Research article

A study on quality control using delta data with machine learning technique

Yufang Liang [a,1], Zhe Wang [b,1], Dawei Huang [a,c], Wei Wang [d], Xiang Feng [b], Zewen Han [b], Biao Song [b], Qingtao Wang [a,e,*], Rui Zhou [a,e,**]

[a] Department of Laboratory Medicine, Beijing Chao-yang Hospital, Capital Medical University, Beijing, PR China
[b] Inner Mongolia Wesure Date Technology Co., Ltd, Inner Mongolia, PR China
[c] Department of Laboratory Medicine, Beijing Longfu Hospital, Beijing, PR China
[d] Department of Blood Transfusion, Beijing Ditan Hospital, Capital Medical University, Beijing, PR China
[e] Beijing Center for Clinical Laboratories, Beijing, PR China

HIGHLIGHTS

• A protocol for data processing by using delta data together with machine learning algorithm, enables to improve data stability.
• After data processing, the performance of QC event prediction surpassed over 50% clinical recognized PBRTQC method, especially for the hard-to-detect error in QC event prediction.

ARTICLE INFO

ABSTRACT

Background: In the big data era, patient-based real-time quality control (PBRTQC), as an emerging quality control (QC) method, is expanding within the clinical laboratory industry. However, the main issue of current PBRTQC methodology is data stability. Our study is aimed to explore a novel protocol for data stability by combining delta data with machine learning (ML) technique to improve the capacity of QC event detection.
Methods: A data set of 423,290 laboratory results from Beijing Chao-yang Hospital 2019 patient results were used as a training set (n = 380960, 90%) and internal validation set (n = 42330, 10%). A further 22,460 results from Beijing Long-fu Hospital 2019 patient results were used as a test set. Three-type data (1) Single-type data processed by truncation limits; (2) delta-type data processed by truncation limits and (3)delta-type data processed by Isolated Forest (IF) algorithm were evaluated with accuracy, sensitivity, NPed, etc., and compared with previously published statistical methods.
Results: The optimal model was based on Random Forest (RF) algorithm by using delta-type data processed by IF algorithm. The model had a better accuracy (0.99), sensitivity (0.99) specificity (0.99) and AUC (0.99) with the dependent test set, surpassing the critical bias of PBRTQC by over 50%. For the LYMPH#, HGB, and PLT, the cumulative MNPed of MLQC were reduced by 95.43%, 97.39%, and 97.97% respectively when compared to the best of the PBRTQC.
Conclusion: Final results indicate that by integrating an innovative ML algorithm with the overall data processing protocol the detection of QC events is improved.

1. Introduction

Laboratory test results play a crucial role in disease screening, diagnosis, prognosis evaluation, treatment monitoring. Traditional QC, due to non-commutable control materials measured, is not treated as real patient samples [1, 2], leading to lower error detection and higher false alarms. PBRTQC, as a newly QC method, because of real testing results used for QC, the commutability issue reduces, however, data stability is outstandingly cumbersome. Several studies have reported methods for data stability. Xincen Duan et al. [3] used the residual of the regression

model as the input for improving univariate statistical process control (SPC) algorithms. Ng et al [4]. developed subpopulation protocols for hospitalized and ambulatory patients by adding additional clinical information to improve method performance. Ichihara et al. [5]. set up the weighted cumulative delta-check (wCDI) method in steps of data transformation, data standardization and index adjustment, for data pre-processing used for detecting specimen mix-up. Cembrowski GS, et al. [3, 6] combined delta check (DC) and moving average (MA) algorithms, developed an average of deltas (AoD) strategy, used for monitoring the mean delta of consecutive, intra-patient results to detect systematic error. In their protocol, a simulated annealing algorithm was used to select the number of patient delta values to calculate the average delta and to determine truncation limits to eliminate the effect from large deltas.

DC, a laboratory information system (LIS)-based quality tool, involves the calculation and evaluation of sequential patients differences, to detect errors derived from total testing process (TTP). The increase of delta values may mainly arise from intra-patient variation, from analytical variation (e.g., instrument) or other out-of analytical variation [7] Previous studies found that delta check was used for error detection, its capability was limited, exclusively suited to detect larger analytic errors or mislabeled specimens [8, 9, 10, 11].

Machine learning (ML), one branch of Artificial Intelligence (AI), is prevailing in various application fields in recent years. It enables users to construct a related learning system in accordance with specific tasks. To be exact, ML can learn useful information from the unknown characteristics of a process and environment, and apply the learned information to develop a prediction, a classification and inform the decision-making process for new unknown problem in the future. ML differs from traditional statistical methods in that ML can cross industry boundaries and summarize solutions to problems that cannot be resolved by simple functions. Our study is aimed to, using delta data, combined with ML technique, explore a new approach for data stability, thus improving the capacity of QC event detection. Our study is aimed to assess explore a novel protocol for data stability by combining delta data with machine learning (ML) technique to improve the capacity of QC event detection. To verify the improvement effect of data stability, we compared both delta-type data and single-type data which were processed by truncation limits in PBRTQC based on statistical method, and further compared delta-type data processed by Isolated Forest in ML and in PBRTQC.

2. Materials and methods

2.1. Data collection

Our data was divided into training set, validation set and test set. The validation set can be understood as a part of the training set used for monitoring the model training process. All data were obtained from two Beijing's Hospitals. 423290 results measured on XN-9000 (Sysmex, Kobe, Japan) were extracted through the laboratory information system (LIS) of Beijing Chao-yang Hospital in 2018, of which the data in the first 10 months were used for model training and the latter two months were used for model validation. 22460 results measured on BC-5390 (Mindray, Shenzhen, China) in the same time interval were obtained from another hospital, Beijing Long-fu Hospital for model testing. The data were filtered by rules, including: (1) patients with only one result in the study interval were excluded; (2) according to Tukey's standard [12] the values less than the overall 25% quantile or greater than the overall 75% quantile were removed as outliers; (3) after Tukey's standard was applied, patients with less than two results were excluded; (4) that the age was 14–60 years old was included; (5) and delta check interval was defined as one year [13]. All data were collected in-control status. Only seven representative test items were selected, including lymphocyte count (lymph#), lymphocyte ratio (lymph%), hemoglobin volume (HGB), mean hemoglobin volume (MCH), mean hemoglobin concentration (MCHC), red blood cell volume distribution width (R–CV) and

platelet count (PLT). The reason for selecting these test items was that they represented different degrees of variation in leukocyte, erythrocyte and platelet series, and the degree of variation was based on their respective GVi/CVg rates. For a pair of results for each patient, one was used for obtaining delta value by calculating the difference between the pair of results in experiment 2 and 3, the other for selecting the second result each patient in experiment 1. Experiment 1–3 were described as followed.

