

Rui Zhou, Yu-fang Liang, Hua-Li Cheng, Wei Wang, Da-wei Huang, Zhe Wang, Xiang Feng, Ze-wen Han, Biao Song, Andrea Padoan, Mario Plebani* and Qing-tao Wang*

A highly accurate delta check method using deep learning for detection of sample mix-up in the clinical laboratory

<https://doi.org/10.1515/cclm-2021-1171>

Received November 5, 2021; accepted November 26, 2021;
published online December 29, 2021

Abstract

Objectives: Delta check (DC) is widely used for detecting sample mix-up. Owing to the inadequate error detection and high false-positive rate, the implementation of DC in real-world settings is labor-intensive and rarely capable of absolute detection of sample mix-ups. The aim of the study was to develop a highly accurate DC method based on designed deep learning to detect sample mix-up.

Methods: A total of 22 routine hematology test items were adopted for the study. The hematology test results, collected from two hospital laboratories, were independently divided into training, validation, and test sets. By selecting six mainstream algorithms, the Deep Belief Network (DBN) was able to learn error-free and artificially (intentionally) mixed sample results. The model's analytical performance was

evaluated using training and test sets. The model's clinical validity was evaluated by comparing it with three well-recognized statistical methods.

Results: When the accuracy of our model in the training set reached 0.931 at the 22nd epoch, the corresponding accuracy in the validation set was equal to 0.922. The loss values for the training and validation sets showed a similar (change) trend over time. The accuracy in the test set was 0.931 and the area under the receiver operating characteristic curve was 0.977. DBN demonstrated better performance than the three comparator statistical methods. The accuracy of DBN and revised weighted delta check (RwCDI) was 0.931 and 0.909, respectively. DBN performed significantly better than RCV and EDC. Of all test items, the absolute difference of DC yielded higher accuracy than the relative difference for all methods.

Conclusions: The findings indicate that input of a group of hematology test items provides more comprehensive information for the accurate detection of sample mix-up by machine learning (ML) when compared with a single test item input method. The DC method based on DBN demonstrated highly effective sample mix-up identification performance in real-world clinical settings.

Keywords: data pre-processing; deep learning; delta check; machine learning; pre-analytical error; sample mix-up.

Rui Zhou and Yu-fang Liang contributed equally to this work.

***Corresponding authors:** Mario Plebani, Department of Laboratory Medicine, University Hospital of Padova, 35218 Padova, Italy, E-mail: mario.plebani@unipd.it; and Qing-tao Wang, Department of Laboratory Medicine, Beijing Chao-yang Hospital, Capital Medical University, Beijing, P.R. China; and Beijing Center for Clinical Laboratories, No. 8 Gongti South Street, Chaoyang District, Beijing, 100020, P.R. China, E-mail: wqt36@163.com

Rui Zhou, Department of Laboratory Medicine, Beijing Chao-yang Hospital, Capital Medical University, Beijing, P.R. China; and Beijing Center for Clinical Laboratories, Beijing, P.R. China

Yu-fang Liang and Hua-Li Cheng, Department of Laboratory Medicine, Beijing Chao-yang Hospital, Capital Medical University, Beijing, P.R. China

Wei Wang, Department of Blood Transfusion, Beijing Ditan Hospital, Capital Medical University, Beijing, P.R. China

Da-wei Huang, Department of Laboratory Medicine, Beijing Longfu Hospital, Beijing, P.R. China

Zhe Wang, Xiang Feng, Ze-wen Han and Biao Song, Inner Mongolia Wesure Date Technology Co., Ltd, Inner Mongolia, P.R. China

Andrea Padoan, Department of Laboratory Medicine, University Hospital of Padova, Padova, Italy.

<https://orcid.org/0000-0003-1284-7885>

Introduction

Reducing patient harm through minimizing the risk of laboratory error is a major safety principle of laboratory practice. In the clinical laboratory testing process, pre-analytical, analytical, and postanalytical phases are the three phases of laboratory practice and are referred to as the total testing process (TTP) [1–3]. However, pre-analytical errors account for approximately 60–70% of all errors found in TTP [4, 5] with the primary source of error being related to the clinical sample. Common causes of errors include patient or sample misidentification, sample labeling errors, sample contamination, and measurement interferences in samples.

Delta check (DC), an error screening tool, calculates the difference between the current and the preceding results, and compares this difference against a predefined limit. If this difference is within a predefined DC limit, the result can be released to the clinical team. Otherwise, if the difference is greater than the predefined DC limit, this raises the possibility of an error in the pre-analytical stage. The concept of DC was introduced by Nosanchuk and Gottman in 1974 as a QC technique to identify misidentified samples [6]. In 1975, Ladenson [7] described the first use of computers to automatically compare patient's current and previous results in real time. With the widespread use of auto-verification in various areas of laboratory medicine, DC is becoming a mandatory component of auto-verification rules to identify results that require additional review before release to the medical record [8].

