A Novel Approach to Maximize G-mean in Nonstationary Data with Recurrent Imbalance Shifts

Radhika Kulkarni[1], S. Revathy[1], and Suhas Patil[2]

[1]Department of Computer Science Engineering, Sathyabama Institute of Science and Technology, India

[2]Department of Computer Science Engineering, Bharati Vidyapeeth's College of Engineering, India

Abstract: *One of the noteworthy difficulties in the classification of nonstationary data is handling data with class imbalance. Imbalanced data possess the characteristics of having a lot of samples of one class than the other. It, thusly, results in the biased accuracy of a classifier in favour of a majority class. Streaming data may have inherent imbalance resulting from the nature of dataspace or extrinsic imbalance due to its nonstationary environment. In streaming data, timely varying class priors may lead to a shift in imbalance ratio. The researchers have contemplated ensemble learning, online learning, issue of class imbalance and cost-sensitive algorithms autonomously. They have scarcely ever tended to every one of these issues mutually to deal with imbalance shift in nonstationary data. This correspondence shows a novel methodology joining these perspectives to augment G-mean in no stationary data with Recurrent Imbalance Shifts (RIS). This research modifies the state-of-the-art boosting algorithms,1) AdaC2 to get G-mean based Online AdaC2 for Recurrent Imbalance Shifts (GOA-RIS) and AGOA-RIS (Ageing and G-mean based Online AdaC2 for Recurrent Imbalance Shifts), and 2) CSB2 to get G-mean based Online CSB2 for Recurrent Imbalance Shifts (GOC-RIS) and Ageing and G-mean based Online CSB2 for Recurrent Imbalance Shifts (AGOC-RIS). The study has empirically and statistically analysed the performances of the proposed algorithms and Online AdaC2 (OA) and Online CSB2 (OC) algorithms using benchmark datasets. The test outcomes demonstrate that the proposed algorithms globally beat the performances of OA and OC.*

Keywords: *Cost-sensitive algorithms, data stream classification, imbalanced data, online learning, population shift, skewed data stream.*

1. Introduction

A data stream is a boundless sequence of real-time data instances with high arrival rate [11]. Predictive models for data stream classification have intense demand in numerous applications like spam separation [5], web monitoring [1], medical diagnosis [27], fraud detection [30] and prediction of financial distress [32]. In such real-world applications often, data may have skewness where the number of examples of one class is very few when contrasted with the number of instances of other [35]. The class with a majority of samples influences the performance of the classifier causing negligence to the minority class [15]. The Imbalance Ratio (IR) is a proportion of the number of majority samples to the number of minority samples [25].

The broad categorization of class imbalance is of two types: intrinsic imbalance and extrinsic imbalance [17]. The inherent nature of sample space causes the intrinsic imbalance. Consider a sample space for fraud detection that has a majority of legitimate transactions and a minority of fraudulent transactions. Some external factors like time, memory capacity result in skewed data in the nonstationary streaming environment. This form of imbalance is categorized as the extrinsic imbalance. Suppose for a specific time interval if data from some sensors is interrupted then it may result in skewed data during that time interval. In the next time interval, smooth transmission from all sensors may resume leading to the balanced data stream. Thus, nonstationary data may have variations in IR. In dynamic data streams its distribution of data changes. The population of classes may vary causing the change in class priors [19]. The most researches focus on the intrinsic imbalance in stationary data [10, 17].

The organization of the remaining contents is as follows. Section 2 overviews the related work. Section 3 explains the background of this research. Section 4 provides a thorough description of the proposed methodology. Section 5 highlights the results of our experiments. Section 6 reports the conclusions and mentions a few points for future work.

2. Related Work

Noting the importance of skewed data problem, many researchers have been managing it by three general classes of techniques [10, 17, 39]:

1. Data level (external)
2. Algorithm level (internal)
3. Cost-sensitive techniques.

Data level strategies pre-process data to balance the uneven class distribution through resampling [2, 22, 23, 28, 43]. Algorithm level procedures develop or change the existing algorithms to focus on examples of positive class [12, 32, 40]. The cost-sensitive approaches use a combination of both data and algorithm level methods [24, 30, 39]. Though cost-sensitive approaches are less popular than re-sampling based data level approaches they are computationally more efficient than re-sampling. Hence in streaming data cost-sensitive approach may be more suitable [15]. This communication presents a novel cost-sensitive approach to deal with skewed data streams.

The ensemble of learners enhances the performance of a single learner. It binds a few classifiers to get a novel classifier that beats all of it [10]. Being one of the topmost ten data mining algorithms [41], AdaBoost ensemble learning [9] has caught the eye of data scientists to build up an assortment of cost-sensitive boosting algorithms [20, 21]. Several such blends are having different viewpoints, basic assumptions, and theories [24]. The most illustrative ensembles of this family are AdaCost [7], AdaC1, AdaC2, and AdaC3 [33], AdaCost (β2), CSB0, CSB1, CSB2 [34], and RareBoost [18]. These algorithms contrast in their weight update rule. The comparative study of various cost-sensitive boosting algorithms with different imbalance levels is available in [42]. The research work in [39] introduces Cost-Sensitive Deep Neural Network Ensemble (CSDE) that applies undersampling to deal with the class imbalance in large stationary data. All these studies consider inherent imbalance in data where the positive class is always in a minority. None of these cost-sensitive boosting techniques incorporates online processing and extrinsic imbalance in nonstationary streaming data.

Adaptive, online processing is essential for dynamic environments to incorporate new incoming data [11]. Online learning can be a situation of incremental learning with a single element in each batch [37]. It achieves fast adaptation to the changing environment as it does not wait for the arrival of a full chunk to update the learning model [37]. The research in [26] describes an online boosting framework using the concept of approximation of the binomial distribution by Poisson distribution when $n \to \infty$, where n is the number of trials. The correct classification of an instance results in a decrease in the Poisson distribution parameter λ associated with it when passed to the next learner and misclassification results in an increase in λ [26]. The study in [35] presents online versions of some state-of-the-art ensemble algorithms to tackle the skewness in data and their performance evaluation by using 5-fold cross-validation. The oversampling and undersampling

based online bagging methods are described in [37, 38]. The work in [28] proposes selection-based resampling ensemble and that in [32] presents time-weighted oversampling by combining Synthetic Minority Oversampling Technique (SMOTE) with AdaBoost-SVM to deal with the intrinsic imbalance in dynamic data streams. Both the algorithms [28, 32] process the data stream in batches. Adaptive Random Forest with weighted resampling [8] handles skewness of nonstationary data stream. Very few studies are based on online cost-sensitive algorithms [14, 36] to classify data streams with intrinsic imbalance.