2.2. Data stimulation

The data filtered by rules were regarded as unbiased data, in order to simulate out-of-control status in the real setting, we artificially introduced 10 biases of different sizes according to the formula below:

$$x' = x + n \times TEa \times \bar{x}$$

where n refers to different multiples ($-3/2$, -1, -1/2, -1/4, -1/6, 1/6, 1/4, 1/2, 1, 3/2), $\bar{x}$ represents the mean of all data for each test item, TEa represents the total allowable error for each test item which is defined as the sum of random error (RE) and systematic error (SE). For delta and single data sets, the simulation methods of biased data were different: 1) for the delta data set, a goal bias was only introduced into the second result of each patient, and then the biased delta was obtained by calculating a difference with the first result of this pair of patient results; 2) for the single data set, a goal bias was directly introduced on unbiased data as biased data. In this paper, a bias represented a shift in the mean.

2.3. MLQC based on delta-type data

2.3.1. Data pre-processing

Isolation Forest (IF) algorithm was used for pre-processing of delta data. IF is an unsupervised anomaly detection method commonly used for continuous numerical data [14] Its principle is to set up multiple isolation trees (iTree), each of which belongs to a binary tree structure to cut in the data space. As usual, the probability of data occurrence is extremely rare in the area with sparse spatial distribution, thus if the data falling in these areas can be considered as an abnormal value [15].

2.3.2. Model construction

Unbiased data and biased data in our study were defined as two kinds of data in ML, and they were distinguished by Random Forest (RF). RF classifier, an integrated supervised learning, is consists of multiple decision trees, each decision tree will give its own prediction, and the final prediction is given by way of voting [14].

Patient data was partitioned as a block size. A block size was taken as a whole. which was called a "machine learning sample". The input feature of RF algorithm model was a sequence composed of multiple values. If the step size was 10, it indicated that the input dimension was 10. This not only introduced the serialization feature, but also met the multi feature requirements of machine learning algorithm. The following described the process of determining the step size and RF algorithm parameters:

Firstly, take an example of one test item, when RF algorithm parameters in the program were set as default values, a block size needed to be pre-defined. A traversal experiment in a range of 5–20 in step of 1 was carried on to determine a proper block size. Secondly, all data included were partitioned by the block size determined, then a new "ML sample" was formed. Thirdly, ML samples were standardized. The samples composed of unbiased data was labeled "0″ and the samples composed of biased data as "1". The same experimental steps above were implemented as unbiased data and biased data. In the following, the RF algorithm parameters were optimized by adjusting the numbers of-trees (N-trees) and max depth of trees (Max Depth). The search scope for N-trees from 100 to 500 in step of 100, and for Max Depth from 50 to 300 in
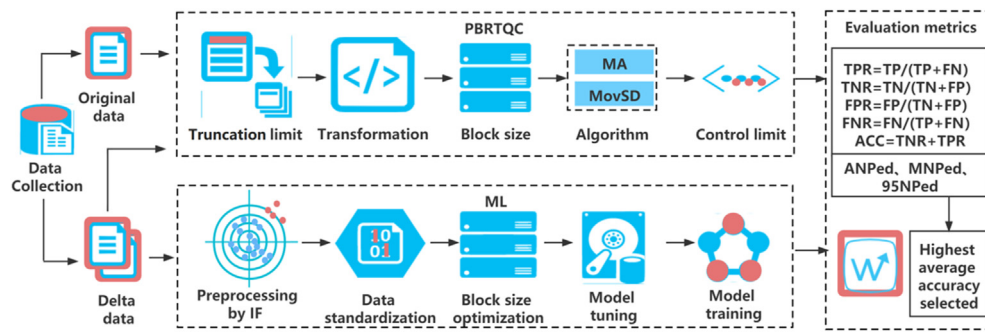
Figure 1. Integrated experimental process diagram.

step of 50. The accuracy of training and testing results for each combination was analyzed, to select the best parameter combination.

2.4. PBRTQC based on statistics

Two PBRTQC algorithms, moving average (MA) and moving standard deviation (MovSD) by only using single-type patient data, were selected as comparative methods. Then, both PBRTQC by using delta-type patient data and ML QC using delta-type patient data were regarded as comparing methods, one to estimate the ability of QC event prediction by using single-type data and delta-type data separately, the other to compare the efficiency of data filtering by the way of transaction limits recommended by The International Federation of Clinical Chemistry and Laboratory (IFCC) [16, 17] and by ML IF algorithm.

For PBRTQC, the optimization of parameters strictly followed the experimental process recommended by IFCC. In the first step, the input data for PBRTQC was filtered by truncation limits, like data filtering of IF in ML, the minimum and maximum values were removed after sorting the original data. In order to reduce the influence of noise caused by extreme values, six data truncation limits were explored (0%, 1%, 5%,15%, 20% and 40%); then the data filtered were dealt with or without Box-Cox transformation.

2.4.1. PBRTQC by using single-type data

For PBRTQC by using single-type patient data, the data were firstly filtered by reference intervals of test items. Both MA and MovSD [18]algorithms were experimented. The optimal range of block size was selected from 10, 30, 50, 90, 110, 130 to 150. the mean and standard deviation (SD) of the two PBRTQC algorithms were calculated for each block size.

All patient results were divided into 20 virtual days according to the original time sequence, and 1150 data were allocated every day, including the first 150 unbiased data and the last 1000 biased data. The control limits based on three calculation methods provided by IFCC, namely: symmetric, all PBRTQC and daily extremes were calculated for each test item [16]. For the symmetric method, we choose two distances: 2.5-time coefficient variance (CV) (equivalent to 2.5-time SD of PBRTQC results, 3-time CV (equivalent to 3-time SD of PBRTQC results). All parameters above mentioned were combined and experimented. The best combined results were obtained.

2.4.2. PBRTQC by using delta-type data

To ensure the comparability with PBRTQC based on single-type patient data, the same experimental steps as method 4.1 were followed. The only change in this experiment was that the input data was replaced by delta-type patient data.

Table 1. Data analysis of the 3 data types for the seven test items.