With more emphasis on proper sample labeling, the prevalence of mislabeled samples may be reduced in certain settings. While efforts to improve labeling practices may mitigate one source of sample mix-up, the ever-expanding scope of tests offered and the sharp increase in sample volumes processed in modern large clinical laboratories introduces high levels of complexity that counteract improvement efforts leaving a sample mix-up rate of 1.2%. Considering the potentially serious health risks posed by unidentified sample mix-up errors to the patient, DC may be as a useful tool to mitigate these risks through early identification of potential sample mix-up errors. Furthermore, DC is unaffected by the prevalence of mislabeled samples.

Issues such as low accuracy of error detection and significant variations in the implementation of DC by different laboratories are, in part, a consequence of the DC method itself and differences for DC limits. Related studies have indicated that the accuracy of DC methods available ranged from 15% to 76% [9]. In addition, DC rules are typically defined for individual analytes of interest. However, in practice, multiple items are often tested and results reported as a group or panel. In such instances, multiple DC rules can be combined according to the common test panel, and the interpretation of DC limits for a grouped test panel should be different from a single analyte, since the number of hypothesis tests (i.e. the number of DC rules) applied is much higher and should be taken into account [8, 10].

A more detailed and formal definition of machine learning (ML), first introduced by Arthur Samuel in 1959, was described as a computer program that by learning from experience (E) with respect to some class of tasks (T) and performance measure (P), if its performance at tasks in T, as measured by P, improved with experience E [11]. In

recent years, the widespread recognition of data-driven methods has made ML algorithms widely used in bioinformatics studies, and biomolecular correlation prediction [12]. However, to our knowledge, there are no related studies demonstrating how to use deep ML technique to establish a DC method to date.

In this work, employing hematology test item results, we tried to establish a highly accurate DC method by using deep ML to detect sample mix-up in clinical laboratories. The performance of the deep ML approach was assessed by comparison with three well-statistical DC methods.

Materials and methods

Data collection and exclusion criteria

In ML, data can be divided into a training set, a validation set, and a test set. The validation set can be understood as a part of the training set to monitor the process of model training. The three datasets are independently separated. In our study, 423,290 deidentified hematology test results measured on the XN-9000 (Sysmex, Kobe, Japan) from 01/2018 to 12/2018 were extracted from the Laboratory Information System (LIS) of the Beijing Chaoyang Hospital. The data from 01/2018 to 10/2018 was used as the training set and the data from 11/2018 to 12/2018 was used as the validation set. Twenty-two thousand four hundred sixty hematology test results from 01/2018 to 12/2018 measured on the BC-5390 (Mindray, Shenzhen, China) were extracted from the LIS of the Beijing Long-fu Hospital to be used as the test dataset. Data filtering rules applied to both the XN and the Mindray datasets. Filter rules included: 1) patients with only one result during the study period were excluded; 2) the first pair of results of each remaining patient was included; 3) Tukey's criteria [13], which defined outliers as values lying three interquartile ranges below the 25th percentile or above the 75th percentile, was applied to remove outlying data; 4) patients with two results after applying Tukey's criteria were included for further analysis; 5) in consideration of gender-dependent and age-dependent differences in distributions of test results, all test results were separated into male and female groups for all test items, and 6) the results of patients aged from 14 years old to 60 years old were included; 7) the time interval of DC was defined to one year [9]. The information of deidentified results included: patient type, sex, age, sample number, sample type and all test item result respective values and units. The test results were randomly sorted by a shuffle function in Python 3.7.3 and then automatically matched the current data and preceding data from different patients to generate a mismatched data, simulating a switched sample scenario. The original paired test results were assumed to error-free. The absolute and relative differences were assessed by original matched and mismatched data.

ML method: data pre-processing

After filtering data by predefined exclusion rules, the data was assessed for consistency of analyte and unit parameters and possible missing

values for each pair of data. Following assessment, the data was normalized with the Standard Scaler tool in soft package python 3.7.3. Then absolute and relative differences of data were calculated. Isolated forest algorithm was used for removing extreme values in delta data.

ML method: algorithm

The classification problem can be implemented by using classifiers with different algorithms. In our work, six mainstream classifiers were tested and evaluated by confusion matrix. They were Deep Belief Network (DBN), Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), K-Nearest Neighbor (KNN), Naive Bayesian Classifier (NBC). The introduction to the six algorithms is depicted in Supplementary Materials and Methods.

DBN belongs to a deeper neural network in the field of deep learning, which consists of Restricted Boltzmann Machine (RBM) and neural network (NN). DBN was selected for establishing our model. It was implemented by deep learning framework Keras in Python 3.7.3. The main tuning parameters included: 1) “learning_rate_rbm” for controlling the rate of learning; “batch_size_rbm” for selecting the number of sample each time; “n_epochs_rbm” for training iterative epochs; “activation_function_nn” for realizing the nonlinearity between the input and output of neuron.