In streaming data, the minority class at a certain time interval may turn to the majority class at the next time interval due to the nonstationary environment. This leads to shifts in class imbalance. Many studies are available on class imbalance problem [10, 15, 17, 20, 21, 24, 42]. However, there is hardly any work that addresses extrinsic imbalance problem in dynamic data, online ensemble method and adaptive cost-sensitive learning altogether. In the case of extrinsic imbalance, both the classes may face a minority problem at different time slots and a classification algorithm needs to adapt to the changes in both class priors. This study presents a joint solution to all these issues. The proposed research modifies the state-of-the-art cost-sensitive boosting algorithms AdaC2 [33] and CSB2 [34] to get four variants namely:

1. GOA-RIS
2. AGOA-RIS
3. GOC-RIS
4. AGOC-RIS.

The proposed research aims at following novelty aspects:

- Online adaptive cost-sensitive boosting algorithms to improve G-mean in RIS.
- G-mean based weighted costs to deal with changing extrinsic or intrinsic imbalance in dynamic data.
- Ageing based approach to incorporate the latest change in class priors in incoming data.
- Empirical and statistical tests to compare G-mean of the proposed algorithms with algorithms AdaC2 and CSB2 on benchmark datasets.

The experimentation results mentioned in this study support the distinguished performance of the proposed novel approach than the state-of-the-art algorithms.

3. Background

The current communication deals with the problem of frequently changing IR due to the streaming nature of nonstationary data. It addresses both the extrinsic and intrinsic imbalance. It refers to online versions of cost-sensitive boosting algorithms AdaC2 and CSB2. It modifies these algorithms to have improved G-mean when data experience RIS.

3.1. Problem Formulation

Consider a data stream $DS=\{(x_t, y_t)\}$ arrived at time step $t = 1, 2\ldots$ and x_t is a data sample in m-dimensional feature space with class label $y_t \in Y = \{y_1, y_2\ldots, y_L\}$; L is the number of class labels. This research assumes a binary class imbalance problem where $Y = \{0, 1\}$. Let DS_t^+ be the set of positive samples with label '1' and DS_t^- be the set of negative samples with label '0' received at time step t such that $DS_t^+ \cup DS_t^- = DS$ and $DS_t^+ \cap DS_t^- = \Phi$. Class imbalance occurs in a data stream at time step t when $|DS_t^+| << |DS_t^-|$, i.e., positive samples are in the minority, and negative samples are in the majority at certain time step t.

As data streams are dynamically changing, they may possess extrinsic imbalance. The changes in class priors at next time step t' may cause imbalance due to $|DS_{t'}^-| << |DS_{t'}^+|$. Over the period, the minority of one class for a certain time step t, $(|DS_t^+| << |DS_t^-|)$ may transform into the majority of the same class for some other time step t', $(|DS_{t'}^-| << |DS_{t'}^+|)$. This type of population shift may lead to RIS in nonstationary data.

3.2. G-mean as an Evaluation Metric in Skewed Domain

The standardized assessment criterion is paramount in the evaluation of a variety of research work in the domain of skewed data. The confusion matrix illustrated in Table 1 describes the binary classification result. Let '0' represents a negative class label and '1' represents a positive class label. Accuracy $((TP+TN)/(TP+TN+FP+FN))$ is less suitable evaluation metric in imbalance domain as poor performance of a learner on minority samples may get obscured by better performance of the learner on majority samples [17]. So, there are some other favourite metrics like sensitivity $(TP/(TP+FN))$, specificity $(TN/(TN+FP))$, precision $(TP/(TP+FP))$, G-mean $\left(\sqrt{Sensitivity \cdot Specificity}\right)$ [17]. As described in section 3.1 when nonstationary data undergo frequent shifts in class imbalance, the performance of both the classes is important. In such cases sensitivity (true positive rate) and specificity (true negative rate) equally, need to be high. Since G-mean measures geometric mean of sensitivity and specificity, it is the appropriate metric for measuring performance in RIS.

Table 1. Confusion matrix for binary classification.

	Actual Class Label '0'	Actual Class Label '1'
Predicted Class Label '0'	TN(True Negative)	FN(False Negative)
Predicted Class Label '1'	FP (False Positive)	TP(True Positive)

3.3. AdaC2 and CSB2

Since the empirical study in [33] claims that AdaC2 outperforms other algorithms of the same family, the current study selects it as one of the state-of-the-art cost-sensitive boosting algorithms. AdaC2 boosts false negative samples more than false positive samples. It reduces lesser weights of true positive samples than true negative samples. Another state-of-the-art algorithm referred in this work is CSB2 [34] that blends AdaBoost and AdaC2. In the case of misclassification, it follows the weight update rule of AdaC2 and that of AdaBoost, otherwise. Both AdaC2 and CSB2 modify the weight update equations to integrate dissimilar costs for misclassification of data belonging to different classes. These algorithms handle an intrinsic imbalance in stationary data using batch processing. Summary of weight update rule and the weight parameter of base learners are given in Table 2.

Table 2. Summary AdaC2 and CSB2 w.r.t. weight update rule and the weight parameter of base learners.

	AdaC2 [33]	CSB2 [34]
Weight update rule	$D_{t+1}^i = \dfrac{C_i \cdot D_t^i \cdot e^{-\alpha_t h_t(x_i) y_i}}{Z_t}$ where Z_t is a normalization factor	$D_{t+1}^i = \dfrac{\gamma_t^i \cdot D_t^i \cdot e^{-\alpha_t h_t(x_i) y_i}}{Z_t}$ where Z_t is a normalization factor and $\gamma_t^i = \begin{cases} C_i\;; if\; y_i \neq h_t(x_i) \\ 1\;; if\; y_i = h_t(x_i) \end{cases}$
Weight parameter	$\alpha_t = \dfrac{1}{2}\log\dfrac{\sum_{i:y_i = h_t(x_i)} C_i \cdot D_t^i}{\sum_{i:y_i \neq h_t(x_i)} C_i \cdot D_t^i}$	$\alpha_t = \dfrac{\sum_{i:y_i = h_t(x_i)}(x_i) D_t^i}{\sum_{i:y_i \neq h_t(x_i)}(x_i) D_t^i}$

3.4. Online AdaC2(OA) and Online CSB2(OC)

In the streaming environment, the whole training dataset is not available at the start; instead, the current work follows the test-then-train approach for each incoming data sample. To have a fair comparison with the state-of-the-art algorithms the current experimentation refers to online versions of AdaC2 and CSB2. Due to unavailability of whole data set in the online environment, tracing of normalization factor becomes impractical. The weight update rules without normalization factors for online processing of AdaC2 and CSB2 are mentioned in [35].