Test item	Algorithm	Mean	SD	Min	25th	50th	75th	Max
LYMPH#	Single	2.0722	0.9855	0.2400	1.4400	1.8600	2.4500	9.4700
	Delta	-0.0138	0.7819	-3.6900	-0.3600	-0.0200	0.2700	11.0600
	IF	-0.0076	0.3249	-0.6800	-0.2600	0.0000	0.2500	0.6000
LYMPH%	Single	31.3645	11.7726	2.9000	23.3000	30.2000	37.3000	77.8000
	Delta	-1.8032	8.8492	-40.9000	-6.5000	-1.5000	3.3000	34.2000
	IF	-0.2161	5.2249	-10.6600	-4.4000	-0.1600	4.0000	9.5600
HGB	Single	129.6815	19.5626	56.0000	118.0000	131.0000	143.0000	184.0000
	Delta	2.0162	11.2371	-64.0000	-4.0000	2.0000	7.0000	68.0000
	IF	0.8813	5.3897	-10.4000	-4.0000	1.0000	5.0000	11.6000
MCH	Single	29.5675	2.7095	17.9000	28.3000	29.8000	31.2000	39.8000
	Delta	0.0941	0.9510	-7.3000	-0.3000	0.1000	0.5000	7.3000
	IF	-0.0140	0.4154	-0.8600	-0.3000	0.0000	0.3000	0.8000
MCHC	Single	325.8178	13.1967	268.0000	318.0000	326.0000	335.0000	364.0000
	Delta	0.8596	12.1156	-48.0000	-7.0000	1.0000	7.0000	75.0000
	IF	1.0480	4.3566	-9.5000	-3.0000	1.0000	5.0000	9.2000
RCV	Single	13.5693	2.1297	10.9200	12.3000	13.0400	14.0700	30.4000
	Delta	-0.0211	1.4696	-10.9000	-0.4900	0.0000	0.4000	18.1600
	IF	0.0243	0.2560	-0.4000	-0.2000	0.0000	0.2000	0.5300
PLT	Single	235.4359	82.7727	30.0000	182.0000	224.0000	275.0000	742.0000
	Delta	2.8489	57.2888	-243.0000	-21.0000	4.0000	27.0000	456.0000
	IF	1.4151	23.3346	-42.0000	-17.0000	1.0000	20.0000	48.0000

Single - single-type data; Delta - delta-type data pre-processed by different truncation limits based on statistical method; IF - delta-type data pre-processed by IF based on ML method; Mean - average value; SD - standard deviation; Min - minimum value; **25th** - 25th quartile; **50th** - 50th quartile; **75th** - 75th quartile; Max - maximum value.

2.5. Evaluation metrics

Four process indicators: true positive (TP), true negative (TN), false positive (FP) and false negative (FN), were recorded, and then five indicators in the confusion matrix commonly used were evaluated the performance of our model, namely: area under ROC curve (AUC), true positive rate (TPR), true negative rate (TNR), false positive rate (FPR), false negative rate (FNR), and accuracy (ACC), here TPR was equivalent to sensitivity, TNR was equivalent to specificity.

When FPR <5%, the number of patients affected (NPed) from the beginning of a bias introduced to the bias detected was used to evaluate the clinical performance of PBRTQC and MLQC. The mean (ANPed), median (MNPed) and 95% quantile (95 NPed) of NPed on 20 virtual days were used as clinical performance indicators. The minimum value of accumulative MNPed was the optimal result.

Data processing and model analysis were performed in Python 3.7.3 package. The model training process depends on "numpy", "Pandas" and other tool kits. Figure 1 showed the integrated experimental process diagram.

3. Results

3.1. Data description

For the seven test items included, three different data pre-processing methods were used. As seen in Table 1, when single-type data was converted to delta-type data, the mean and SD of the data narrowed significantly, and the concentration and stability of the data were also increased. In terms of the values adopting different quartiles, the distribution was more uniform to the results of delta-type data than to that of single-type data, although the sign introduced by delta-type data increased the value threshold for some test items. Further, the data was more concentrated when filtered by IF based on delta-type data. Taking LYMPH# as an example, the SD of delta-type data after dealt with by IF was reduced by 67.03% compared with the single-type data processed by truncation limits, and the reduction rates of the remaining six test items were 55.62%, 72.45%, 84.67%, 66.99%, 87.98%, and 71.81%, respectively. To visualize of data distribution characteristics of the three types of data, principal component analysis (PCA) technique was
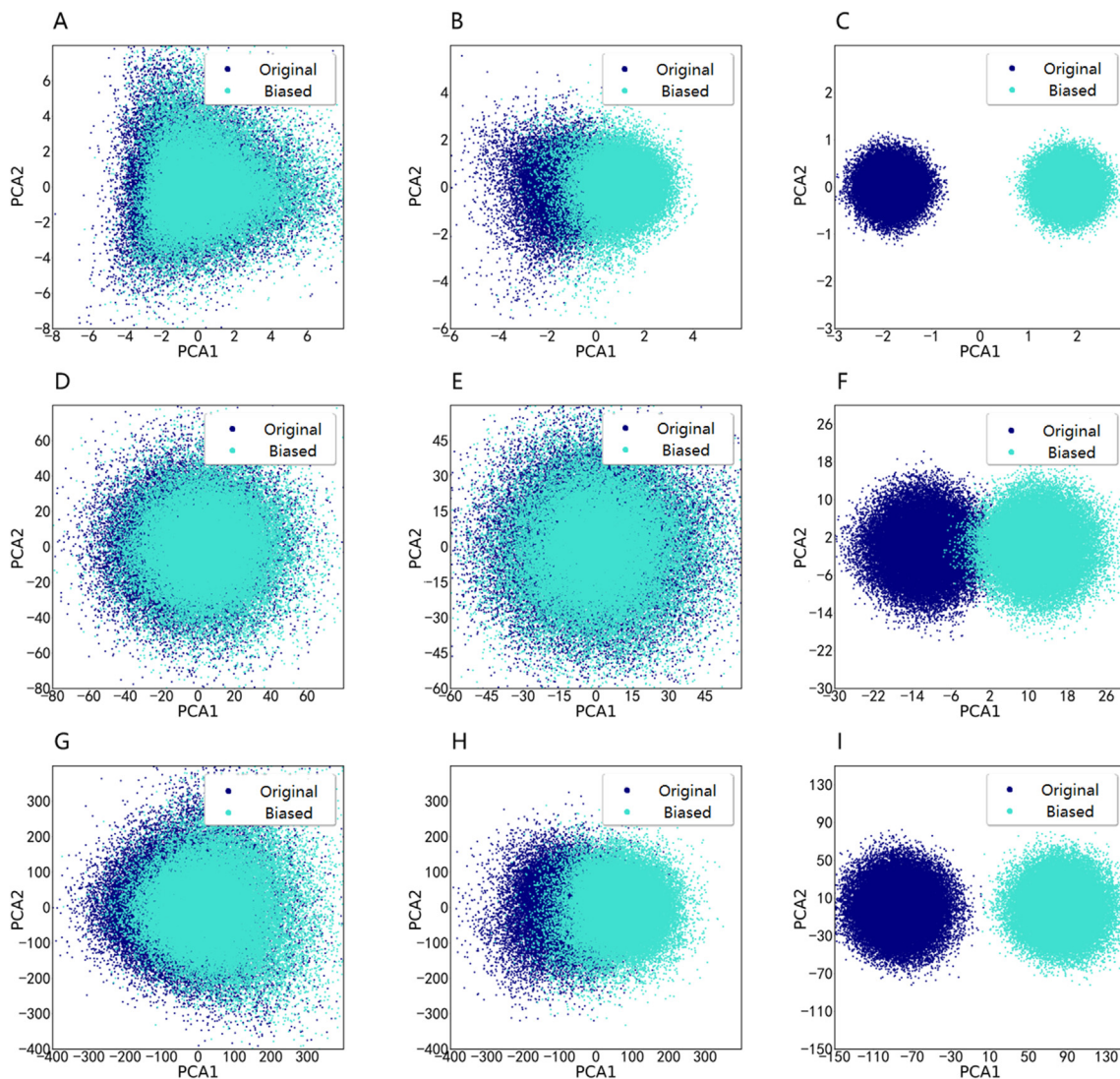

