., 2019. *"A New Combination of Diversity Techniques in Ensemble Classifiers for Handling Complex Concept Drift". Learning from Data Streams in Evolving Environments* Springer, Cham, pp. 39–61.

Kifer, D., Ben-David, S. and Gehrke, J. 2004. Detecting change in data streams. In Proceedings of the International Conference on Very Large Data Bases, Toronto, Canada, Morgan Kaufmann, pp. 180–191.

Kitchenham, B. 2004. "Procedures for performing systematic reviews," Department of Computer Science, Keele University, ST5 5BG, U.K., Tech. Rep. TR/SE-0401.

Kitchenham, B. A. and Charters, S. 2007. Guidelines for performing systematic literature reviews in software engineering, Tech. Rep. EBSE-2007-01, Keele University and University of Durham.

Krawczyk, B. 2015. Reacting to Different Types of Concept Drift One Class Classifiers. 2nd International Conference on Cybernetics, IEEE, Gdynia, Poland, pp. 30–35.

Kuncheva, L. I. 2004. "Classifier Ensembles for Changing Environments", In Roli, F., Kittler, J. and Windeatt, T. (Eds), *Multiple Classifier Systems. MCS. LNCS* 3077, Springer, Berlin and Heidelberg, pp. 1–15.

Lan, Y., Soh, Y. C. and Huang, G. 2009. "A constructive enhancement for Online Sequential Extreme Learning Machine," 2009 International Joint Conference on Neural Networks, Atlanta, GA, pp. 1708–1713, doi: 10.1109/IJCNN.2009.5178608.

Lavaire, J. D. D., et al. 2015. "Dimensional scalability of supervised and unsupervised concept drift detection: An empirical study." 2015 IEEE International Conference on Big Data (Big Data). IEEE.

Liang, N., Huang, G., Saratchandran, P. and Sundararajan, N. 2006. A Fast and Accurate Online Sequential Learning Algorithm for Feedforward Networks. *IEEE Transactions Neural Networks* 17(6): 1411–1423.

Liu, N. and Wang, H. 2010. Ensemble based Extreme Learning Machine. IEEE. *Signal Process* 17(8): 754–757.

Liu, Z., Loo, C. K. and Seera, M. 2019. "Meta-cognitive Recurrent Recursive Kernel OS-ELM for concept drift handling". *Applied Soft Computing* 75: 494–507.

Mehta, S. 2017. Concept drift in Streaming Data Classification: Algorithms, Platforms, and Issues. *Procedia computer science* 122: 804–811.

Minku, L. L., White, A. P. and Yao, X. May 2010. The impact of diversity on online ensemble learning in the presence of concept drift. *IEEE Transactions on Knowledge and Data Engineering* 22(5): 730–742, doi: 10.1109/TKDE.2009.156.

Mouss, H., Mouss, D., Mouss, N. and Sefouhi, L. 2004. Test of Page-Hinkley, an Approach for Fault Detection in an Agro-Alimentary Production System. *Proceedings of the 5th Asian Control Conference* 2: 815–818.

Nishida, K. 2008. "Learning and Detecting Concept Drift", A Dissertation: Doctor of Philosophy in Information Science and Technology, Graduate School of Information Science and Technology, Hokkaido University].

Nishida, K., et al. 2008. "Detecting sudden concept drift with knowledge of human behavior." 2008 IEEE International Conference on Systems, Man and Cybernetics. IEEE.

Page, E. S. 1954. Continuous Inspection Schemes. *Biometrika* 41: 100–115.

Petersen, K., Feldt, R., Mujtaba, S. and Mattsson, M. 2008. "Systematic mapping studies in software engineering," in Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering (EASE 2008).

Pfleeger, S. L. 2005. Soup or art? The role of evidential force in empirical software engineering. *IEEE Software* 22(1): 66–73.

Raza, H., Prasad, G. and Li, Y. 2014. "Adaptive learning with covariate shift-detection for nonstationary environments." 2014 14th U.K. Workshop on Computational Intelligence (UKCI). IEEE.

Ross, G. J., et al. 2012. "Exponentially weighted moving average charts for detecting concept drift". *Pattern recognition letters* 33(2): 191–198.

Rouse, M. 2009. Predictive Analytics Definition.

Saurav, S., et al. 2018. "Online anomaly detection with concept drift adaptation using recurrent neural networks." Proceedings of the ACM India Joint International Conference on Data Science and Management of Data. ACM.