Assume (x_t, y_t) be the tth data instance available for training of learner l. Cost of a False Positive (*CFP)* and Cost of a False Negative *(CFN)* are the costs associated with the negative and positive samples, respectively. $\hat{y}_t$ is the predicted class of a data instance x_t by a learner l. As proposed in [35], Equations (1) and (2) mentioned below respectively define weight update rule for Online AdaC2 (OA) and Online CSB2 (OC). OA traces weighted accuracy and weighted error while OC traces both unweighted and weighted errors. In online boosting [26], the Poisson parameter λ is needed to be updated. The presented algorithms calculate weighted λ parameters for true positive (W_l^{TP}), true negative (W_l^{TN}), false positive (W_l^{FP}) and false negative (W_l^{FN}) referring to lth learner such that

Equation (3) gives their weighted total (W_l^{Total}). Equations (4) and (5) respectively formulate weighted accuracy (W_l^{acc}) and weighted error (W_l^{err}) of the learner *l*. Equation (6) defines unweighted error (ε_l) where (W_l^{MS}) represents the weighted λ parameter for misclassified samples by the *l*th learner.

$$D_{l+1}^t = D_l^t \times \begin{cases} \hat{y}_t = y_t \begin{cases} y_t = 1; \left(CFN/2 \cdot W_l^{acc}\right) \\ y_t = 0; \left(CFP/2 \cdot W_l^{acc}\right) \end{cases} \\ \hat{y}_t \neq y_t \begin{cases} y_t = 1; \left(CFN/2 \cdot W_l^{err}\right) \\ y_t = 0; \left(CFP/2 \cdot W_l^{err}\right) \end{cases} \end{cases} \quad (1)$$

$$D_{l+1}^t = D_l^t \times \begin{cases} \hat{y}_t = y_t ; \left(\varepsilon_l / \left((1-\varepsilon_l) \cdot (\varepsilon_l + W_l^{err})\right)\right) \\ \hat{y}_t \neq y_t \begin{cases} y_t = 1; \left(CFN/2 \cdot W_l^{err}\right) \\ y_t = 0; \left(CFP/2 \cdot W_l^{err}\right) \end{cases} \end{cases} \quad (2)$$

$$W_l^{Total} = \left(W_l^{TP} + W_l^{TN} + W_l^{FP} + W_l^{FN}\right) \quad (3)$$

$$W_l^{acc} = \left(W_l^{TP} + W_l^{TN}\right)/W_l^{Total} \quad (4)$$

$$W_l^{err} = \left(W_l^{FP} + W_l^{FN}\right)/W_l^{Total} \quad (5)$$

$$\varepsilon_l = W_l^{MS}/W_l^{Total} \quad (6)$$

4. Proposed Methodology

The objective of the proposed methodology is to address imbalanced data streams, adaptive cost-sensitive algorithms and online ensemble learning simultaneously to improve G-mean in skewed data streams with changing class priors. It introduces G-mean based weighted cost-functions for false positive and false negative samples resulted in online learning of skewed data stream. These proposed cost-functions combined with online AdaC2 and CSB2 boosting algorithms cater to online learning of nonstationary data with recurrent imbalance shifts.

4.1. G-mean based Cost-Sensitive Boosting Approach

Being computationally more efficient, a cost-sensitive approach is highly suitable to handle the imbalance problem in online data streams [15]. The cost-sensitive approach assigns higher misclassification cost to minority instances than that of majority instances. The critical task in online cost-sensitive approach is the setting of misclassification costs for unknown data [14]. Very few online cost-sensitive algorithms [14, 36] are available. But all of them assumes intrinsic imbalance. Though there are variations in IR these works assume that one class is always in minority and other is in majority and hence *CFN* is always larger than *CFP* to focus on a positive class which is in a minority.

In dynamic data with the extrinsic imbalance, any class may turn in minority or majority in different time intervals as explained in section 3.1. In such scenarios of changing class imbalance, setting misclassification cost of one class always larger than that of other may

not focus on both the classes. G-mean represents the balanced performance of a learner between two classes. Hence, this research presents a G-mean based cost-sensitive boosting approach to deal with frequent shifts in class imbalance. The research objective is to set the cost value for each class such that it maximizes the G-mean. The cost function of the traditional approach is changed to G-mean based adaptively weighted cost function so that it decreases the significance of negative class only when the performance of positive class degrades. Thus, it emphasizes both the classes in case of RIS caused by changing class priors.

Let $|DS_t^+|$ be the total number of positive data samples with label '1' and $|DS_t^-|$ be the total number of negative samples with label '0' received at time *t*. Let Se_t and Sp_t be the sensitivity and specificity respectively, at time *t*.

$$Maximize\ G-mean_t = \left(\sqrt{Se_t \cdot Sp_t}\right) = \sqrt{\left(\frac{TP}{TP+FN} \cdot \frac{TN}{TN+FP}\right)_t}$$

$$\Rightarrow Max\left(\left(1 - \frac{FN}{|DS_t^+|}\right) \cdot \left(1 - \frac{FP}{|DS_t^-|}\right)\right)$$

$$\Rightarrow Min\left(\frac{FN}{|DS_t^+|} - \frac{FP}{|DS_t^-|} + \frac{FN}{|DS_t^+|} \cdot \frac{FP}{|DS_t^-|}\right) \quad (7)$$

$$\Rightarrow Min\left(\left(|DS_t^-|/|DS_t^+|\right) \cdot FN + Se_t \cdot FP\right)$$

$$\Rightarrow Min\sum_{y_t=1}\left(|DS_t^-|/|DS_t^+|\right) \cdot I(\hat{y}_t \neq y_t) + \sum_{y_t=0} Se_t \cdot I(\hat{y}_t \neq y_t)$$

where *I* is an indicator function.

The proposed online cost-sensitive boosting approach sets different cost values for the negative and the positive class referring to Equation (7) to have maximum G-mean. Let CFN_t and CFP_t be the costs of false negative and false positive instances respectively, at time t. Equations (8) and (9) give the G-mean based weighted cost-functions for CFN_t and CFP_t respectively.

$$CFN_t \propto |DS_t^-|/|DS_t^+| \quad (8)$$

$$CFP_t \propto Se_t \quad (9)$$

At any time *t*, if the received data possess skewness, then the IR at that time is ≥ 1, increasing the cost of a positive class. As per Equation (9), the cost of a negative class is directly proportional to the true positive rate. If the learning model is performing satisfactorily with a higher true positive rate, then the cost associated with the negative class does not decrease further. It pulls down the prominence of the negative class only when the true positive rate decreases.

4.2. Proposed Algorithms

This research contributes four algorithms that effectively tackle population shift due to change in class priors in an imbalanced data stream:

1. GOA-RIS
2. AGOA-RIS
3. GOC-RIS

4. AGOC-RIS.

The novel approach reported in this paper modifies the cost-functions in OA and OC [35] to maximize G-mean as mentioned in section 4.1 Also, it applies the test-then-train approach with and without ageing factor to deal with a nonstationary imbalanced data stream.

4.2.1. Algorithms GOA-RIS and GOC-RIS

The algorithms GOA-RIS and GOC-RIS present the modified versions of OA and OC by incorporating G-mean based cost-functions (Equations (8) and (9)) and the test-then-train approach without ageing factor.

