Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Application of concept Drift Detection and Adaptive Framework for Non Linear Time Series Data from Cardiac Surgery

Rajarajan Ganesan[1] | Tarunpreet Kaur[2] | Alisha Mittal[1] | Mansi Sahi[2] | Sushant Konar[1] | Tanvir Samra[1] | Goverdhan Dutt Puri[1] | Shayam Kumar Singh Thingnum[3] | Nitin Auluck[2]

[1]Department of Anaesthesia and Intensive Care, Post Graduate Institute of Medical Education and Research, Chandigarh, Punjab, India

[2]Department of Computer Science and Engineering, Indian Institute of Technology, Ropar, Punjab, India

[3]Department of Cardiovascular and Thoracic Surgery, Post Graduate Institute of Medical Education and Research, Chandigarh, Punjab, India

Correspondence
Corresponding author Nitin Auluck,
Email: nitin@iitrpr.ac.in

Abstract

The quality of machine learning (ML) models deployed in dynamic environments tends to decline over time due to disparities between the data used for training and the upcoming data available for prediction, which is commonly known as drift. Therefore, it is important for ML models to be capable of detecting any changes or drift in the data distribution and updating the ML model accordingly. This study presents various drift detection techniques to identify drift in the survival outcomes of patients who underwent cardiac surgery. Additionally, this study proposes several drift adaptation strategies, such as adaptive learning, incremental learning, and ensemble learning. Through a detailed analysis of the results, the study confirms the superior performance of ensemble model, achieving a minimum mean absolute error (MAE) of 10.684 and 2.827 for predicting hospital stay and ICU stay, respectively. Furthermore, the models that incorporate a drift adaptive framework exhibit superior performance compared to the models that do not include such a framework.

KEYWORDS
Concept drift, Adaptation, Incremental, Ensemble, Machine Learning, Cardiac surgery

1 | INTRODUCTION

Drift detection refers to the identification of changes in a time series that occur gradually, or suddenly, over time. These changes can affect the accuracy of predictions made by machine learning (ML) models. Drift is also known as the occurrence where the statistical characteristics of datasets undergo unforeseen changes as time progresses. Drift can occur due to a variety of factors, including alterations in patient demographics, advancements in medical technology, shifts in clinical protocols, or the evolution of disease trends. In the context of medical data, the presence of drift can adversely affect the precision and reliability of predictive models built using past data. This unfolds due to the fact that the assumptions formulated during model training may become invalid, resulting in suboptimal adaptability to new and unfamiliar data.[1].

Drift can take on various forms, each with distinct implications for predictive modeling:

1. Concept Drift: This occurs when the associations between predictor variables and the target variable undergo changes over time. To illustrate, a diagnostic feature once strongly linked to a particular ailment might lose relevance due to shifts in medical knowledge, or the incorporation of novel diagnostic methods.
2. Population Drift: Changes in the demographics and characteristics of the patient population being studied can lead to disparities in data distributions. This phenomenon might occurs due to changes in patient demographics, shifts in disease prevalence, or evolving patient characteristics.
3. Data Source Drift: Medical data, drawn from diverse origins such as hospitals, clinics, or geographic regions, can exhibit discrepancies in data collection practices or patient compositions. These variations across data sources can introduce dissimilarities in data distributions, consequently influencing the performance of predictive models.

However, concept drift within medical applications holds an immense significance when compared to alternative forms of drift, such as data source or population drift, primarily due to the dynamic nature inherent to medical research, clinical trials, and treatment methodologies. Within the healthcare domain, where precise predictions wield a direct influence over patient outcomes, the significance of concept drift is further elevated.

Improvement in anaesthesia and surgical techniques can lead to a decrease in the length of intensive care unit (ICU) stay and hospital stay, causing changes in the statistical distributions of ICU stay and hospital stay, which can affect the prediction of survival outcomes for cardiac surgery patients. For predictive maintenance, a drift detector is used to identify statistical differences in the data distribution. This allows for the retraining of the existing model or replacement with a new one to effectively accommodate the upcoming batch of datasets. Therefore, detecting and adapting to concept drift are the major challenges associated with the high variability issue in nonlinear time series data[2].

Conventional ML models are unable to adapt to concept drift, necessitating the development of adaptive frameworks that can handle both predictable and unpredictable changes in non-linear time series data. ML model adaptation strategies are broadly classified into passive and active methods. Passive approaches require updating the model every time new data are made available. Besides achieving accurate prediction, passive approaches are challenging to implement, since the model updating requires significant computational resources. On the other hand, active approaches strive to achieve a trade-off between update frequency and model accuracy by modifying the ML model solely when there is evidence of concept drift[3].

As a result, this article explores various active approaches for a drift adaptive framework in time series data analytics, specifically for predicting survival outcomes using invasive blood pressure (IBP) and heart rate (HR) data collected during cardiac surgery. The aim of this study is to design a drift adaptive framework that employs IBP and HR time series parameters to forecast the duration of ICU and hospital stays following cardiac surgeries.

The paper is structured as follows — in Section 2, the techniques for detecting concept drift are discussed. The pro- posed methodology for different drift adaptation frameworks is presented in Section 3, while Section 4 is dedicated to presenting the results of the experiments. Section 5 provides the conclusion of the study.