ML method: implementation

Data pre-processing and model analysis were performed by “numpy” and “pandas” tools in Python 3.7.3 and by “sklearn” and “tensorflow” encoding frameworks in Python 3.7.3. All software packages were accessed from the sklearn library_2.4.0 in the public Python. Python is a computer language that can be used in scientific computing and data analysis, and is currently a mainstream programming tool of artificial intelligence.

Reference change value (RCV) method

RCV limits of each test item dependent on biological variability (BV) [14] were estimated using the following formula:

$$RCV (\%) = K * \sqrt{2} * \sqrt{(CV_a^2 + CV_i^2)}$$

Coverage factor K was varied from 1.5 to 3.3 in steps of 0.1, coefficient of variation (CV_a) was analytical imprecision, and CV_i was within-subject BV. CV_a was calculated from the mean CV , which was considered a representative interval of long-term imprecision. There were two-type CV_a ($CV_{a,1}, CV_{a,2}$) calculated. $CV_{a,1}$ used whole data. For the data of $CV_{a,2}$, we excluded pairs of test results if both test results constituting the pair were within the reference interval (CL). Extended CL here referred to twice the upper limit value of the CL.

Empirical delta check (EDC) method

EDC limits of each test item were calculated using the absolute or the absolute difference. For each patient, the relative difference for patient, Δx_r , was given by:

$$\Delta x_r = \frac{|x_1 - x_2|}{x_1},$$

where x_1 and x_2 corresponded to the early and later dates of the patient, respectively.

The absolute difference for each patient, Δx_a , was given by:

$$\Delta x_a = |x_1 - x_2|.$$

For relative difference, the DC limits were varied from 1% to 200% range in steps of 0.1%, whereas for absolute difference, the DC limits were varied from 1% to 200% of the average test result in the same step.

Revised weighted delta check (RwCDI) method

For all test items, a distribution of values for each test was transformed into approximately Gaussian form by using the Box-Cox formula [15]. To make data comparable and unaffected by measurement units, all the transformed test results were standardized to a uniform scale on the basis of reference interval (RI) as described by Ichihara [16]. As a next step, we used Formula (1) to get the absolute difference for each test item and calculated a new index termed weighted cumulative delta index (wCDI). We got three panels (including 5-item, 10-item and 22-item) to compute new parameter, and continued following the EDC method. The details of the procedure are described in Figure 1.

Evaluation metrics

The four parameters were defined below as [17]: 1) True Positive (TP): delta check limit was exceeded CL of mismatched queue; 2) False Positive (FP): delta check limit was exceeded CL of matched queue; 3) False Negative (FN): delta check limit was not exceeded CL of mismatched queue; 4) True Negative (TN): when the delta check limit was not exceeded CL of matched queue.

The parameters on confusion matrix were calculated, including true positive rate (TPR), true negative rate (TNR), false positive rate (FPR), false negative rate (FNR), accuracy rate (ACC). We evaluated our model using receiver operating characteristic (ROC) analysis, and the area under the curve (AUC) was calculated, which ranges between 0.0 and 1.0, with values of 0.5 for random classification and 1.0 for perfect classification.

Results

Data distribution

Total 445,750 data was included from two hospitals, 123,365 pairs of data in matched queue and 123,365 sets of data in mismatched queue. We split the 423,290 data of Beijing Chao-yang Hospital dataset into a training set from 1/2018 to 10/2018 and a validation set from 11/2018 to 12/2018. We used the 22,460 data of Long-fu Hospital from 1/2018 to 12/2018 as test set.