Algorithm 1: GOA-RIS

Input: (x_t, y_t) is an incoming data instance at time $t = \{1, 2,...\}$ of data stream DS, Ensemble of B learners, initial costs of CFN_t and CFP_t of positive and negative samples respectively.
Output:

$$H(x_t) = \arg\max_{y_t \in \{0,1\}} \sum_{l=1}^{B} \log\left(W_l^{acc}/W_l^{err}\right) \cdot I\left(h_t(x_t) = y_t\right)$$

Initialize: $M_t^+ = 0, M_t^- = 0, Se_t = 0, W_l^{acc} = 0, W_l^{Total} = 0,$
$W_t^{err} = 0, W_t^{TP} = 0, W_t^{TN} = 0, W_t^{FP} = 0, W_t^{FN} = 0,$ for all base learners $l \in \{1,2,..,B\}$.
Do for each incoming instance (x_t, y_t)
　Set $\lambda = 1$;
　for $l = 1$ to B do

$$W_l^{Total} = W_l^{Total} + \lambda ;$$

Let $r \sim Poisson(\lambda)$;
By test-then-train approach repeat r times
training of the base learner l; $h_t(x_t) = \hat{y}_t$;

if $\left(\hat{y}_t == 1 \,\&\, \& \, y_t == 1\right)$ then

$$W_l^{TP} = W_l^{TP} + \lambda \cdot CFN_t ;$$

Update W_l^{acc} using Equation

$(4); \lambda = \lambda \cdot CFN_t / 2 \cdot W_l^{acc}$;
else if $\left(\hat{y}_t == 0 \,\&\, \& \, y_t == 0\right)$ then
$$W_l^{TN} = W_l^{TN} + \lambda \cdot CFP_t ;$$

Update W_l^{acc} using Equation

$(4); \lambda = \lambda \cdot CFP_t / 2 \cdot W_l^{acc}$;
else if $\left(\hat{y}_t == 0 \,\&\, \& \, y_t == 1\right)$ then

$$W_l^{FN} = W_l^{FN} + \lambda \cdot CFN_t ;$$

Update W_l^{err} using Equation

$(5); \lambda = \lambda \cdot CFN_t / 2 \cdot W_l^{err}$;
else if $\left(\hat{y}_t == 1 \,\&\, \& \, y_t == 0\right)$ then

$$W_l^{FP} = W_l^{FP} + \lambda \cdot CFP_t ;$$

Update W_l^{err} using Equation

$(5); \lambda = \lambda \cdot CFP_t / 2 \cdot W_t^{err}$;
　　end if
　end for
　Calculate Se_t ;

Update $M_t^+, M_t^-, CFN_t, CFP_t$ using Equations (11), (12), (13), (14) respectively;
end

In online data streaming as all *n* instances are unavailable at the start, the IR and sensitivity are updated with each incoming instance. Let M_t^+ and M_t^- be the metrics indicating the percentages of positive and negative class samples arrived by time *t*. The function *f* (Equation (10)) returns 1 for the correct prediction of incoming data instance and 0 otherwise. Equations (11), (12), (13) and (14) respectively describe $M_t^+, M_t^-, CFN_t, CFP_t$.

$$f\left(h_t(x_t)\right) = \begin{cases} 1, \hat{y}_t = y_t \\ 0, \hat{y}_t \neq y_t \end{cases} \tag{10}$$

$$M_t^+ = \frac{(t-1) \cdot M_{t-1}^+ + f\left(h_t(x_t)\right)}{t} \tag{11}$$

$$M_t^- = \frac{(t-1) \cdot M_{t-1}^- + f\left(h_t(x_t)\right)}{t} \tag{12}$$

$$CFN_t = M_t^- / M_t^+ \tag{13}$$

$$CFP_t = \frac{(t-1) \cdot Se_{t-1} + f\left(h_t(x_t)\right)}{t} \tag{14}$$

The Algorithms 1 and 2 describe GOA-RIS and GOC-RIS respectively.

4.2.2. Algorithms AGOA-RIS and AGOC-RIS

The population shift may occur in dynamic data stream leading to change in class priors of the arriving classes. This variation in prior probabilities may result in misclassification. When currently arrived instances are approaching the large number, the contribution of function *f* (Equation (10)) in the calculation of class-percentage Equations (11) and (12) becomes insignificant. Hence, it takes a long span to notice the change in class-percentage by techniques mentioned in Equations (11) and (12). The usage of ageing factors results in adaptation to varying prior probabilities of the classes. More is the age of an instance; lesser is its prominence in class-percentage and cost-functions. Another set of proposed algorithms (AGOA-RIS and AGOC-RIS) uses two ageing factors, namely, data-ageing factor Ψ, ($0 < \Psi < 1$) and sensitivity-ageing factor Φ, ($0 < \Phi < 1$) to more emphasise the latest data and sensitivity.

The Equations (15), (16), (17), and (18) formulate the ageing-based parameters $M_t^+, M_t^-, CFN_t, CFP_t$.

$$M_t^+ = \Psi \cdot M_{t-1}^+ + (1-\Psi) \cdot f\left(h_t(x_t)\right) \tag{15}$$

$$M_t^- = \Psi \cdot M_{t-1}^- + (1-\Psi) \cdot f\left(h_t(x_t)\right) \tag{16}$$

$$CFN_t = M_t^- / M_t^+ \tag{17}$$

$$CFP_t = \Phi \cdot Se_{t-1} + (1-\Phi) \cdot f\left(h_t(x_t)\right) \tag{18}$$

The algorithms AGOA-RIS and AGOC-RIS are same as GOA-RIS and GOC-RIS respectively excluding

line number 19. For ageing based algorithms, line number 19 in both the Algorithms 1 and 2 update parameters $M_t^+, M_t^-, CFN_t, CFP_t$ using Equations (15), (16), (17), and (18), respectively.

Algorithm 2: GOC-RIS

Input: (x_t, y_t) is an incoming data instance at time $t = \{1, 2,...\}$ of data stream DS, Ensemble of B learners, initial costs of CFN_t and CFP_t of positive and negative samples respectively.