2 | RELATED WORK

The term 'Drift Detection' pertains to methods and mechanisms used for identifying shifts or alterations by pinpointing instances of change or brief periods in which modifications take place. These alterations render existing models ineffective in forecasting the behavior of the current data[2]. Detectors for concept drift are techniques capable of indicating shifts in data distributions, relying on insights from classifier performance or incoming data. Such indications typically prompt the requirement to update, substitute, or retrain the model. Several researchers have proposed various frameworks in order to deal with drift detection issue. In[4], the authors have discussed various strategies to adapt to the concept drift situation. Lin et al.[5] have used ensemble learning in industrial IoT for detecting drift. The framework forecasts the moment at which machines begin to function abnormally, and proactively manage or substitute their components in order to prevent the production of substantial faulty products. By combining a variety of multiple classifiers, ensemble learning offers an effective solution for tackling concept drift with strong performance. In[6], Agrahari S et al. used the Light Gradient Boosting model for detecting drift in IoT data streams. The framework was tested on an anomaly dataset. The proposed model performed well with very high accuracies on the benchmark NSL-KDD and IoTID20 datasets. In[3], Fields T et al. have used various ML models like MLP, LSTM, GRU, and CNN for detecting drift in time-series data. The results show that the RNN model performed well in comparison to other models. Also, in order to deal with drift, in[3], the authors added noise to the training data, which improved the model accuracy.[7], Yu S et al. proposed a hierarchical hypothesis testing framework which detects as well as adapts to different kinds of concept drifts such as abrupt, recurrent etc. The experiments were performed on both real-world datasets and simulated data in order to demonstrate the effectiveness of the proposed framework. in[8], the authors introduced a concept drift detection method, named DetectA, tailored for identifying sudden concept shifts. The primary innovation of this approach lies in its proactive nature, that aims to identify an upcoming concept drift, in contrast to the majority of drift detection methods that only recognize concept shifts after they have occured.

3 | CONCEPT DRIFT DETECTION

At a particular moment, contemplate a data instance com- prising feature vectors represented by X and corresponding output values, denoted as y, in the feature space. The concept drift, which is represented as $p(X, y)$, refers to the joint distribution of

these feature vectors and output values. This occurs when the original joint distribution transforms into a new data distribution, which might happen gradually, or suddenly among two concepts[9]. In practical situations, the length of time of patient stay in hospitals and ICUs changes dynamically over time, for example, during the Covid-19 pandemic, the duration of ICU and hospital stays increased. These local changes can affect the entire hospital environment, leading to a change in the concept. Various statistical methods are utilized to identify concept drift by contrasting the distribution of previous and present data instances, which will be discussed in the following sections.

3.1 | Population Stability Index (PSI) Test

The PSI test is a statistical technique that can be utilized to identify alterations or drifts in the distribution of data instances over time. To compare the distributions of a historical dataset and a new dataset, the PSI test computes an index value that indicates the degree of variation among the two distributions. A PSI value of less than 0.1 usually indicates a stable population, while values ranging from 0.1 to 0.25 signify a moderate level of change, and values over 0.25 show considerable change or drift in the distribution between two populations over time. Thus, the larger the dissimilarities among the distributions, the higher the PSI value, indicating a significant drift[5].

3.2 | Kolmogorov-Smirnov (KS) Test

In the field of drift detection, the KS test is a non-parametric method that examines the cumulative distribution functions (CDFs) of past and current datasets to detect any notable dissimilarities. The primary objective of this method is to ascertain whether two sets of data samples share the same distribution. To determine the statistical significance of the observed change in the p-value, a hypothesis test is utilized. If the p-value is lower than the predetermined significance level (generally 0.05), the null hypothesis (which states that the test statistic will not indicate a significant difference among the historical and current data) is rejected, and it is concluded that concept drift has taken place[10].

To evaluate the statistical importance of the change detected in the p-value, a hypothesis test is employed. If the p-value is lower than the predetermined level of significance (often 0.05), the null hypothesis, which presumes that the test statistic will not reveal a significant difference between the previous and current data, is rejected. The sentence conveys the same meaning, but the words are rephrased.

To evaluate the statistical importance of the observed change in the p-value, a hypothesis test is utilized. If the p-value is lower than the predetermined level of significance, typically 0.05, the null hypothesis is rejected. The null hypothesis assumes that there will be no significant difference between the old and new data, as demonstrated by the test statistic.

3.3 | Wasserstein Distance Metric (WDM) Test

The WDM, also refereed to as Earth Mover's Distance (EMD), is a metric used to measure the distance between historical and current data distributions, and determine if there is a major shift or drift in either distribution. If the variance is found to be statistically significant, it will indicate the occurrence of a considerable concept drift. The distance values vary from 0 to 1, and the degree of concept drift increases with an increase in the distance value. A distance of "1" denotes that the new concept is entirely distinct from the prior one, while a distance of "0" indicates that the two data samples are same[6].

3.4 | Cramer Von Mises (CVM) Test

The CVM is a statistical hypothesis test that examines whether a given dataset adheres to a specific probability distribution. It calculates CVM statistics which assess the difference between the empirical CDF (ECDF) of the available data, and the theoretical CDF of the distribution under consideration. If the CVM statistic surpasses the critical value at a predetermined significance level, it implies that there is a drift in the time series. A higher CVM statistic indicates a greater deviation among the two distributions[6].

3.5 | Mann-Whitney U (MWU) Test

The MWU test is a statistical hypothesis test that is non- parametric in nature and determines if two independent samples are derived from a similar distribution. The test can also be applied for detecting deviations by comparing incoming datasets to the baseline datasets. The resulting p-value of the test statistic measures the extent of divergence. When the p-value is less than the pre-defined significance level (usually 0.05), it rejects the null hypothesis, which assumes that both datasets come from the same distribution, and it is concluded that the two distributions are distinct[6].

4 | ML- BASED DRIFT ADAPTIVE FRAMEWORKS

The distinctive characteristics of the healthcare sector, wherein predictions hold a direct sway over patient results, emphasize the importance of tackling concept drift. When delving into the prediction of ICU stays and hospital stays, the incorporation of drift adaptation takes on substantial significance.