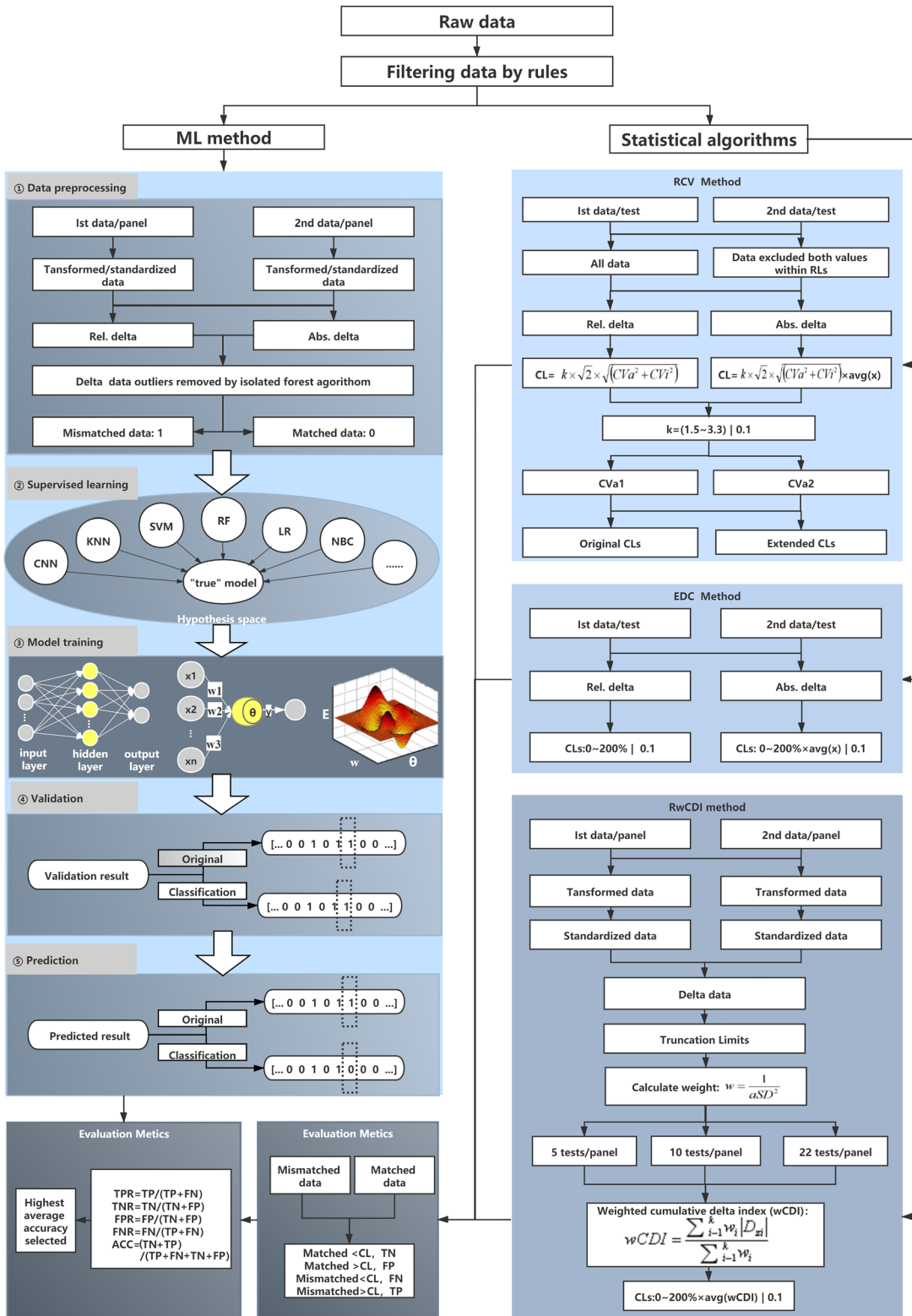


Figure 1: A comprehensive process and architecture of DC detection of sample mix-up.

Prior to conducting further analysis, data distribution characteristics were examined; the distribution of MCH and MCHC had a skewness (Sk) close to zero (−0.15 and −0.02)

resembling a normal distribution. The other test items examined had skewed distribution with $|Sk| > 0.3$ ranging from 0.31 (MCV) to 2.02 (NEUT). All items kurtosis

Table 1: Prediction scores of models created by different ML algorithms.

Algo	TPR	TNR	FPR	FNR	ACC	AUC	Model dimensions
DBN	0.9295	0.9325	0.0675	0.0705	0.9310	0.9773	22
	0.9045	0.9230	0.0770	0.0955	0.9137	0.9544	10
KNN	0.9309	0.8878	0.1122	0.0691	0.9094	0.9455	22
	0.9014	0.8681	0.1319	0.0986	0.8848	0.9334	10
SVM	0.9261	0.9196	0.0804	0.0739	0.9229	0.9678	22
	0.9103	0.8861	0.1139	0.0897	0.8982	0.9556	10
RF	0.9117	0.9285	0.0715	0.0883	0.9201	0.9689	22
	0.8930	0.8967	0.1033	0.1070	0.8948	0.9504	10
LR	0.9247	0.8972	0.1028	0.0753	0.9110	0.9698	22
	0.9158	0.8580	0.1420	0.0842	0.8869	0.9550	10
NBC	0.9164	0.8322	0.1678	0.0836	0.8743	0.9509	22
	0.9229	0.7876	0.2124	0.0771	0.8552	0.9404	10

Algo, algorithm; DBN, Deep Belief Network; RF, Random Forest; SVM, Support Vector Machine; LR, Logistic Regression; KNN, K-Nearest Neighbor; NBC, Naive Bayesian Classifier; TPR, true positive rate; FPR, false positive rate; TNR, true negative rate; FNR, false negative rate; ACC, accuracy; AUC, area under the receiver operating curve.