Figure 2. Data separability between critical biased and unbiased data for three-type data by PCA. The 3 rows from top to bottom represented LYMPH #, HGB and PLT of Cell Blood Count, the 3 columns from left to right represented single-type data, delta-type delta, and delta-type data processed by IF. Every point in each diagram represented a ML sample with the same block size consisting of 10 patient raw data.

Table 2. AUC for different block sizes and the results of RF tuning for R–CV at the critical bias.

Block size	AUC	N-trees	Max Depth	Training accuracy	Testing accuracy
5	0.9122	100	150	0.90	0.89
6	0.9352	200	100	0.91	0.92
7	0.9449	200	300	0.92	0.93
8	0.9556	300	100	0.96	0.94
9	0.9768	300	300	0.93	0.93
10	0.9862	400	100	0.94	0.93
11	0.9841	400	300	0.94	0.93
12	0.9832	500	100	0.94	0.93
13	0.9865	500	300	0.95	0.92

N-trees - number of trees; Max Depth - max depth of trees.

used to show the difference between biased data at the critical level and unbiased data, when a block size was set as 10. In Figure 2, the 3 rows from left to right representing LYMPH #, HGB and PLT, and the 3 columns from top to bottom representing single data, delta delta, and delta by IF. Every point in each diagram represented a ML sample consisting of 10 patient data as a block size. The results proved that the separability increased when the single-type data sequentially was processed by delta and then IF.
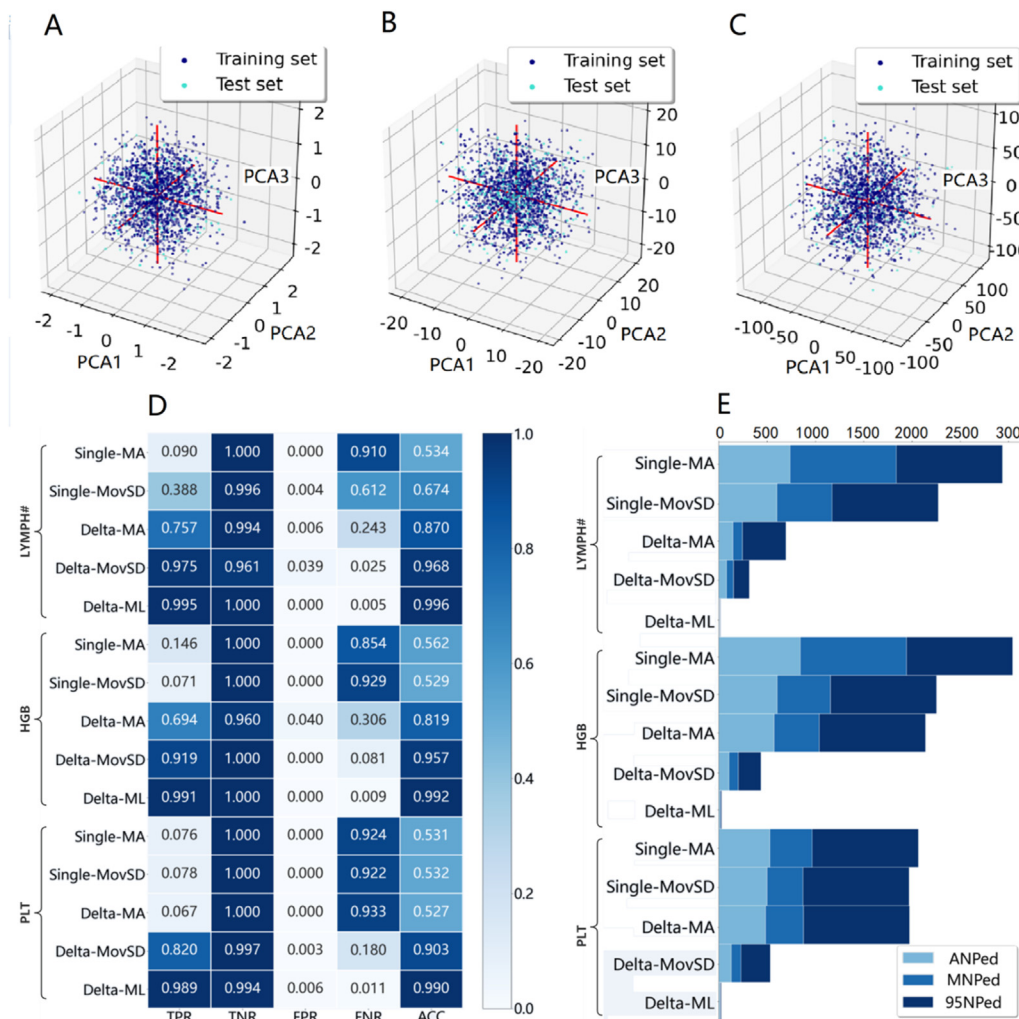
3.2. MLQC results by using delta-type data