To address the challenge of drift adaptation, three active approaches, namely, adaptive learning, incremental learning, and ensemble learning techniques, can be employed, as discussed in[7]. The objective of this study is to design an ML-based adaptive framework capable of handling time and memory constraints, as well as addressing concept drift problems mentioned in the previous section.

4.1 | Adaptive Learning

Adaptive learning is a method of handling concept drift by retraining the learning model on a modified dataset, which includes both new and old data, after detecting a drift[4,11].

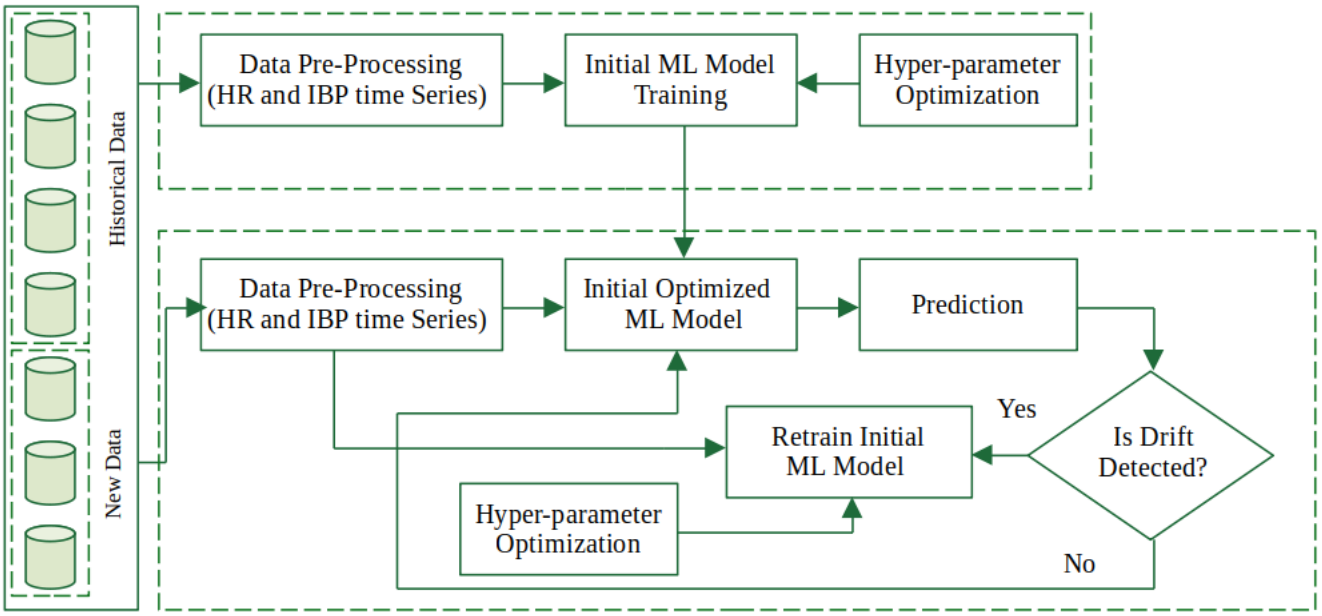

FIGURE 1 The overall methodology followed in Adaptive Learning.

Figure 1 demonstrates the overall architecture of the adaptive learning based drift framework. Initially, the available datasets undergo data pre-processing which includes artifacts removal and normalization, as described in Konar et al.[12]. Next, an ML model is trained on the available datasets using a subset of hyperparameters that provide the best accuracy. The trained ML model is then used to predict the ICU and hospital stay for the upcoming batch of data. After the survival outcomes are predicted,

drift detection methods are used to determine if a concept drift exists. If no drift is detected, the historical model is used to make predictions on the latest batch of datasets. However, if a drift is detected, the regression ML model issues a warning signal, indicating a change in the environment, such as a change in user behavior or data distribution. The new datasets are then merged with the subset of historical datasets, and the ML model is retrained from scratch on the combined dataset. Finally, the retrained model is used for subsequent predictions and evaluated for accuracy in the new environment.

Despite being fast and easy to implement, adaptive learning has a forgetting mechanism that discards old data samples, which may contain useful historical information.

4.2 | Incremental Learning

Incremental learning is a method of continuously updating a predictive model as new data becomes available, allowing it to adapt to changes in the underlying data distribution. Unlike adaptive learning, which often retrains the entire learning model on the altered dataset, incremental learning is capable of preserving the historical trends and patterns of the entire dataset in the ML model without the need to store all the data. This enables the model to adjust to new data patterns by making partial updates to the original model within a short execution time [13,14].