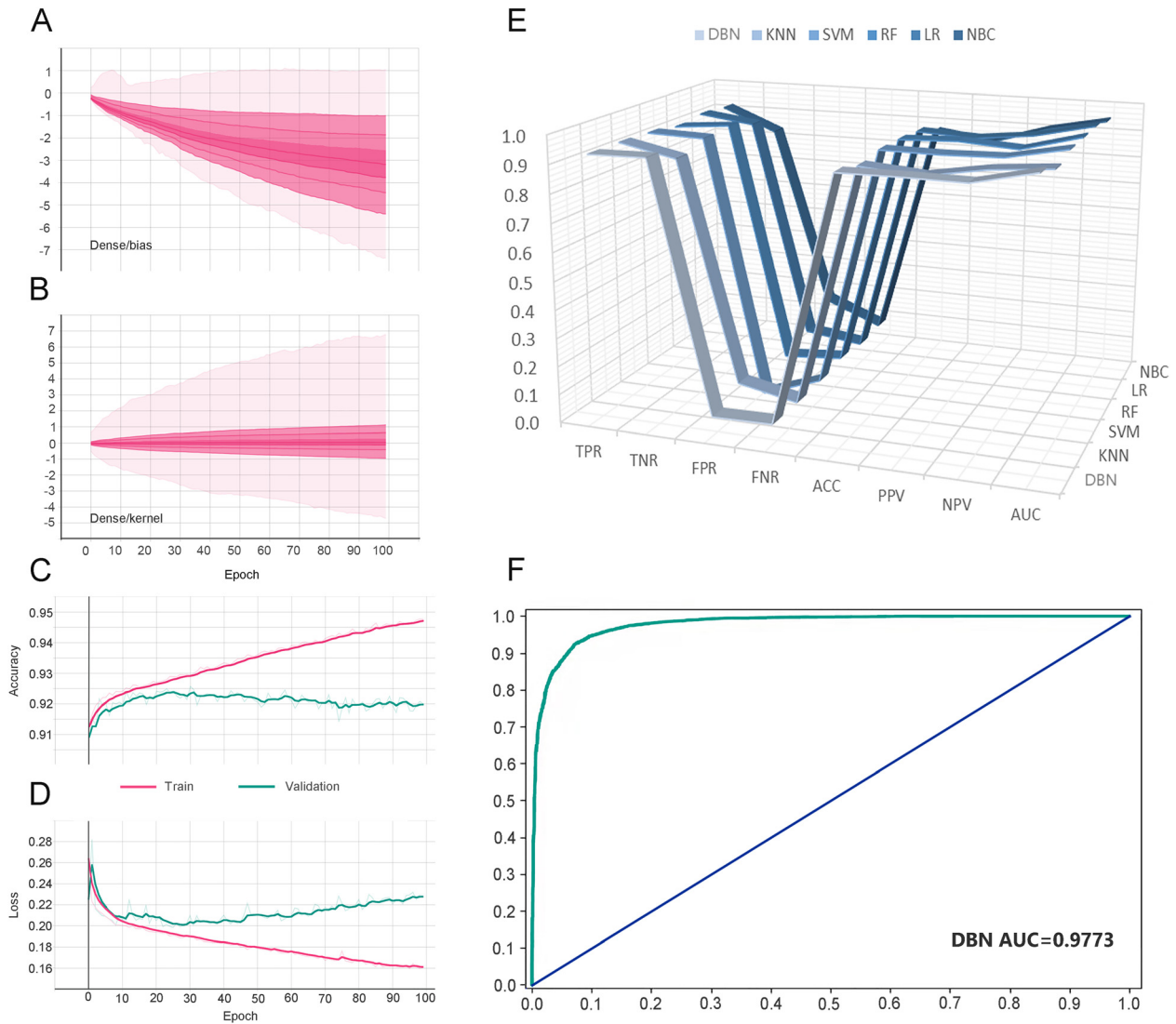


Figure 2: DBN training process flowchart.

(A–B) Represents the change of parameters with time for certain layer in the training dataset. (C–D) Represents the change of the accuracy and loss value with time in the training dataset and validation dataset from Beijing Chao-yang Hospital. In each diagram, red colored line represents the training dataset; green colored line the validation dataset. (E) Represents the results of ML algorithm selection. (F) Represents DBN ROC curve of the test dataset from Beijing Long-fu Hospital.

(peakedness of distribution) ranged from -0.83 (RBC) to 16.73 (MCHC).

Performance evaluation of six ML algorithms

We evaluated six types of classifiers: SVM, KNN, RF, LR, NBC and DBN model. The evaluation metrics of each model in absolute male data are shown in Table 1 and the detailed ROC curves are depicted in Supplementary Figure 1. We also evaluated the performance of six ML methods for different number of combinations of hematology test items (10-item and 22-item). The performance of all ML methods combined by 22-item was better, as shown in Table 1. As a result, we selected the 22-test item ML model for model training. The 22 hematology test items were input as a multi-label classification task in the ML method, as shown in Figure 1. Figure 2C and D shows the change curve of the accuracy and loss value with time in the training set and the validation set at the current training model. The DBN model clearly achieved the highest accuracy on the test dataset, shown in Figure 2F. RF achieved closely competitive performance for current dataset. As shown in Figure 2E, the performance of DBN was obviously superior to those of the other five ML algorithms in DC method.