Output:

$$H(x_t) = \arg\max_{y_t \in \{0,1\}} \sum_{l=1}^{B} \log\left(\frac{1-\varepsilon_l}{\varepsilon_l}\right) \cdot I\left(h_t(x_t) = y_t\right)$$

Initialize: $M_t^+ = 0, M_t^- = 0, Se_t = 0, W_l^{MS} = 0, W_l^{Total} = 0,$

$W_l^{FP} = 0, W_l^{FN} = 0, W_l^{err} = 0, \varepsilon_l = 0$ *for all base learners $l \in \{1,2,..,B\}$.*

Do for each incoming instance (x_t, y_t)

　Set $\lambda = 1$;

　for $l = 1$ to B do

　　$W_l^{Total} = W_l^{Total} + \lambda$;

　　Let $r \sim Poisson(\lambda)$;

　　By test-then-train approach repeat r times training of the base learner l; $h_t(x_t) = \hat{y}_t$;

　　if $\left(\hat{y}_t == y_t\right)$ then

　　　$\lambda = \varepsilon_l / \left((1 - \varepsilon_l) \cdot \left(\varepsilon_l + W_l^{err}\right)\right)$;

　　else

　　　if $\left(\hat{y}_t == 0 \&\& y_t == 1\right)$ then

　　　　$W_l^{FN} = W_l^{FN} + \lambda \cdot CFN_t$; $W_l^{MS} = W_l^{MS} + \lambda$;

　　　　Update W_l^{err}, ε_l using Equations (5), (6) respectively; $\lambda = \lambda \cdot CFN_t / \left(\varepsilon_l + W_l^{err}\right)$;

　　　else

　　　　$W_l^{FP} = W_l^{FP} + \lambda \cdot CFP_t$;

　　　　$W_l^{MS} = W_l^{MS} + \lambda$;

　　　　Update W_l^{err}, ε_l using Equations (5), (6) respectively; $\lambda = \lambda \cdot CFP_t / \left(\varepsilon_l + W_l^{err}\right)$;

　　　end if

　　end if

　end for

　Calculate Se_t ;

　Update $M_t^+, M_t^-, CFN_t, CFP_t$ using Equations (11), (12), (13), (14) respectively;

end

5. Experimental Results

The reported study focuses on a binary imbalance problem with recurrent changes in IR. It studies different semi-synthetic scenarios in which any of the classes may become a minority class due to change in class priors in that specific time interval.

5.1. Preparation of Datasets for Different Scenarios of RIS

The experiments are performed on one synthetic and two real benchmark datasets with binary class imbalance:-SEA [31], Electricity pricing [16], Weather-

NOAA as introduced in [6]. Table 3 gives the summary of datasets used for the experimentation.

Table 3. Summary of datasets used for experimentation.

Dataset	No. of instances	No. of features	No. of Classes	% Negative	% Positive	IR
SEA	50000	3	2	62.84	37.16	1.69
Electricity Pricing	45312	7	2	57.55	42.45	1.36
Weather	18159	8	2	69	31	2.23

The present empirical study considers different scenarios of imbalance shifts by recurrently changing class priors. We equally divide each data set into four slots. Each slot has chunks of randomly chosen ten samples from an original dataset. We apply specific IR to all chunks in each slot to get semi-synthetic datasets with RIS. By abruptly varying class priors of each class in each slot, three different data sequences are produced resulting in three scenarios of RIS as described in Table 4. Figure 1 portrays the variations in the percentages of positive and negative data samples in each slot of different scenarios of RIS. In RIS-Extreme both the classes experience minority in alternate slots whereas in the remaining two scenarios (RIS-Mild and RIS-Variable) only positive class is always in a minority but with abruptly changing IR in each slot.

Table 4. Scenarios of RIS.

Scenario	Data Description (% negative, % positive, IR)			
	1st Slot	2nd Slot	3rd Slot	4th Slot
RIS-Mild	(50,50,1)	(90,10,9)	(50,50,1)	(90,10,9)
RIS-Extreme	(90,10,9)	(10,90,0.11)	(90,10,9)	(10,90,0.11)
RIS-Variable	(50,50,1)	(70,30,2.33)	(80,20,4)	(90,10,9)

Figure 1. Data samples distribution in different scenarios of RIS.

5.2. Experimentation Setup

The research has implemented the state-of-the-art algorithms OA and OC by the test-then-train approach to have a fair evaluation of four proposed algorithms GOA-RIS, GOC-RIS, AGOA-RIS and AGOC-RIS against them. All these algorithms are implemented using an ensemble of five Hoeffding trees as base learners (i.e., $B = 5$). The initial costs of a false positive and a false negative are 0.7 and 1

respectively. We set a data-ageing factor (Ψ), and a sensitivity-ageing factor (Φ) to 0.9. The most popular analytic tool [29] R is used for this research work.

5.3. Empirical Results

In online streaming data, prediction of the total number of instances *n* is unfeasible. Hence, instead of using any cross-validation technique, this study has opted prequential analysis [3] by which each data sample is tested on the learning model before its training, and based on it the performance metrics are incrementally calculated. At each time step, the G-mean is incrementally updated by referring to the recent performance. As per the mentioned scenarios of RIS (section 5.1), the shift in imbalance occurs in all slots. We reset G-mean values after each slot to avoid the influence of pre-shift performance on post-shift values.

Table 5. G-mean (%) of all algorithms in all scenarios of RIS.

Dataset	Slot	AdaC2 Family			CSB2 Family		
		OA	GOA-RIS	AGOA-RIS	OC	GOC-RIS	AGOC-RIS
RIS-Mild (Slots 1&3: IR=1; Slots 2&4: IR=9)							
SEA	1	74.62	**78.12**	77.56	79.50	80.05	**80.42**
	2	77.45	77.39	**78.68**	75.87	75.61	**76.00**
	3	75.09	76.76	**78.64**	80.09	79.64	79.67
	4	78.35	**79.04**	77.14	76.96	75.73	76.01
Electricity	1	**77.44**	69.80	69.29	70.71	70.84	**70.96**
	2	68.86	69.33	**76.95**	66.29	**69.39**	68.73
	3	72.08	71.17	**72.92**	73.27	73.12	**73.72**
	4	**73.66**	72.05	72.20	69.74	70.41	**72.17**
Weather	1	**77.70**	67.19	69.71	64.73	**69.07**	66.92
	2	66.25	67.79	**73.35**	52.36	65.31	**71.28**
	3	69.65	71.53	**72.22**	71.07	71.09	**72.32**
	4	65.79	**67.97**	**71.69**	63.92	63.49	68.71
RIS-Extreme (Slots 1&3: IR=9; Slots 2&4: IR=0.11)							
SEA	1	71.31	**79.75**	76.18	64.14	**74.88**	71.72
	2	48.57	48.42	**66.38**	59.10	50.91	**69.71**
	3	77.25	**78.52**	77.65	**78.44**	77.66	76.60
	4	51.11	56.64	**69.30**	63.26	64.24	**70.74**
Electricity	1	61.10	**79.84**	71.88	56.01	**67.90**	64.90
	2	48.31	51.09	**63.03**	54.95	55.80	**66.18**
	3	67.05	68.95	**69.98**	68.25	**70.37**	69.32
	4	57.43	55.66	**65.21**	63.80	65.54	**68.44**
Weather	1	50.28	**72.65**	71.14	38.97	**75.79**	67.68
	2	49.34	46.32	**62.11**	57.31	50.23	**65.49**
	3	59.99	64.06	**67.18**	61.80	66.96	63.82
	4	54.66	55.93	**65.41**	61.67	59.22	**65.39**
RIS-Variable (Slot 1: IR=1; Slot 2: IR=2.33; Slot 3: IR=4; Slot 4:IR=9)							
SEA	1	72.87	**78.39**	77.67	78.94	**80.02**	79.63
	2	79.23	**80.35**	79.07	**81.06**	80.35	80.40
	3	**80.21**	80.12	79.85	**80.37**	80.27	79.60
	4	**77.55**	76.33	77.03	73.04	74.09	**75.65**
Electricity	1	**77.74**	70.60	71.25	70.77	**70.87**	69.84
	2	70.72	**71.57**	71.46	73.17	72.59	72.21
	3	72.23	73.03	**74.16**	72.64	72.12	**72.87**
	4	68.15	70.70	**71.14**	65.29	68.83	**69.23**
Weather	1	**70.05**	67.41	67.32	65.90	68.33	**68.79**
	2	70.10	70.33	**75.95**	68.15	70.48	**71.16**
	3	68.59	71.39	**74.08**	65.20	67.55	**67.95**
	4	62.21	67.58	**73.39**	53.09	64.52	**69.67**
Original							
SEA		79.62	81.25	**81.35**	83.27	83.27	83.06
Electricity		85.68	**85.83**	85.12	84.79	**85.09**	84.70
Weather		70.93	70.71	**72.51**	70.91	**72.89**	71.93