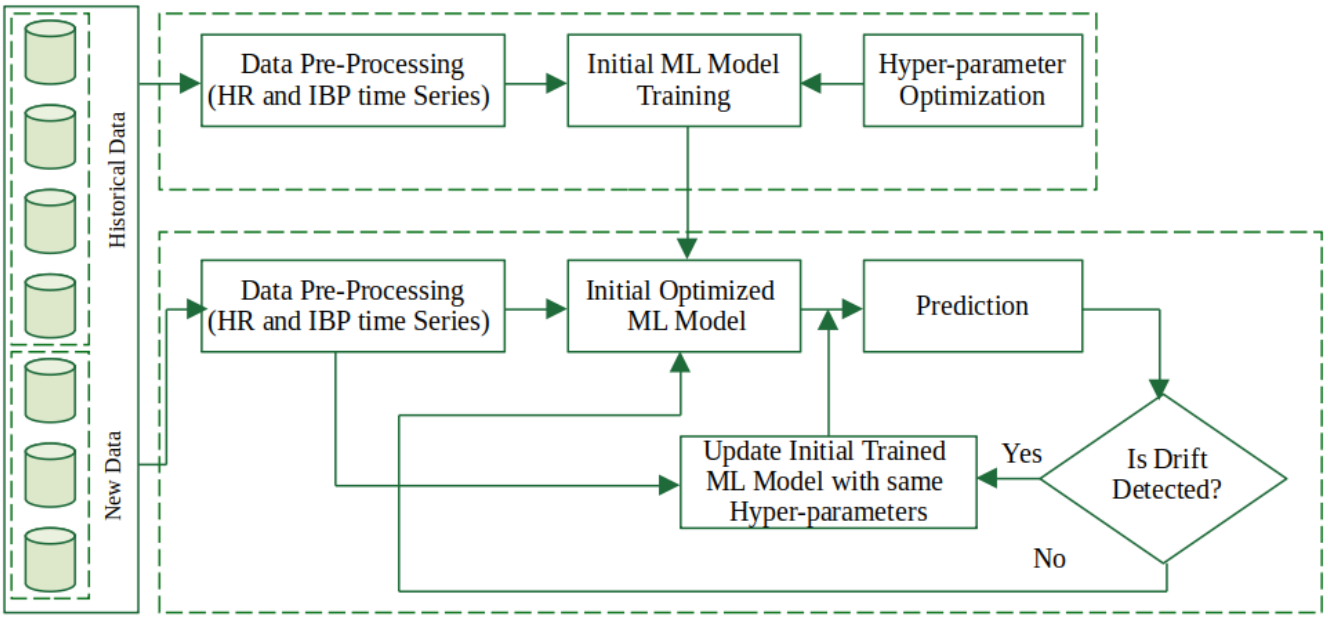

FIGURE 2 The overall methodology followed in Incremental Learning.

Figure 2 shows the architecture of an incremental learning based drift adaptive framework. In the initial phase, the available dataset instances undergo pre-processing and are used to train the regression ML model with optimal hyper-parameters, as described in Konar et al. [12]. In the next phase, the optimized ML model makes predictions on incoming data instances. After predicting the output values, drift detection techniques are employed to identify any concept drift in the forthcoming data samples. If no drift is detected, the initial ML model is used to make further predictions on the current data instances. However, when significant drift is detected, the ML model partially updates itself using either the warm start or partial fit method on the current data instances, which may differ from the original training data due to variation in the data distribution. During the last phase, the updated model is used to make predictions on subsequent datasets. The predictions are then compared to the actual values, and the model's accuracy is evaluated.

Therefore, by continuously updating the model and monitoring its performance, the system can adapt to changes in the underlying distribution, while still retaining information from the past.

4.3 | Ensemble Learning

To handle recurring concept drift, using ensemble learning can be an effective approach where multiple ML models with better generalization ability are combined. The core idea is to preserve and reuse old models, which can save significant effort in retraining a new model [15]. In ensemble learning, averaging, voting and stacked ensemble methods are the potential solutions for ML model updating. In averaging ensembles, the individual model's predictions are combined by taking the mean or weighted mean of their outputs. Voting ensembles combine the output of each base classifier by taking a majority or weighted vote of their outputs to predict newly arrived data. In contrast, stacked ensembles use the individual model's predictions as inputs to a meta-model, which learns to combine their outputs in a more efficient way [16,17].

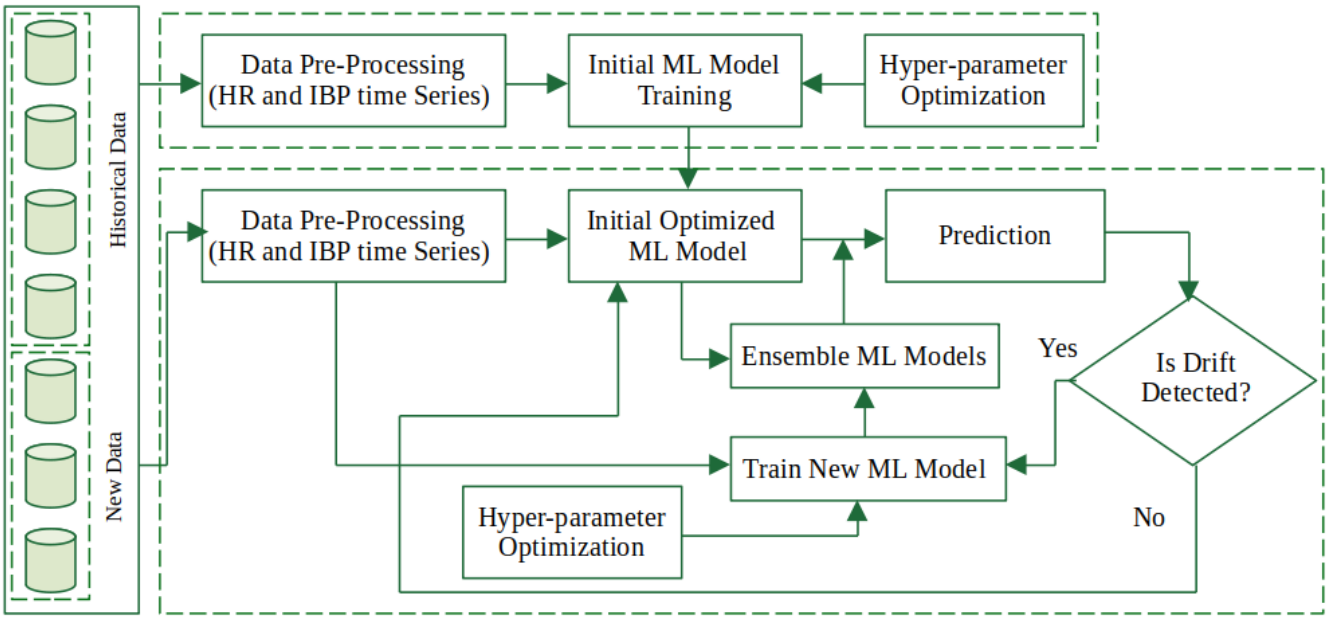

FIGURE 3 The overall methodology followed in Ensemble Learning.