Performance of improved DBN model

The robustness and fault tolerance of RF, KNN and SVM to noise data were low, the learning ability of LR and NBC to multi-attribute nonlinear data was weak as well. Compared with the above 5 ML algorithms, NN performed stronger robustness and fault tolerance to noise data, stronger

learning ability to complex nonlinear correlation, and higher classification accuracy. However, NN algorithm was also not omnipotent, with the shortcomings of slackness of learning rate or relatively inadequate accuracy. We designed an improved DBN with restricted Boltzmann machine (RBM) as shown in Figure 3. DBN consisted of two parts: a feature learner with multi-layer RBMs and a classifier with a back propagation (BP). Model training initialized, RBM enabled to be self-encoded to strengthen data features, thus enlarging significant difference between positive data and negative data. Intra-and-inter RBM learning method not only dramatically improved learning rate, but also prevented exploding gradient and vanishing gradient problems, thus to assure capturing the higher accuracy than traditional NN as much as possible.

Comparison with three statistical DC methods

To evaluate the performance of the DBN model, it was compared with three statistical methods which had been proven to have high performance in their respective domains.

Absolute difference and relative difference of all test items were shown on male/female dataset. Nineteen thousand eight hundred seventy-six test results of Long-fu hospital were used to compare the DBN model parameters with three DC methods. Figure 4 demonstrated that seven parameters of four methods collected including TPR, TNR, FPR, FNR, PPV, NPV and ACC. Meanwhile, Figure 4 depicted the absolute difference results in male data among four methods. For the sake of space, the absolute and relative difference results in male and in female are

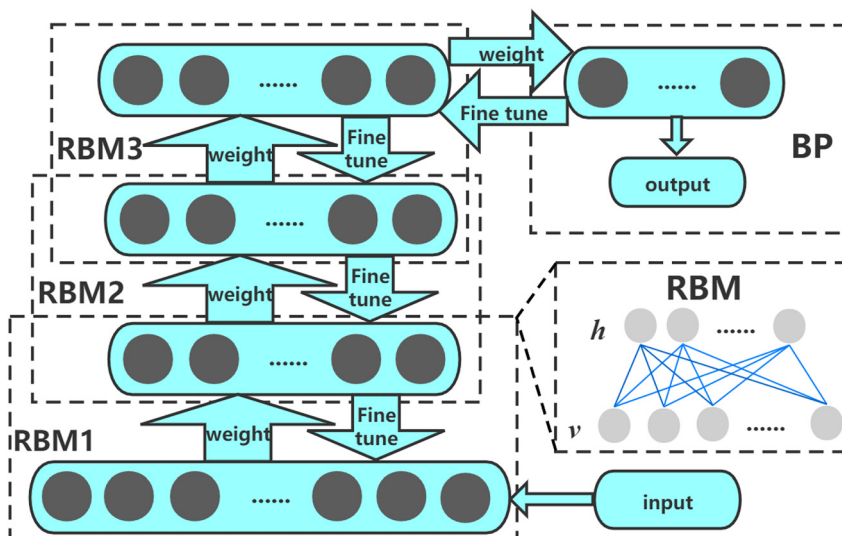


Figure 3: DBN parameter tuning chart. DBN consisted of two parts: a feature learner with multi-layer RBMs and a classifier with a back propagation (BP). Parameter tuning was realized at RBM and BP parts separately.