Table 5 gives the percentages of G-mean of OA, GOA-RIS, AGOA-RIS, OC, GOC-RIS and AGOC-RIS

observed in each slot of different scenarios of imbalance shifts of each dataset. The proposed G-mean based weighted cost function assures better classification performance over both the classes. It reduces the influence of the negative class only when the misclassification of positive class increases. Hence the proposed algorithms result in improved post-shift (slots 2, 3, and 4) G-mean than that of OA and OC in all scenarios of RIS. Especially, in the case of RIS-Extreme where both the classes undergo minority, the rise in G-mean values of proposed algorithms is more prominent. Influence of old data in cost calculations has resulted in competent G-mean values of GOA-RIS and GOC-RIS when experimenting with an original dataset.

Table 6. Scenario wise average G-mean (%).

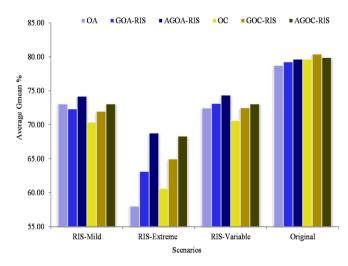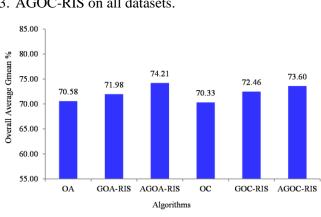| Scenario	AdaC2 Family			CSB2 Family		
	OA	GOA-RIS	AGOA-RIS	OC	GOC-RIS	AGOC-RIS
RIS-Mild	73.08	72.34	**74.20**	70.38	71.98	**73.08**
RIS-Extreme	58.03	63.15	**68.79**	60.64	64.96	**68.33**
RIS-Variable	72.47	73.15	**74.36**	70.63	72.50	**73.08**
Original	78.74	79.26	**79.66**	79.66	**80.42**	79.89
Overall Average	70.58	71.98	**74.25**	70.33	72.46	**73.60**

Figure 2. Scenario wise average G-mean (%) of all algorithms.

Table 6 presents the percentages of average G-mean of six algorithms in all slots of all datasets for each scenario of experimentation. It shows that G-mean values achieved by the proposed novel approach are superior to algorithms OA and OC. More contribution of recent data in cost updating has given a better performance of ageing-based algorithms in all scenarios of RIS. Figure 2 depicts the scenario wise percentages of average G-mean of three algorithms of AdaC2 family:

1. OA
2. GOA-RIS
3. AGOA-RIS.

and three algorithms of CSB2 family:

1. OC

2. GOC-RIS
3. AGOC-RIS on all datasets.

Figure 3. Overall average G-mean (%) of all algorithms.

Figure 3 presents the percentages of overall average G-mean of all six algorithms OA, GOA-RIS, AGOA-RIS, OC, GOC-RIS, AGOC-RIS. These values are computed by averaging the scenario wise percentages of average G-mean of all algorithms. The empirical analysis exhibits that ageing and G-mean based online AdaC2 algorithm AGOA-RIS achieves the highest percentage of overall average G-mean.

Table 7. G-mean based overall average ranking.

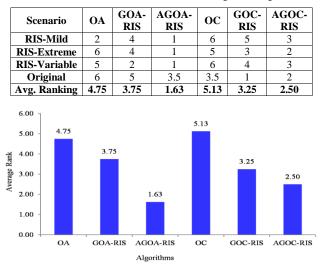| Scenario	OA	GOA-RIS	AGOA-RIS	OC	GOC-RIS	AGOC-RIS
RIS-Mild	2	4	1	6	5	3
RIS-Extreme	6	4	1	5	3	2
RIS-Variable	5	2	1	6	4	3
Original	6	5	3.5	3.5	1	2
Avg. Ranking	**4.75**	**3.75**	**1.63**	**5.13**	**3.25**	**2.50**

Figure 4. Comparison of G-mean based average ranking of all algorithms.

Table 7 computes the G-mean based overall average ranking of each algorithm. These rank values are based on the results described in Table 6. Figure 4 portrays the comparison of G-mean based average ranking of six algorithms. It shows that the proposed four algorithms GOA-RIS, AGOA-RIS, GOC-RIS and AGOC-RIS beat both, OA and OC. AGOA-RIS has the least and OC has the highest value of average rank indicating that AGOA-RIS is the best performer and OC is the worst performer among these six variations of ensembles.

5.4. Statistical Results

To ensure better comparative analysis and to test whether mentioned six algorithms have noteworthy differences, this study has used nonparametric statistical tests as per the recommendations in [4, 13]. We apply Iman-Davenport test to check a significant difference among these algorithms. Table 8 provides the result of Iman-Davenport test on G-mean in different RIS. The resultant p-value for RIS-Extreme indicates that there is at least one algorithm that behaves differently and rejects a null hypothesis ($H_{0:}$ All algorithms are similar) at significant level $\alpha=0.05$. Hence for RIS-Extreme, we opt for Aligned Rank posthoc test with Bergmann and Hommel's correction [4, 13]. Table 9 presents the result of posthoc test.

Except for RIS-Extreme, all scenarios describe a minority in just positive class. For such scenarios, all six algorithms do not show a significant distinction in statistical tests (Table 8). It underpins that the proposed G-mean based adaptive weighted cost-functions (section 4.1) are equally competent to the traditional cost-sensitive approach when only positive class is consistently in a minority. As both, the classes become minor in different slots of RIS-Extreme, the behaviour of proposed algorithms due to G-mean based adaptive weighted cost-function discern from OA and OC. The least values (boldfaced) in Table 9 resulted by Aligned Rank posthoc test applied to RIS-Extreme infer the same.