Sayed, S., Ansari, S. A. and Poonia, R. 2018. "Overview of Concept Drifts Detection Methodology in Data Stream" Handbook of Research on Pattern Engineering System Development for Big Data Analytics. *IGI Global*, pp. 310–317, doi: 10.4018/978-1-5225-3870-7.ch018.

Schaik, A. and van. Tapson, J. 2015. Online and Adaptive Pseudoinverse Solutions for ELM Weights. *Neurocomputing* 149(A): 233–238.

Sidhu, P. and Bhatia, M. P. S. 2018. "A novel online ensemble approach to handle concept drifting data streams: diversified dynamic weighted majority". *International Journal of Machine Learning and Cybernetics* 9(1): 37–61.

Spinosa, E. J., de Carvalho, A. P. de L. F. and Gama, J. 2007. "Olindda: A cluster-based approach for detecting novelty and concept drift in data streams." Proceedings of the 2007 ACM symposium on Applied computing. ACM.

Street, W. N. and Kim, Y. 2001. "A streaming ensemble algorithm (SEA) for large-scale classification," in Proc. 7th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 377–382.

Tsymbal, A. 2004. The problem of concept drift: definitions and related work. Technical Report TCD-CS-2004-15, The University of Dublin, Trinity College, Department of Computer Science, Dublin, Ireland.

Uddin, V., Rizvi, S. S. H., Hashmani, M. A., Jameel, S. M. and Ansari, T. 2019. September. *A Study of Deterioration in Classification Models in Real-Time Big Data Environment. In International Conference of Reliable Information and Communication Technology*, Springer, Cham, pp. 79–87.

Wadewale, K. and Desai., S. 2015. "Survey on method of drift detection and classification for time varying data set". *International Journal of Research in Engineering and Technology* 2(9): 709–713.

Wang, H., Fan, W., Yu, P. S. and Han, J. 2003. "Mining concept-drifting data streams using ensemble classifiers", In Getoor, L., Senator, T. E., Domingos, P. and Faloutsos, C. (Eds), *Association for Computing Machinery*, ACM Press, New York, NY, pp. 226–235.

Webb, G. I., et al. 2016. "Characterizing concept drift". *Data Mining and Knowledge Discovery* 30(4): 964–994.

Webb, G. I., et al. 2018. "Analyzing concept drift and shift from sample data". *Data Mining and Knowledge Discovery* 32(5): 1179–1199.

Xu, S. and Wang, J. 2016. A Fast-Incremental Extreme Learning Machine Algorithm for Data Streams Classification. *Expert Systems with Applications* 65: 332–344.

Xu, S. and Wang, J. 2017. Dynamic Extreme Learning Machine for Stream Classification. *Neurocomputing* 238(A): 433–449.

Yasumura, Y., Kitani, N. and Uehara, K. 2007. *"Quick Adaptation to Changing Concepts by Sensitive Detection." International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems* Springer, Berlin and Heidelberg.

Zang, W., Zhang, P., Zhou, C. and Guo, L. 2014. Comparative Study Between Incremental and Ensemble Learning on Data Stream: Case Study. *Journal of Big Data* 1(1): 1–16.

Zeira, G., Maimon, O., Last, M. and Rokach, L. 2004. "Data mining in time series databases", In Last, M., Kandel, A. and Bunke, H. (Eds), *Data Mining in Time Series Databases, Volume 57, Chapter Change Detection in Classification Models Induced from Time-Series Data*, World Scientific, Singapore, pp. 101–125, Available at: https://www.worldscientific.com/page/about/corporate-profile.

Zhai, J., Wang, J. and Wang, X. 2014. "Ensemble Online Sequential Extreme Learning Machine for Large Dataset Classification", *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, San Diego, CA, pp. 2250–2255, doi: 10.1109/SMC.2014.6974260.

Zliobaite, I. 2010. Learning under Concept Drift: an Overview. *Cornell University Library*, pp. 1–36, doi: arxiv.org/abs/1010.4784.

Zliobaite, I., Bifet, A., Pechenizkiy, M. and Bouchachia, A. 2014. A Survey on Concept Drift Adaptation. *ACM Computer Survey* 46(4): 1–37.

Zliobaite, I., et al., 2012. Next Challenges for Adaptive Learning Systems. *ACM SIGKDD Explorations Newsletter* 14(1): 48.