Figure 3 demonstrates the overall architecture of the ensemble learning based drift adaptive framework. The initial stage involves gathering time series data from the present to generate a historical dataset. Following the collection of data, it is pre-processed and employed to train an initial machine learning model, which is subsequently included in the ensemble. Moreover, the hyper-parameters of the ML model are optimized, as described in Konar et al. [12]. In the next stage, when a new batch of datasets arrive, the initial model is used to predict the outcome values, and drift detection techniques are employed to detect any changes in the incoming batch of datasets. If significant drift is detected, a new ML model is trained on new concept datasets, and the system combines the selected models' outputs to generate an ensemble prediction using techniques such as averaging, voting, or stacking. The final stage consists of evaluating the performance of the ensemble. To achieve this, the evaluation of the ensemble model on the most recent batch of data is compared with that of the individual models.

Ensemble learning utilizes the advantages of multiple models, making it an effective method to adjust to new data patterns that constantly change, and to preserve accurate prediction results without having to completely retrain a model.

5 | RESULTS

In this section, the performance of various drift detection and drift adaptive techniques has been evaluated. For experimental evaluation, the time series datasets of patients undergoing cardiac surgery collected by the Anaesthesia Information Management System (AIMS) installed at the host institute (PGIMER Chandigarh) were used. As described in Konar et al. [11], we have used

Robustscaler for normalization, Gradient boosting regressor for hospital stay prediction, and XGBoost regressor for ICU stay prediction.

5.1 | Performance Metrics for Evaluation

$$MAE = \frac{1}{m}\sum_{i=1}^{m} \|X_i - Y_i\| \tag{1}$$

$$MSE = \frac{1}{m}\sum_{i=1}^{m} (X_i - Y_i)^2 \tag{2}$$

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m} (X_i - Y_i)^2} \tag{3}$$

$$R^2score = 1 - \frac{\sum_{i=1}^{m}(X_i - Y_i)^2}{\sum_{i=1}^{m}(\overline{Y} - Y_i)^2} \tag{4}$$

$$s\_MRE = \frac{n}{N} \tag{5}$$

$$d\_MAE = 1 - \frac{MAE}{\overline{Y}} \tag{6}$$

Here, X_i represents the predicted i^{th} value, Y_i represents the actual i^{th} value, m represents the total number of test samples, $\overline{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$ is the mean of actual values, n is the number of predicted values that satisfy $\frac{\|X_i - Y_i\|}{X_i} < \theta$, N is the number of elements in the test set, and θ is the relative change between actual and predicted value that can be accepted. In this paper, we have used $\theta = 0.2$.

5.2 | Performance Evaluation of Drift Detecting Methods

The various drift detection methods such as PSI, KS, WDM, CVM, MWT were used to identify concept drift in incoming batch of datasets. For experimental evaluation, the HR and IBP time series data of 945 patients undergoing cardiac surgeries were utilized as the historical dataset, while the batch of 100 new patients undergoing the same surgery was used as the test dataset. A virtual drift was induced in the test datasets by changing the data distribution of outcome values. Table 1 shows the comparative analysis of various statistical tests for detecting drift in both input features.

TABLE 1 Comparison of Drift Detection Techniques for input and output parameters

Drift Detection Methods	Values			
	ICU Stay	Hospital Stay	HR parameter	IBP parameter
Population Stability Index Test	0.3717	1.1301	0.2627	0.1145
Kolmogorov Smirnov Test	0.0944	4.0209 e^{-32}	0.7641	0.3734
Wasserstein Distance Metric	1.5654	6.2543	1.9971	1.2627
Cramer Von Mises Test	6.6206 e–10	3.3838 e^{-09}	0.0579	0.2203
Mann-Whitney U Test	0.8010	8.7810 e^{-31}	0.9318	0.1351

When comparing the performance of various statistical methods for detecting drift in a new batch of datasets, as displayed in Table 1, it was observed that most of the methods were able to detect the drift induced in the hospital stay parameter. The WDM,

MWU, and KS tests exhibited greater efficiency in drift detection, thus validating the stated hypothesis, in comparison to other statistical methods. However, the CVM test exhibited false drift detection in ICU stay, leading to inadequate performance in this particular dataset. This is likely due to its reliance on a large number of samples to accurately estimate the distribution, which could be challenging in real- world situations where data is limited.

5.3 | Performance Evaluation of Active Approaches for Drift Adaptive Framework

Table 2 exhibits the experimental results for the drift adaptive methods used for predicting hospital stay. These techniques were evaluated based on various performance evaluation metrics such as mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), and R2 score. When predicting the hospital stay on the latest batch of datasets, the voting ensemble technique achieved the lowest RMSE of 13.683, followed by stacked ensembles with 13.973, incremental learning with 18.446, and adaptive learning with 10.36. However, in terms of both MAE and R2-score, the voting ensemble was found to be the best technique for adapting drift, with values of 10.684 and 0.87, respectively. Therefore, the voting ensemble-based learning technique demonstrated the best performance in adapting concept drift, as indicated by the metrics of MSE, RMSE, MAE, and R2-score.