Take RCV as an example, the influencing degree of the block size to the ability of QC event detection was explored by RF algorithm in ML. Here RCV was selected as the arithmetic example from the seven test items included because its within-individual and between individual ratio (CVi/CVg) was the largest, and this indicated that the degree of variation of this test item was considered relatively complex. The biased data at the critical level and unbiased data were used for model training. When RF algorithm parameters were set as default values, the block size was gradually increased starting from 5, and the AUC value of the model also increased; however, the change trend in AUC was no longer significant for block sizes above 10, so 10 was used as the block size for all subsequent experiments. When further adjusting the RF parameters, the search range of the parameter of N-trees was from 100 to 500 in step of 100; for each N-trees, the Max Depth was set from 50 to 300 in step of 50, and the accuracy of corresponding training and testing was counted in each group of experiments. The following Table 2 showed some data of the traversal experiment process, and the results showed that when the N-trees was 300 and the Max Depth was 200 was the optimal parameter combination.

3.3. The performance of QC event prediction for three-type data

For each test item, three types of data were experimented, namely: (1) PBRTQC by using single-type data processed by statistical truncation

Figure 3. The visualization of data distribution feature for the training and the test sets and the performance parameters of five experiments at critical bias for LYMPH #, HGB and PLT. A-C take examples of LYMPH #, HGB and PLT ordered from left to right, represented principal component analysis (PCA) plots of the training set and internal validation set. D represented the TPR, TNR, FPR, FNR and ACC of the five algorithms (TPR - true positive rate; TNR - true negative rate; FPR - false positive rate; FNR - false negative rate; ACC - accuracy). E represented ANPed, MNPed, 95NPed of them (ANped - average of Nped; MNped - median of Nped; 95Nped - 95 quantile of Nped).

Table 3. Test results of 5 algorithms at the critical level in leucocyte lineage.

Test item	Algorithm	TL (%)	Transfor-mation	BS	CL	CL_l	CL_U	TPR	TNR	FPR	FNR	ACC	ANPed	MNPed	95NPed
LYMPH#	Single-MA	1	BC	130	daily extremes	1.6912	1.9715	0.0899	1.0000	0.0000	1.0000	0.5335	736.6693	1100	1100
	Single-MovSD	5	BC	50	daily extremes	0.3741	0.6351	0.3876	0.9955	0.0045	0.9955	0.6741	599.3228	573	1100
	Delta-MA	10	-	110	daily extremes	-0.4356	0.3840	0.7572	0.9945	0.0055	0.9945	0.8704	145.3333	98	451
	Delta-MovSD	15	-	90	2.5CV	0.8371	1.7768	0.9750	0.9610	0.0390	0.9610	0.9685	77.5542	72	166
	Delta-ML		Processing	10	RF-model	-	-	0.9950	1.0000	0.0000	1.0000	0.9957	5.0000	5	6
LYMPH%	Single-MA	10	neat	130	daily extremes	28.4640	32.6476	0.5614	0.9989	0.0011	0.9989	0.7676	591.4567	494	1100
	Single-SD	10	neat	110	daily extremes	4.7516	9.4391	0.3686	0.9989	0.0011	0.9989	0.6691	635.5118	657	1100
	Delta-MA	5	-	90	3CV	-6.3065	5.2503	0.8771	1.0000	0.0000	1.0000	0.9343	330.3810	237	951
	Delta-SD	1	-	130	daily extremes	9.7270	22.4594	0.9590	0.9966	0.0034	0.9966	0.9765	63.9565	57	140
	Delta-ML		Processing	10	RF-model	-	-	0.9678	0.9847	0.0153	0.9847	0.9700	4.0000	4	5

TL - truncation limit; BC - Box–Cox transformation; BS-block size; CL_l - Control limit lower; CL_U - Control limit_upper; TPR - true positive rate; TNR - true negative rate; FPR - false positive rate; FNR - false negative rate; ACC - accuracy; ANped - average of Nped; MNped - median of Nped; 95Nped - 95 quantile of Nped.

limits; (2) PBRTQC by using delta-type data processed by statistical truncation limits; and (3) MLQC by using delta-type data processed by IF, where (1) and (2) experiments included 2 different algorithms, MA and MovSD, respectively. For MA and MovSD algorithms, 280 permutations derived from 5 truncation limits, 7 block sizes, 4 control limits, and 2 data transformation, respectively, were experimented. Due to paper space limitation, take examples of LYMPH #, HGB and PLT of CBC, Figure 3A, B, C showed the distribution feature of the training set and the test set in the MLQC. The coverage of the two sets was consistent for the three items in the scatter plots of A-C.

In Figure 3D, E, for the LYMPH#, HGB, and PLT, the experimental results for the positive bias at the critical level (equivalent to adding a new factor n with values between 1/4 and 1/2 to the previous simulation equation) showed that the model had a better accuracy (0.99), sensitivity (0.99), specificity (0.99) in the in dependent test set. Its performance surpassed over 50% by the optimal method by MA and MovSD. The MNPed of the ML QC results using delta-type data were all within 10, which were reduced by 93.06%, 90.72%, 93.20%, respectively, relative to the best results of PBRTQC. The remaining of test items, the tuning parameters, and their performance in different experimental conditions at the critical biases, were shown in Tables 3, 4, and 5.

In Figure 4 A, B, C, the different colored lines indicated the MNPed for each experiment, and the corresponding colored area indicated ANPed and 95NPed. The top horizontal line of the MNPed curve indicated the current error is not detected. The PBRTQC for both single-type and delta-type data were inferior to MLQC with delta-type data and showed asymmetric error detection curves for both positive and negative errors. However, MLQC was not only more sensitive than both PBRTQC, but also the MNPed curves were symmetrically distributed, suggesting that MLQC showed stronger ability for error detection, especially at critical levels.