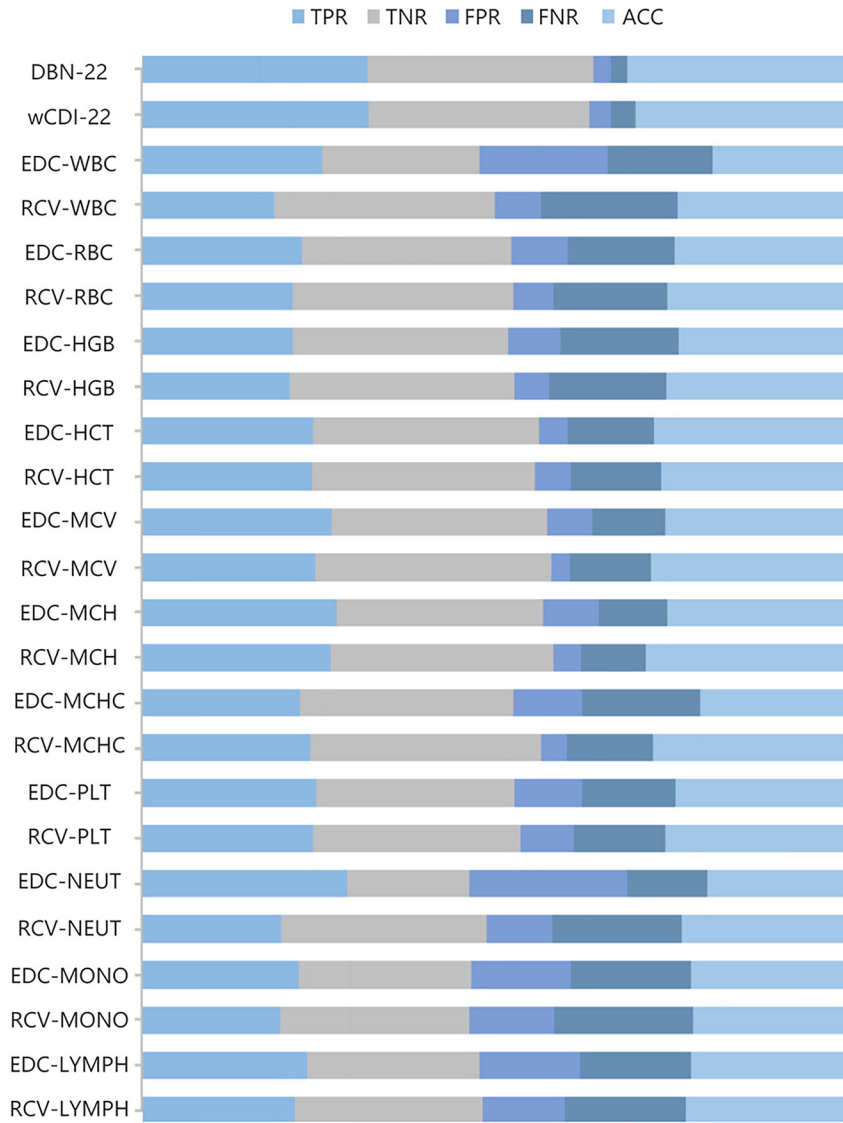


Figure 4: Comparison of DBN method with three optimized DC methods using absolute difference results of male samples.

shown in Supplementary Tables 1 and 2 and in Supplementary Tables 3 and 4. Experimental results illustrated DBN was better than the three statistical methods. Of all test items performed, absolute difference DC yielded higher accuracy than relative difference for all methods. The same simulation study was performed by artificially generating cases of female samples.

Discussion

Our model enabled the accurate detection of sample mix-up in real-world settings, illustrating powerful performance when compared to previous studies [10, 16, 18]. The main reasons for these results were as follows:

- DC methods reported were prone to be affected by the data distribution patterns of test results, DC limits, and the amount of test items.
- Dramatically heterogeneous and extreme results exist in real-world clinical laboratory data and individual biological variations enlarge data fluctuation.
- Assuming analytical variation was ignored, matched data was mainly affected by within-individual biological variation, data distribution pattern and extreme values. Viceversa, mismatched data was mainly affected by between-individual biological variation, data distribution patterns, and extreme values.
- Simple statistical analysis was not use in uncovering cases of sample mix-up. For both DBN and the improved RwCDI, at the first, raw data was filtered by pre-defined rules, and further a series of subsequent

data preprocessing was adopted, mainly including data transformation and removal of extreme values for delta data. The difference was that DBN got rid of extreme values by isolated forest algorithm, while that RwCDI by simple truncation limits. Isolated forest algorithm was a relative robust method to remove extreme values. Its working principle was similar to the density map method. The number of extreme values were able to be adjusted according to the degree of density and balance of data. In this study, isolated forest algorithm in this step removed about 3% of the extreme data, while RwCDI excluded about 1% of the extreme data. For RCV and EDC, the original data only filtered by the first-step rules. The experimental results showed that the accuracy of DBN and RwCDI was 0.9310 and 0.9089 separately. DBN was better than RwCDI and was significantly superior to RCV and EDC.

DC limit setting was the key step to detect sample mix-up. Due to the different control limit settings in various laboratories, the maximum variation in the error detection rate of sample mix-up among laboratories reached up to 76% [9]. In this study, two types of DC control limit setting methods were compared. The control limit for EDC was optimized by a dense grid search within a broad range of 0–200% in steps of 0.1. The control limit for RCV was calculated according to individual biological variation and optimized by adjusting k value or excluding pairs of test results within reference intervals or directly extending the original control limits. Our results illustrated that the accuracy for different test items for EDC after optimization ranged from 0.5825 to 0.7804, while for RCV from 0.5631 to 0.8145, which was similar to the results reported in the literature [18]. The accuracy of EDC and RCV far lagged behind that of DBN. This might be related with method itself. The working principle of both methods was based on simple DC control limits to distinguish error samples from correct samples. Thus, they are difficult to capture nonlinear effects and interaction in real-world clinical scenarios.