TABLE 2 Comparison of Drift Adaptive Techniques for Hospital Stay

Drift Adaptation Techniques	MAE	MSE	RMSE	R2_score
Adaptive Learning	14.208	374.91	19.36	-0.826
Incremental Learning	13.590	340.28	18.446	-0.657
Voting Ensemble Learning	10.684	187.24	13.683	0.87
Stacked Ensemble Learning	11.100	195.26	13.973	0.48

Table 3 presents the experimental results for the drift adaptive methods used for predicting ICU stay. Different performance evaluation metrics, including MAE, MSE, RMSE, and R2 score, were used to assess the effectiveness of these techniques. When predicting ICU stay on the latest batch of datasets, the voting ensemble technique achieved the lowest RMSE of 4.546, followed by stacked ensembles with 4.658, incremental learning with 6.111, and adaptive learning with 6.681. However, in terms of both MAE and R2-score, the voting ensemble was found to be the best technique for adapting drift, with values of 2.827 and 0.042, respectively. Therefore, the voting ensemble-based learning technique demonstrated the best performance in adapting concept drift, as indicated by the metrics of MSE, RMSE, MAE, and R2-score.

TABLE 3 Comparison of Drift Adaptive Techniques for ICU Stay

Drift Adaptation Techniques	MAE	MSE	RMSE	R2_score
Adaptation Learning	3.721	44.64	6.681	-0.285
Incremental Learning	3.622	37.351	6.111	-0.075
Voting Ensemble Learning	2.827	20.66	4.546	0.042
Stacked Ensemble Learning	2.874	21.69	4.658	0.075

When comparing the performance of all drift adaptive techniques for hospital stay and ICU stay analysis, as displayed in Table 2 and Table 3 respectively, it was observed that the ensemble learning technique was the most effective and accurate in adapting

to concept drift. This is because the weighted voting approach of the ensemble allowed for reusing the old model to predict values from a familiar environment by adjusting the weights of the ensemble members. Additionally, if old regressor ML models become relevant again in the future, they can be re-weighted based on their predicted expertise in the current environment. However, other techniques ignore previous data and only learn from the current environment, which restricts their capability to handle substantial concept drift. As a result, adaptive and incremental learning methods may not perform as well in cases of rigorous concept drift.

5.4 | Performance Evaluation of ML Models with and without Drift Adaptive Technique

Table 4 presents the experimental results of regression models to predict the Hospital stay for both 'with' and 'without drift' adaptation techniques. The model's performance was evaluated using various metrics such as MAE, MSE, RMSE, R2 score, s_MAE, and d_MAE. While evaluating using all regression models, it was observed that the performance of prediction system has been significantly improved by incorporating ensemble based drift adaptive technique. However, for predicting Hospital stay, the lowest RMSE was achieved using GB regressor (13.683), followed by XGB regressor (14.612), and random forest (RF) (17.514). When considering other evaluation metrics including MAE, R2-score, s_MRE(%), and d_MAE(%), the GB regressor was the best model for the analysis of Hospital stay, with values of 10.684, 0.87, 80%, and 7.13% respectively.

TABLE 4 Performance Evaluation of Hospital Stay with and without Drift Adaptation Technique

Evaluation Metrics	Without Drift Adaptation			With Drift Adaptation		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
MAE	12.360	13.247	12.927	12.463	11.523	10.684
MSE	21710	545.87	374.75	306.744	213.523	187.24
RMSE	147.34	23.363	19.358	17.514	14.612	13.683
R2-Score	-4.149	-1.034	-0.635	-0.494	-0.040	0.87
s_MRE(%)	0.00	6.67	6.67	46.67	53.33	80.00
d_MAE(%)	31.67	37.62	36.64	78.22	82.17	87.13

Model 1=Random Forest, Model 2=XGBRegressor, Model 3=Gradient Boosting Regressor

Table 5 presents the experimental results of regression models to predict the ICU stay for both 'with' and 'without drift' adaptation techniques. The model's performance was evaluated using various metrics such as MAE, MSE, RMSE, R2 score, s_MAE, and d_MAE. When evaluating using all regression models, it was observed that the performance of prediction system has been significantly improved by incorporating ensemble based drift adaptive technique. However, for predicting ICU stay, the lowest RMSE was achieved using XGB regressor (4.546), followed by GB regressor (5.205)), and RF (5.758). When considering other evaluation metrics including MAE, R2-score, s_MRE(%), and d_MAE(%), the XGB regressor was the best model for the analysis of ICU stay, with values of 2.874, 0.042, 73.33%, and 87.73 % respectively.

5.5 | Performance Evaluation of Models with and without drift using F-test

When fitting data using nonlinear regression, there are often times when one must choose between two models that both appear to fit the data well. After looking at the R2 scores for each model, both models may appear to fit the data. In this case, an F-test can be conducted to see which model is statistically better. It gives a definitive answer and does not rely on arbitrary interpretation of

TABLE 5 Performance Evaluation of ICU Stay with and without Drift Adaptation Technique

Evaluation Metrics	Without Drift Adaptation			With Drift Adaptation		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
MAE	11.610	4.764	3.856	3.278	2.874	3.0278
MSE	162.96	125.37	34.194	33.16	20.66	27.09
RMSE	12.765	11.197	5.847	5.758	4.546	5.205
R2-Score	-4.799	-3.461	-0.216	0.44	0.042	0.219
s_MRE(%)	26.67	46.67	40.00	60.00	73.33	60.00
d_MAE(%)	40.52	78.94	70.26	84.21	87.73	85.97

Model 1= Random Forest, Model 2= XGBRegressor, Model 3= Gradient Boosting Regressor

an R2 score or any other error metrics. If both models have the same number of parameters, the formula for the F statistic is

$$F = \frac{RSS1}{RSS2} \tag{7}$$

Here, RSS1 is the residual sum of squares for the model without drift, RSS2 is the residual sum of squares for the model with drift.