Table 8. Iman-Davenport test on G-mean ($\alpha = 0.05$).

Scenario	p-value	Remark
RIS-Mild	0.1437	Accept H_0
RIS-Extreme	0.00001104	**Reject H_0**
RIS-Variable	0.8412	Accept H_0
Original	0.663	Accept H_0

Table 9. Aligned Rank posthoc test with Bergmann and Hommel's correction on G-mean in RIS-Extreme ($\alpha=0.05$).

	OA	GOA-RIS	AGOA-RIS	OC	GOC-RIS	AGOC-RIS
OA	NA	0.30	**0.00**	0.30	**0.03**	**0.00**
GOA-RIS	0.30	NA	**0.04**	0.92	0.86	0.18
AGOA-RIS	**0.00**	**0.04**	NA	**0.02**	0.30	0.92
OC	0.30	0.92	**0.02**	NA	0.86	0.12
GOC-RIS	**0.03**	0.86	0.30	0.86	NA	0.86
AGOC-RIS	**0.00**	0.18	0.92	0.12	0.86	NA

6. Conclusions and Future Scope

The presented research contributes to a novel approach that combines class imbalance, cost-sensitive algorithms, online, ensemble learning to construct a joint solution. Through this joint solution, the reported work addresses the problem of recurrent imbalance shift by changing class priors in nonstationary data. The paper presents four online adaptive cost-sensitive boosting algorithms as modifications to algorithms AdaC2 and CSB2 as below:

1. GOA-RIS
2. GOC-RIS
3. AGOA-RIS
4. AGOC-RIS.

To cope with the online streaming environment the study follows test-then-train approach and Poisson parameter λ that approximates the weight updating rule of AdaC2 and CSB2. The proposed algorithms assign different costs to positive and negative classes with the objective of G-mean maximization in various scenarios of imbalance shifts. The work introduces the forgetting of old data through ageing factors to adapt to recent changes in data.

The empirical analysis shows that the proposed algorithms work globally better than the state-of-the-art algorithms in all scenarios of RIS. The G-mean based overall average ranking presents AGOA-RIS as the best performer among all six mentioned algorithms. Referring to the statistical tests the proposed algorithms show distinguished performance when both the classes exhibit minority due to the extrinsic imbalance in the streaming environment. However, they show similar behaviours as that of state-of-the-art cost-sensitive approaches when data possess changing intrinsic imbalance. It indicates the suitability of proposed G-mean based adaptively weighted cost-functions even in conventional cost-sensitive approach.

Concerning the current work there are some points we would like to focus on in the future. We have considered a few scenarios of RIS with changing IR, but there is a variety of concept drifting scenarios in online streaming data. Also, this work is limited to a binary class imbalance problem and Hoeffding tree as a base learner. We will test our algorithms with different base learners using a variety of real data streams of multiclass imbalance. More efficient online adaptive algorithms to handle different types of concept drifts in multiclass imbalanced streaming data constitute future scope.

References

[1] Anastasovski G. and Goseva-Popstojanova K., "Classification of Partially Labeled Malicious Web Traffic in the Presence of Concept Drift," *in Proceedings of 8th International Conference on Software Security and Reliability-Companion*, San Francisco, pp. 130-139, 2014.

[2] Ancy S. and Paulraj D., "Handling Imbalanced Data with Concept Drift by Applying Dynamic Sampling and Ensemble Classification Model," *Computer Communications*, vol. 153, pp. 553-560, 2020.

[3] Dawid A., "Present Position and Potential Developments: Some Personal Views: Statistical Theory: the Prequential Approach," *Journal of the Royal Statistical Society, Series A (General)*, vol. 147, no. 2, pp. 278-292, 1984.

[4] Demˇsar J., "Statistical Comparisons of Classifiers over Multiple Data Sets," *Journal of Machine Learning Research*, vol. 7, pp. 1-30, 2006.

[5] El-Halees A., "Filtering Spam E-Mail from Mixed Arabic and English Messages: A Comparison of Machine Learning Techniques," *The International Arab Journal of Information Technology*, vol. 6, no. 1, pp. 52-59, 2009.

[6] Elwell R. and Polikar R., "Incremental Learning of Concept Drift in Nonstationary Environments," *IEEE Transactions on Neural Networks*, vol. 22, no. 10, pp. 1517-1531, 2011.

[7] Fan W., Stolfo S., Zhang J., and Chan P., "AdaCost: Misclassification Cost-Sensitive Boosting," *in Proceedings of the 6th International Conference on Machine Learning*, San Francisco, pp. 97-105, 1999.

[8] Ferreira L., Gomes H., Bifet A., Bifet A., and Oliveira L., "Adaptive Random Forests with Resampling for Imbalanced Data Streams," *in Proceedings of International Joint Conference on Neural Networks*, Budapest, pp. 1-6, 2019.

[9] Freund Y. and Schapire R., "A Decision-Theoretic Generalization of On-line Learning and an Application to Boosting," *Journal of Computer System Sciences*, vol. 55, no. 1, pp. 119-139, 1997.

[10] Galar M., Fernandez A., Barrenechea E., Bustince H., and Herrera F., "A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches," *IEEE Transactions on Systems, Man and Cybernetics-Part C: Applications and Reviews*, vol. 42, no. 4, pp. 463-484, 2012.

[11] Gama J., Žliobaitė I., Bifet A., Pechenizkiy M., and Bouchachia A., "A Survey on Concept Drift Adaptation," *ACM Computing Surveys*, vol. 1, no. 1, pp. 1-35, 2013.

[12] Gao X., Chen Z., Tang S., Zhang Y., and Li J., "Adaptive Weighted Imbalance Learning with Application to Abnormal Activity Recognition," *Neurocomputing*, vol. 173, pp. 1927-1935, 2016.

[13] García S., Fernández A., Luengo J., and Herrera F., "Advanced Nonparametric Tests for Multiple Comparisons in the Design of Experiments in Computational Intelligence and Data Mining: Experimental Analysis of Power," *Information Sciences*, vol. 180, no. 10, pp. 2044-2064, 2010.

[14] Ghazikhani A., Monsefi R., and Yazdi H., "Online Cost-Sensitive Neural Network Classifiers for Nonstationary and Imbalanced Data Streams," *Neural Computing and Applications*, vol. 23, pp. 1283-1295, 2013.

[15] Haixiang G., Yijing L., Shang J., Mingyun G., Yuanyue H., and Bing G., "Learning from Class-Imbalanced Data: Review of Methods and Applications," *Expert Systems with Applications*, vol. 73, pp. 220-239, 2017.

[16] Harries M., "Splice-2 Comparative Evaluation: Electricity Pricing," Technical Report, The University of South Wales, 1999.

[17] He H. and Garcia E., "Learning from Imbalanced Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263-1284, 2009.