For MLQC, all results of the MNPed to the seven test items for all biases (except for zero bias), were below 10 except for the MCHC item. The cumulative ANPed, MNPed and 95NPed corresponding to each of the 10 experiments for all biases (except for bias at critical level for each test item) showed significant differences among the methods. Taking LYMPH# as an example, the best performance for using single-type data was the MovSD algorithm with the above three cumulative values of 5044.92, 4871, and 8476.5, respectively; the best performance for using delta-type data was the MA algorithm with the three cumulative values of 1629.52, 1094, and 5213, respectively; and the three cumulative values for using delta-type data by ML QC were 45.07, 50, 60. Taking HGB as an example, the best performance for using single-type data was the MovSD algorithm with three cumulative values of 4550.80, 4751.5, and 8015.3; the best performance for using delta-type data was the MovSD algorithm with three cumulative values of 3515.60, 3504, and 5341.5; and the three cumulative values for delta-type data by ML QC were 89.42、91.5、104.55. Taking PLT as an example, the best performance for using single-type data was MovSD algorithm with 3 cumulative values of 4479.43, 3818, 8260.3; the best performance for using delta-type data was MovSD algorithm with 3 cumulative values of 3454.44, 3798, 5362.1; and the three cumulative values for delta-type data by ML QC were 70.33, 72, 102.2. It was obvious that the experimental results using delta-type data were better than single-type data, and the cumulative ANPed and MNPed of delta data ML QC for all 10 biases were reduced to within 100 from the previous 1000 or beyond. For the LYMPH#, HGB, and PLT, the cumulative MNPed of ML QC results relative to the best PBRTQC results were reduced by 95.43%, 97.39%, and 97.97%, respectively.

4. Discussion

Levey and Jennings introduced QC by using QC materials into the clinical laboratory in 1950 as the primary way to improve poor analytical performance [19]. The fundamental objective of QC program in the clinical laboratory is to characterize the analytical process accurately and thereby to provide information regarding the quality of results reported

Table 4. Test results of 5 algorithms at the critical level in erythrocyte lineage.

Test item	Algorithm	TL (%)	Transfor-mation	BS	CL	CL_l	CL_U	TPR	TNR	FPR	FNR	ACC	ANPed	MNPed	95NPed
HGB	Single-MA	10	BC	130	daily extremes	125.9305	130.8100	0.1459	1.0000	0.0000	1.0000	0.5622	842.0866	1100	1100
	Single-MovSD	5	neat	50	daily extremes	2.4035	14.5398	0.0709	1.0000	0.0000	1.0000	0.5286	602.7619	551	1100
	Delta-MA	15	-	110	daily extremes	-3.8918	5.1418	0.6943	0.9596	0.0404	0.9596	0.8193	570.3228	469	1100
	Delta-MovSD	5	-	30	daily extremes	11.4405	26.4159	0.9191	1.0000	0.0000	1.0000	0.9568	103.1467	97	236
	Delta-ML		Processing	10	RF-model	-	-	0.9912	0.9997	0.0003	0.9997	0.9923	8.8500	9	11
MCH	Single-MA	15	BC	90	3CV	29.2380	30.2543	0.4386	0.9722	0.0278	0.9722	0.7015	643.4252	643	1100
	Single-MovSD	15	Neat	30	daily extremes	1.2504	2.1272	0.4785	0.9558	0.0442	0.9558	0.7136	547.8189	454	1100
	Delta-MA	5	-	30	2.5CV	-1.1558	1.2430	0.1658	0.9939	0.0061	0.9939	0.5759	568.5952	426	1100
	Delta-MovSD	1	-	20	3CV	1.4054	4.1035	0.4086	0.9989	0.0011	0.9989	0.6900	196.8989	136	494
	Delta-ML		Processing	10	RF-model	-	-	0.9909	0.9923	0.0077	0.9923	0.9911	8.3500	8	12
MCHC	Single-MA	20	BC	130	daily extremes	329.3763	334.2094	0.5485	0.9990	0.0010	0.9990	0.7704	630.4646	559	1100
	Single-MovSD	0	Neat	30	3CV	6.7953	11.7294	0.5884	0.9990	0.0010	0.9990	0.7907	567.5827	429	1100
	Delta-MA	1	-	30	3CV	-9.1536	9.6639	0.4066	0.9989	0.0011	0.9989	0.6890	268.1905	200	669
	Delta-MovSD	5	-	90	daily extremes	5.6724	10.6232	0.9570	0.9977	0.0023	0.9977	0.9760	83.3125	83	120
	Delta-ML		Processing	10	RF-model	-	-	0.9913	0.9930	0.0070	0.9930	0.9915	8.7500	9	11
R–CV	Single-MA	10	BC	90	daily extremes	12.5823	13.1440	0.3866	0.9674	0.0326	0.9674	0.6697	476.0472	293	1100
	Single-MovSD	5	BC	50	3CV	0.1928	1.1664	0.4026	0.9568	0.0432	0.9568	0.6756	187.3571	94	921
	Delta-MA	20	-	30	daily extremes	-0.4530	0.4332	0.2248	0.9630	0.0370	0.9630	0.5884	473.9370	283	1100
	Delta-MovSD	1	-	30	2.5CV	0.8907	3.1528	0.7922	0.9978	0.0022	0.9978	0.8902	83.4404	76	185
	Delta-ML		Processing	10	RF-model	-	-	0.9897	0.9843	0.0157	0.9843	0.9890	10.3000	10	12

TL - truncation limit; BC - Box–Cox transformation; BS-block size; CL_l - Control limit_lower; CL_U - Control limit_ upper; TPR - true positive rate; TNR - true positive rate; FPR - false positive rate; FNR - false negative rate; ACC - accuracy; ANped - average of Nped; MNped - median of Nped; 95Nped - 95 quantile of Nped.

Table 5. Test results of 5 algorithms at the critical level in platelet lineage.