Also, there are N-V degrees of freedom, where N is the number of data points, and V is the number of parameters being estimated. A p-value less than $\alpha = 0.05$ indicates that the model with drift (denominator of F-statistic) fits the data significantly better than the model without drift.

Table 6 presents the calculated F-value and p-value for comparing models with and without drift for predicting hospital stay. Gradient Boosting Regression has a slightly lower F-value compared to the other two models, indicating a relatively smaller difference in performance between the cases of with and without drift.

TABLE 6 Comparison for ML Models without and with Drift for Hospital Stay

ML Models	Calculated F-value	p-value
Model 1	12.02	$2.926839e^{-242}$
Model2	12.78	$3.505184e^{-252}$
Model 3	9.62	$5.076535e^{-207}$

Model 1= Random Forest, Model 2= XGBRegressor, Model 3= Gradient Boosting Regressor

TABLE 7 Comparison for ML Models without and with Drift for ICUl Stay

ML Models	Calculated F-value	p-value
Model 1	21.11	< 0.000001
Model2	4.30	$1.0723e^{-95}$
Model 3	2.57	$3.581843e^{-43}$

Model 1= Random Forest, Model 2= XGBRegressor, Model 3= Gradient Boosting Regressor

Table 7 presents the calculated F-value and p-value for comparing model with and without drift for predicting ICU stay. In comparison to the other two models, XGB Regressor exhibits a lower F-value, implying a relatively modest variance in performance when considering the scenarios with and without drift.

6 | DISCUSSION

Cardiac surgery risk prediction models have been in existence since 1970s[18]. The risk scores are usually based on regression equations. They have been regularly updated to reflect the heterogeneity in patient population as well as the improvement in care. In fact, the Society of Thoracic Surgeons risk score is updated every 3 months. This need for regular updation needs to be recognized when applying a machine learning model to risk prediction.

When machine learning becomes increasingly applied to clinical practice, causality, and missingness can alter the performance of prediction models[18]. Causality is attribution of the occurrence of an outcome to a particular parameter in the input dataset. The parameter could be an intervention or an observational variable. While it is possible that the intervention may have led to the outcome, it is also possible that the intervention was included as a protective measure to prevent the outcome. Therefore, the intervention will be more often included when the physician recognizes the risk of the occurrence of an outcome. This misinterpretation of causality could be amplified when a static model is used for prediction. This is possible in our present application too, in the prediction of outcome and stay duration.

For example, a sick patient will have higher variability in the IBP(M) since the blood pressure often decreases and the physician intervenes to increase the blood pressure frequently. Here, the variability is an association with poor outcomes. Technology is being developed that could help reduce the variability in blood pressure[19]. Suppose, the use of this technology reduces the variability in blood pressure, the static model could fail to identify a sick patient. On the other hand, the inclusion of newer patients in the model development through adaptive learning could account for the incorporation of newer medical technology.

Moreover, healthcare systems are complex with multiple factors affecting outcome. The factors could also be outside the purview of an individual patient attribute. These factors are also dynamic and include staffing, resource availability, technical skills, practice variability and newer medications. For example, the concept of fast-tracking in cardiac surgery includes multiple interventions before, during and after the cardiac surgery to reduce the ICU stay and hospital stay of patients[20]. When implemented, fast-track protocols can decrease the stay duration of the patients. Moreover, not all patients are eligible for fast-track protocols. Therefore, the data distribution of stay duration will also change following fast-track protocol implementation. Again, in this case, use of static models would be inaccurate in predicting the stay duration. If the machine learning models are updated using the data from patients after fast-tracking, it could better predict the stay durations.

Another important challenge is data missingness[18].With a change in clinical practice, some investigations or monitoring parameters could become irrelevant. Therefore, the data required for the model could be missing. One example is in the procedure of coronary artery bypass grafting. With increased experience, the surgeon could remove the aortic cross clamp at a later stage such that the difference in time and the amount of data between the aortic cross clamp off time and cardiopulmonary bypass off time could become lesser. This necessitates retraining the model to the new distribution of data.

In this study, different drift detection techniques, including PSI, KS, WDM, CVM, and MWT, were employed to detect instances of concept drift within nonlinear time series data. By referring to Table 1, we observe that the choice of the most appropriate drift detection technique depends on the desired sensitivity to drift, the selection of the most appropriate drift detection method depends on the research context, the desired level of sensitivity, and the specific characteristics of the data under investigation.

In the realm of medical data, accurately predicting outcomes like ICU and hospital stays following cardiac surgeries becomes essential. The study underscores the importance of adapting to concept drift in such scenarios. To address the issue of drift adaptation, three dynamic strategies—adaptive learning, incremental learning, and ensemble learning techniques have been utilized. Adaptive learning involves retraining models on modified datasets after drift detection, while incremental learning continually updates models as new data arrives. Ensemble learning, on the other hand, combines models to handle recurring drift and better adapt to changing data patterns. By referring to Table 2 and 3, we observe that Voting Ensemble Learning and Stacked Ensemble Learning demonstrate better adaptation to drift and offer improved predictive performance for hospital stay and ICU stay duration compared to Adaptive Learning and Incremental Learning.

In summary, incorporating the Drift Adaptation Technique leads to improvements in prediction accuracy across various evaluation metrics for all three models (Random Forest, XGBRegressor, and Gradient Boosting Regressor), as shown in Table 4 and 5. The models with drift adaptation consistently outperform their counterparts in terms of MAE, MSE, RMSE, R2-Score,

s_MRE, and d_MAE, indicating the effectiveness of the drift adaptation approach in enhancing ICU stay and hospital stay prediction.