[18] Joshi M., Kumar V., and Agarwal R., "Evaluating Boosting Algorithms to Classify Rare Classes: Comparison and Improvements," *in Proceedings of IEEE International Conference Data Mining*, San Jose, pp. 257-264, 2001.

[19] Kelly M., Hand D., and Adams N., "The Impact of Changing Populations on Classifier Performance," *in Proceedings of 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Diego, pp. 367-371, 1999.

[20] Landesa-Vázquez I. and Alba-Castro J., "Revisiting AdaBoost for Cost-Sensitive Classification Part I: Theoretical Perspective," *Computing Research Repository (CoRR)*, Cornell University, vol. abs/1507/04125, 2015.

[21] Landesa-Vázquez I. and Alba-Castro J., "Revisiting AdaBoost for Cost-Sensitive Classification Part II: Empirical Analysis," *Computing Research Repository (CoRR)*, Cornell University, vol. abs/1507/04126, 2015.

[22] Lim P., Goh C., and Tan K., "Evolutionary Cluster-Based Synthetic Oversampling Ensemble (ECO-Ensemble) for Imbalance Learning," *IEEE Transactions on Cybernetics*, vol. 47, no. 9, pp. 2850-2861, 2017.

[23] Lu Y., Cheung Y., and Tang Y., "Adaptive Chunk-based Dynamic Weighted Majority for Imbalanced Data Streams with Concept Drift," *IEEE Transactions on Neural Networks Learning Systems*, vol. 31, no. 8, pp. 2764-2778, 2020.

[24] Nikolaou N., Edakunni N., Kull M., Flach P., and Brown G., "Cost-Sensitive Boosting Algorithms: Do We Really Need Them?," *Machine Learning*, vol. 104, pp. 359-384, 2016.

[25] Orriols-Puig A. and Bernad´o-Mansilla E., "Evolutionary Rule-based Systems for Imbalanced Data Sets," *Soft Computing*, vol. 13, no. 3, pp. 213-225, 2009.

[26] Oza N., "Online Ensemble Learning," PhD Dissertation, University of California, 2001.

[27] Park Y., Luo L., Parhi K., and Netoff T., "Seizure Prediction with Spectral Power of EEG Using Cost-Sensitive Support Vector Machines," *Epilepsia*, vol. 52, no. 10, pp. 1761-1770, 2011.

[28] Ren S., Zhu W., Liao B., Li Z., Wang P., Li K., Chen M., and Li Z., "Selection-based Resampling Ensemble Algorithm for Nonstationary Imbalanced Stream Data Learning," *Knowledge-Based Systems*, vol. 163, pp. 705-722, 2019.

[29] Rexer K., "A Decade of Surveying Analytic Professionals: Survey Highlights," Rexer Analytics Survey Summary Report, 2017.

[30] Somasundaram A. and Reddy S., "Parallel and Incremental Credit Card Fraud Detection Model to Handle Concept Drift and Data Imbalance," *Neural Computing and Applications*, vol. 31, no. 1, pp. 3-14, 2019.

[31] Street W. and Kim Y., "A Streaming Ensemble Algorithm (SEA) for Large-Scale Classification," *in Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, pp. 377-382, 2001.

[32] Sun J., Li H., Fujita H., Fu B., and Ai W., "Class-imbalanced Dynamic Financial Distress Prediction based on Adaboost-SVM Ensemble Combined with SMOTE and Time Weighting," *Information Fusion*, vol. 54, pp. 128-144, 2020.

[33] Sun Y., Kamel M., Wong A., and Wang Y., "Cost-Sensitive Boosting for Classification of Imbalanced Data," *Pattern Recognition*, vol. 40, no. 12, pp. 3358-3378, 2007.

[34] Ting K., "A Comparative Study of Cost-Sensitive Boosting Algorithms," *in Proceedings of the 17th International Conference on Machine Learning*, Stanford University, pp. 983-990, 2000.

[35] Wang B. and Pineau P., "Online Bagging and Boosting for Imbalanced Data Streams," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 12, pp. 3353-3366, 2016.

[36] Wang J., Zhao P., and Hoi S., "Cost-Sensitive Online Classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 10, pp. 2425-2438, 2014.

[37] Wang S., Minku L., and Yao X., "A Learning Framework for Online Class Imbalance Learning," *in Proceedings of IEEE Symposium on Computational Intelligence and Ensemble Learning*, Singapore, pp. 36-45, 2013.

[38] Wang S., Minku L., and Yao X., "Resampling-Based Ensemble Methods for Online Class Imbalance Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 5, pp. 1356-1368, 2015.

[39] Wong M., Seng K., and Wong P., "Cost-Sensitive Ensemble of Stacked Denoising Autoencoders for Class Imbalance Problems in Business Domain," *Expert Systems with Applications*, vol. 141, pp. 112918, 2020.

[40] Wu D., Wang Z., Chen Y., and Zhao H., "Mixed-kernel based Weighted Extreme Learning Machine for Inertial Sensor based Human Activity Recognition with Imbalanced Dataset," *Neurocomputing*, vol. 190, pp. 35-49, 2016.

[41] Wu X., Kumar V., Quinlan J., Ghosh J., Yang Q., Motoda H., McLachlan G., Ng A., Liu B., Yu P., Zhou Z., Steinbach M., Hand D., and

Steinberg D., "Top 10 Algorithms in Data Mining," *Knowledge and Information Systems*, vol. 14, no. 1, pp. 1-37, 2008.

[42] Yin Q., Zhang J., Zhang C., and Liu S., "An Empirical Study on the Performance of Cost-Sensitive Boosting Algorithms with Different Levels of Class Imbalance," *Mathematical Problems in Engineering*, vol. 2013, 2013.

[43] Zhu T., Lin Y., and Liu Y., "Synthetic Minority Oversampling Technique for Multiclass Imbalance Problems," *Pattern Recognition*, vol. 72, pp. 327-340, 2017.

Radhika Kulkarni received the M.Tech. degree in Computer Science and Technology from the Shivaji University, Kolhapur, India in 2011. She is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai India. She is working as an Assistant Professor in the Department of Information Technology, Pune Institute of Computer Technology, Pune, India. Her current research interests include Machine Learning, Data Analytics and Big Data.

Subramanion Revathy (S. Revathy) is presently working as an Associate Professor in the Department of Information Technology, Sathyabama Institute of Science and Technology, Chennai India. Her research interest includes Machine Learning, Data Analytics and Big Data. She has published over twenty papers in refereed journals.

Suhas Patil received the Ph.D. degree in Computer Science and Engineering from Bharati Vidyapeeth Deemed University, Pune, India in 2009. He is currently working as a Professor in the Department of Computer Science and Engineering, Bharati Vidyapeeth Deemed University College of Engineering, Pune, India. His research area includes Machine Learning, Expert System, Computer Network, Operating System, System Software. He has published over 65 papers in international journals, 36 papers in international conferences and 42 papers in national conferences.