Test item	Algorithm	TL (%)	Transfor-mation	BS	CL	CL_l	CL_U	TPR	TNR	FPR	FNR	ACC	ANPed	MNPed	95NPed
PLT	Single-MA	15	neat	110	3CV	212.2105	242.0452	0.0759	1.0000	0.0000	1.0000	0.5312	525.9528	440	1100
	Single-MovSD	15	BC	30	daily extremes	4.7925	19.7785	0.0779	1.0000	0.0000	1.0000	0.5322	498.0866	372	1100
	Delta-MA	0	-	30	2.5CV	-48.8773	46.9443	0.0669	1.0000	0.0000	1.0000	0.5266	482.6190	392	1100
	Delta-MovSD	1	-	30	3CV	64.3340	178.1932	0.8202	0.9966	0.0034	0.9966	0.9033	126.3592	103	303
	Delta-ML		Processing	10	RF-model	-	-	0.9894	0.9937	0.0063	0.9937	0.9900	7.4500	7	11

TL - truncation limit; BC - Box–Cox transformation; BS-block size; CL_l - Control limit_lower; CL_U - Control limit upper; TPR - true positive rate; TNR - true positive rate; FPR - false positive rate; FNR - false negative rate; ACC - accuracy; ANped - average of Nped; MNped - median of Nped; 95Nped - 95 quantile of Nped.

for clinical specimens [20]. Patient-based real-time quality control (PBRTQC) is a generic term for the use of patient results for real-time quality-control purpose as an alternative tool for insufficient or inefficient QC. Recent studies by Xincen Duan, et al. [3] used an additional regression adjustment before using a common algorithm in the RARTQC framework removed autocorrelation in the test results, and allowed researchers to add additional variables, and to improve data transformation; Ichihara et al. [5] set up the weighted cumulative delta-check (wCDI) method, applying a series of techniques for data stability.

Otherwise, in real settings, clinical testing data takes on significant heterogeneous feature and contains a number of extreme data. And some independent variables from population impact patient data-oriented QC method performance, such as age, sex, patient type, within-or-between biological variability, sample mislabeling, patient misidentification, distribution patterns of test results.

Delta check is a quantity of change expressed as a magnitude or ratio that can be determined by calculating continuous paired data from representative patients. It has been used for monitoring quality issues in total testing process, such as patient identification errors, sample identification errors or sample mishandling in the pre-analytical phase, as well as for QC in analytical phase [21]. While DC is still limited by a simple linear transformation for handling the noise from data. ML, as one of the main tools for data mining, can seek structural features of the data from complex dimensions and big data volume. In this paper, ML technique is introduced into QC, combined with delta data, to explore a newly overall protocol for data stability, thus improving patient data-based QC effectiveness.

An overall protocol for stabilizing data is set up by using delta data in combination with IF algorithm in ML. First, inputs are changed from single-type data to delta-type data, the paired-data weakens the perturbation from single patient data although additional intra-individual variation information introduced, which is equivalent to denoising to the single data, making the characteristics of the bias expected to be identified more significant, thus improving the accuracy of QC event detection. The experimental results in Table 1 showed that the SD was reduced by 31.85% on average after the single data of the seven test items were converted to delta data. Second, for the stabilization of sample sources, ML IF shows a powerful advantage, which is essentially based on the distribution of samples in a high-dimensional space to remove outliers in samples. It transforms a sample pre-processing issue into a classification problem based on density and distance by the spatial location of the samples. The IF algorithm enables to recursively segment data set randomly until all sample points are completely isolated. With this segmentation strategy, every data point is effectively utilized, thus improving data utilization, reducing information loss, and preventing denoising failure caused by removing outlier data directly by means of setting statistical truncation limits. When the delta-type data processed by IF, the SD was reduced by an average of 72.36% compared to the single-type data processed only by a truncation limit on statistics.

As while, RF algorithm is also used to further improve the effective utilization of data, which in turn improves the QC effectiveness. In this study, a ML RF model is established, which maps every single data point within a moving window length to a high-dimensional space, and the data points within the same moving window are regarded as a whole and mapped in the high-dimensional space with a divisible population effect. Here, the data points within each moving window are characterized as serialized information after feature engineering process in ML, and the data points within each moving window produce a horizontal cross-correlation into the RF model in a multidimensional parallel manner, instead of by the way of the mean calculated by using the data points in one block size data in PBRTQC. The RF model is trained through many data iterations, and there exists a technique similar to IF in its process, which summarizes a relatively reasonable delimited hyper-curve. And when a new unknown input sequence needs to be predicted, the sequence will be compared with the delimited hyper-curve, a final anomaly probability is output by calculating the size relationship of each element
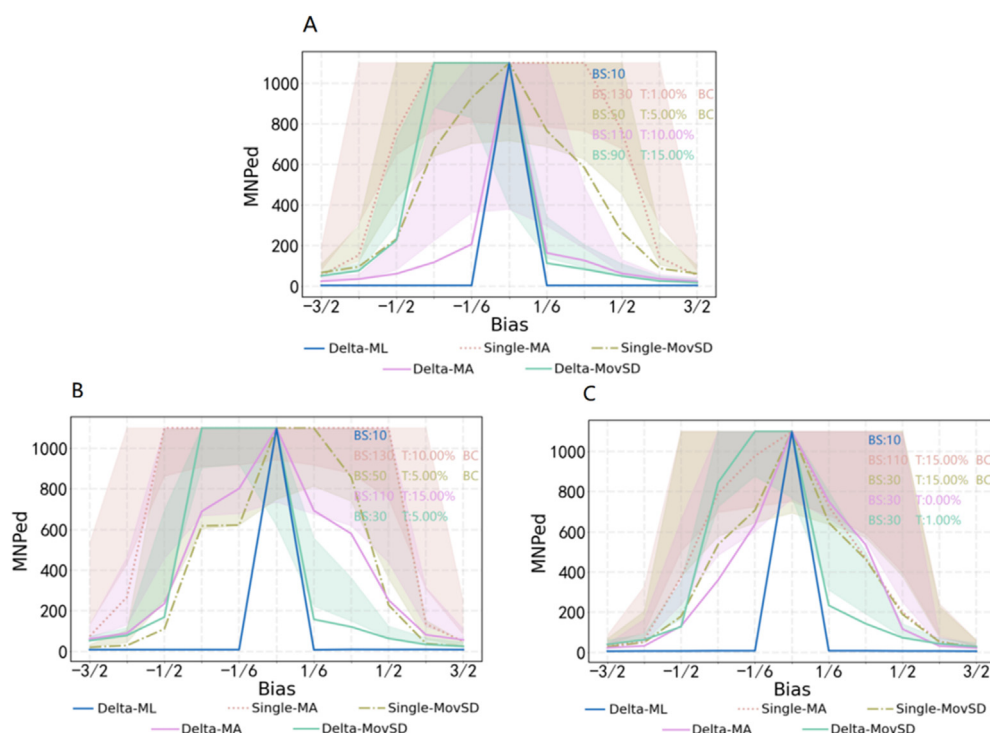
Figure 4. The curves for the comparison performance of 5 experiments. A,B,C corresponded to LYMPH #, HGB, and PLT, respectively. Colored lines represented MNPed for each bias, colored area represented the associated 95NPed. Parameters were displayed in the top corner (BS: block size; T: truncation limit, BC: with Box–Cox transformation).

in it. Our experimental results showed that ANPed, MNPed and 95NPed of MLQC using delta data, were basically within 10, which were reduced by 96.39%, 95.34% and 96.37% respectively compared to the optimal results of PBRTQC. The sensitivity and specificity of the MLQC were also both better than the best results of PBRTQC. Here, TPR refers to the sensitivity, which represents out-of-control status of QC, and TNR refers to the specificity, which represents in-control status of QC. High sensitivity and specificity of our model indicates that the MLQC is rarely possible for misclassification and omission of QC event detection as well as without delay and labor-intensive to false alarm.