The F-values and p-values obtained from all three models, as shown in Tables 6 and 7, indicate that accounting for drift significantly influences how well each model predicts hospital stay and ICU stay. The very low p-values highlight a high level of statistical importance, further confirming that the differences in observed performance are not random occurrences.

7 | CONCLUSION

This research emphasizes the use of different frameworks for adapting to concept drift in nonlinear time series data for predictive maintenance. The experimental results demonstrate the effectiveness of using statistical tests to detect concept drift. The comparative analysis indicates that dynamic ensemble learning is a more efficient and accurate method for adapting to concept drift.

AUTHOR CONTRIBUTIONS

SK designed the study, performed data preprocessing and revised and approved the manuscript. AM collected the data and performed model evaluation. TK and MS drafted the manuscript. NA designed the study, revised and approved the manuscript. TS and RG analyzed the data, revised and approved the manuscript. GDP conceived the study, revised and approved the manuscript. TK and MS performed training and testing, drafted and approved the manuscript. TS and SKST interpreted the data, revised and approved the manuscript.

ACKNOWLEDGMENTS

The authors wish to acknowledge the Department of Science and Technology (DST) Government of India, the National Supercomputing Mission (NSM), and the Indian Council of Medical Research (ICMR) for providing funding that supported this research.

FINANCIAL DISCLOSURE

This work is supported by a research grant from the Department of Science and Technology (DST) under the National Supercomputing Mission (NSM) via letter no. DST/NSM/R&D-HPC-Application/2021/03.02, dated 23rd March 2021. The work in the present article was also supported by funding provided by Indian Council of Medical Research (ICMR), New Delhi via letter no. 5/3/8/33/ITR-F/2020-ITR dated 26–10-2020.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

REFERENCES

1. Lu J, Liu A, Dong F, Gu F, Gama J, Zhang G. Learning under concept drift: A review. *IEEE Transactions on Knowledge and Data Engineering.* 2018;31(12):2346-2363.
2. Lewis G, Echeverría S, Pons L, Chrabaszcz J. Augur: A Step Towards Realistic Drift Detection in Production ML Systems. In: 2022:37-44.
3. Fields T, Hsieh G, Chenou J. Mitigating drift in time series data with noise augmentation. In: IEEE. 2019:227-230.
4. Gama J, Žliobaitė I, Bifet A, Pechenizkiy M, Bouchachia A. A survey on concept drift adaptation. *ACM Computing Surveys (CSUR).* 2014;46(4):1-37.
5. Lin C, Deng D, Kuo C, Chen L. Concept drift detection and adaption in big imbalance industrial IoT data using an ensemble learning method of offline classifiers. *IEEE Access.* 2019;7:56198-56207.
6. Bayram F, Ahmed B, Kassler A. From concept drift to model degradation: An overview on performance-aware drift detectors. *Knowledge-Based Systems.* 2022:108632.
7. Yu S, Abraham Z, Wang H, Shah M, Wei Y, Príncipe J. Concept drift detection and adaptation with hierarchical hypothesis testing. *Journal of the Franklin Institute.* 2019;356(5):3187-3215.
8. Escovedo T, Koshiyama A, Cruz dAA, Vellasco M. DetectA: abrupt concept drift detection in non-stationary environments. *Applied Soft Computing.* 2018;62:119–133.
9. Yang L, Shami A. A lightweight concept drift detection and adaptation framework for IoT data streams. *IEEE Internet of Things Magazine.* 2021;4(2):96-101.
10. Agrahari S, Singh A. Concept drift detection in data stream mining: A literature review. *Journal of King Saud University-Computer and Information Sciences.*
11. Celik B, Vanschoren J. Adaptation strategies for automated machine learning on evolving data. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2021;43(9):3067-3078.
12. Konar S, Auluck N, Ganesan R, et al. A non-linear time series based artificial intelligence model to predict outcome in cardiac surgery. *Health and Technology.* 2022:1-13.

13. Zang W, Zhang P, Zhou C, Guo L. Comparative study between incremental and ensemble learning on data streams: Case study. *Journal of Big Data.* 2014;1(1):1-16.
14. Yuan X, Wang R, Zhuang Y, Zhu K, Hao J. A concept drift based ensemble incremental learning approach for intrusion detection. In: IEEE. 2018:350-357.
15. Mera C, Orozco-Alzate M, Branch J. Incremental learning of concept drift in Multiple Instance Learning for industrial visual inspection. *Computers in Industry.* 2019;109:153-164.
16. Yang Z, Al-Dahidi S, Baraldi P, Zio E, Montelatici L. A novel concept drift detection method for incremental learning in nonstationary environments. *IEEE Transactions on Neural Networks and Learning Systems.* 2019;31(1):309-320.
17. Mulimani D, Totad S, Patil P, Seeri S. Adaptive Ensemble Learning with Concept Drift Detection for Intrusion Detection. In: Springer Singapore. 2021:331-339.
18. Ghassemi M, Naumann T, Schulam P, Beam AL, Chen IY, Ranganath R. A review of challenges and opportunities in machine learning for health. *AMIA Summits on Translational Science Proceedings.* 2020;2020:191.
19. Kumar S, Puri GD, Mathew PJ, Mandal B. Evaluation of indigenously developed closed-loop automated blood pressure control system (claps): a preliminary study. *Journal of Clinical Monitoring and Computing.* 2022;36(6):1657–1665.
20. Wong WT, Lai VK, Chee YE, Lee A. Fast-track cardiac care for adult cardiac surgical patients. *Cochrane Database of Systematic Reviews.* 2016(9).
21. Pittams AP, Iddawela S, Zaidi S, Tyson N, Harky A. Scoring systems for risk stratification in patients undergoing cardiac surgery. *Journal of cardiothoracic and vascular anesthesia.* 2022;36(4):1148–1156.