In summary, by implementing an overall protocol for data processing, together with ML algorithm innovation, an effective tool for QC error detection is established.

Declarations

Author contribution statement

Yufang Liang, Rui Zhou: Conceived and designed the experiments, Analyzed and interpreted the data; Wrote the paper.

Zhe Wang: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Qingtao Wang: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data.

Xiang Feng, Zewen Han: Performed the experiments.

Biao Song: Performed the experiments; Contributed reagents, materials, analysis tools or data.

Dawei Huang: Analyzed and interpreted the data Contributed reagents, materials, analysis tools or data.

Wei Wang: Contributed reagents, materials, analysis tools or data.

Data availability statement

Data included in article/supplementary material/referenced in article.

Declaration of interest's statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

[1] W.G. Miller, A. Erek, T.D. Cunningham, O. Oladipo, M.G. Scott, R.E. Johnson, Commutability limitations influence quality control results with different reagent lots, Clin. Chem. 57 (1) (2011) 76–83.

[2] M.A. Thaler, R. Iakoubov, A. Bietenbeck, P.B. Luppa, Clinically relevant lot-to-lot reagent difference in a commercial immunoturbidimetric assay for glycated hemoglobin A1c, Clin. Biochem. 48 (16-17) (2015) 1167–1170.

[3] X. Duan, B. Wang, J. Zhu, C. Zhang, W. Jiang, J. Zhou, W. Shao, Y. Zhao, Q. Yu, L. Lei, K.L. Yiu, K.T. Chin, B. Pan, W. Guo, Regression-adjusted real-time quality control, Clin. Chem. 67 (10) (2021) 1342–1350.

[4] D. Ng, F.A. Polito, M.A. Cervinski, Optimization of a moving averages program using a simulated annealing algorithm: the goal is to monitor the process not the patients, Clin. Chem. 62 (10) (2016) 1361–1371.

[5] T. Yamashita, K. Ichihara, A. Miyamoto, A novel weighted cumulative delta-check method for highly sensitive detection of specimen mix-up in the clinical laboratory, Clin. Chem. Lab. Med. 51 (4) (2013) 781–789.

[6] G.S. Cembrowski, Q. Xu, M.A. Cervinski, Average of patient deltas: patient-based quality control utilizing the mean within-patient Analyte variation, Clin. Chem. 67 (7) (2021) 1019–1029.

[7] D.V. Tran, G.S. Cembrowski, T. Lee, T.N. Higgins, Application of 3-D Delta check graphs to HbA1c quality control and HbA1c utilization, Am. J. Clin. Pathol. 130 (2) (2008) 292–298.

[8] F.G. Strathmann, G.S. Baird, N.G. Hoffman, Simulations of delta check rule performance to detect specimen mislabeling using historical laboratory data, Clin. Chim. Acta 412 (21-22) (2011) 1973–1977.

[9] K. Ovens, C. Naugler, How useful are delta checks in the 21 century? A stochastic-dynamic model of specimen mix-up and detection, J. Pathol. Inf. 3 (2012) 5.

[10] E.W. Randell, S. Yenice, Delta Checks in the clinical laboratory, Crit. Rev. Clin. Lab Sci. 56 (2) (2019) 75–97.

[11] R.Z. Tan, C. Markus, T.P. Loh, An approach to optimize delta checks in test panels - the effect of the number of rules included, Ann. Clin. Biochem. 57 (3) (2020) 215–222.

[12] R.Z. Tan, C. Markus, T.P. Loh, Relationship between biological variation and delta check rules performance, Clin. Biochem. 80 (2020) 42–47.

[13] R.Z. Tan, C. Markus, K.W. Choy, J.C.G. Doery, T.P. Loh, Optimized delta check rules for detecting misidentified specimens in children, Am. J. Clin. Pathol. 153 (5) (2020) 605–612.

[14] Z.H. Zhou, Machine Learning, first ed., Tsinghua University Press, Bei Jing, 2016.

[15] B. Liu, Research on Classification Model and Algorithm of Big Data, first ed., Yunnan University Press, Yunnan, 2019.

[16] A. Bietenbeck, M.A. Cervinski, A. Katayev, T.P. Loh, H.H. van Rossum, T. Badrick, o.b.o.t.I.F.o.C. Chemistry, L.M.C.o.A. Quality, Understanding patient-based real-time quality control using simulation modeling, Clin. Chem. 66 (8) (2020) 1072–1083.

[17] X. Duan, B. Wang, J. Zhu, W. Shao, H. Wang, J. Shen, W. Wu, W. Jiang, K.L. Yiu, B. Pan, W. Guo, Assessment of patient-based real-time quality control algorithm performance on different types of analytical error, Clin. Chim. Acta 511 (2020) 329–335.

[18] J. Liu, C.H. Tan, T. Badrick, T.P. Loh, Moving standard deviation and moving sum of outliers as quality tools for monitoring analytical precision, Clin. Biochem. 52 (2018) 112–116.

[19] S. Levey, E.R. Jennings, The use of control charts in the clinical laboratory, Am. J. Clin. Pathol. 20 (11) (1950) 1059–1066.

[20] R. Rej, R.W. Jenny, J.P. Bretaudiere, Quality control in clinical chemistry: characterization of reference materials, Talanta 31 (10 Pt 2) (1984) 851–862.

[21] Y. Fei, W. Wang, Z. Wang, Delta checks based on individual patients' data, Lab. Med. 32 (1) (2017) 64–68. https://kns.cnki.net/kcms/detail/31.1915.R.2017021 3.1244.030.html.