Previous studies reported that the amount of test items affected the accuracy of error detection for sample mix-up [19]. Most of DC methods only used a single test item as an input index. If a combination of test items was used as input indexes, ML features would be strengthened. Here k was introduced, which represented the number of test items ($k=5-22$). Our results proved that the accuracy of DBN adopting 22 test items ($k=22$) as input indexes reached up to 0.9310, which was higher than 10 test items ($k=10$). Teppei's study stated that AUC and sensitivity increased

proportionately for test items $k < 10$ but remained almost unchanged for $k > 10$, and the cut off value decreased until $k=10$ and remained unchanged for $k > 10$. This might be related with the way of weighting in the calculation. In Teppei's method [16], a weighting factor was converted by standard deviation of a given test item. But correlations among test items involved in the calculation did not be taken into consideration.

For DBN model established in this study, the accuracy was regarded as the primary evaluation metric. The most basic component of DBN model was a neuron. Neurons receiving output signals from other neurons ($x_1 \dots x_n$) regarded as next input signals, these input signals transferred between neurons by connections with different weights ($\omega_1 \dots \omega_n$). A total input value received by neurons would be compared with a threshold, called θ . Then, the output of neurons was processed by an "activation function" (y) (Figure 1). RCV and EDC were mainly optimized by adjusting DC limits at different strength. Our experimental data showed that EDC was better than RCV in DC limit optimization, but the input signal of the two methods was only a single dimension, that was x_1 . In the Teppei's method, the input signal was multi-dimensional, i.e. $x_1 \dots x_n$. This was similar to DBN method. But Teppei's method was one-way correlation to input signals, the number of weight (ω) was the same as the input dimension, and the size of each weight was related to the dispersion of each input signal x , that was $\omega = \frac{1}{\text{ASD}}$. In our DBN model, input signals were transferred in a multi-layer and cross-structured way, and the number of weight was tremendous and complicated. In general, parameters ω_i and θ obtained by the way of on-going ML. In particular, perceptron (that was, it had only one layer of neurons) had limited learning ability and mainly solved the linear separable problem. For the nonlinear indivisible problem, we needed to consider the use of multi-layer functional neurons. The learning process was actually to adjust the "connection weight" between neurons and the threshold θ of each functional neuron according to the training set data. The results showed that the accuracy of the four methods was $\text{DBN} > \text{RwCDI} \gg \text{EDC} > \text{RCV}$.

The generalization of the model was another important evaluation metrics for assuring a valuable clinical application. In this study, hematology test results were selected due to high testing frequency and high levels of standardization. Data from two laboratories in different hospitals were used to establish our training dataset, validation dataset, and test dataset. The test dataset came from one hospital, the training dataset and the validation dataset set data were from the other hospital. The training

dataset and validation dataset were separated independently to avoid overestimation of the accuracy of unknown data by the established model. The experimental results did show that the accuracy training dataset from Chaoyang Hospital was approximately 93%, equal to the accuracy of test dataset from Long-fu Hospital. In addition, in the process of ML algorithm selection, it was found that both the RF algorithm and the DBN algorithm demonstrated acceptable performance characteristics. The DBN algorithm was slightly better than the RF algorithm on the current dataset. However, in clinical complex scenarios, when the data distribution difference became smaller, RF algorithm might be prone to worse, while DBN would represent stronger generalization ability.

In conclusion, our data demonstrate that utilizing the full panel of all available hematology test result items provides more significant information for sample mix-up detection by ML than what is offered by a single test item input. The DC method based on the DBN has demonstrated highly effective sample mix-up identification performance in real-world clinical settings.

Acknowledgments: The authors thank Mr. Chen Chao of Inner Mongolia Wesure Data Technology Co., Ltd for administrative management and hardware support for our experiment.

Research funding: Major innovation support project for high-tech industries, Beijing Chaoyang District Science and Technology Plan (CYGX2112), entitled by “constructing a newly artificial intelligence quality control platform for traceable testing results by exclusive routine test item data”; Excellence Project of key clinical specialty in Beijing; The 1351 Talent Training Plan (grant numbers CYMY-2017-01).

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Competing interests: Authors state no conflict of interest.

Informed consent: Not applicable.

Ethical approval: The project was approved by the local hospital Ethics Committee.

References

1. Da Rin G. Pre-analytical workstations: a tool for reducing laboratory errors. *Clin Chim Acta* 2009;404:68–74.
2. Lundberg GD. Acting on significant laboratory results. *JAMA* 1981; 245:1762–3.

3. Plebani M. The detection and prevention of errors in laboratory medicine. *Ann Clin Biochem* 2010;47(2 Pt):101–10.
4. Lippi G, Guidi GC, Mattiuzzi C, Plebani M. Preanalytical variability: the dark side of the moon in laboratory testing. *Clin Chem Lab Med* 2006;44:358–65.
5. Lippi G, Bowen R, Adcock DM. Re-engineering laboratory diagnostics for preventing preanalytical errors. *Clin Biochem* 2016;49:1313–4.
6. Nosanchuk JS, Gottmann AW. CUMS and delta checks. A systematic approach to quality control. *Am J Clin Pathol* 1974;62: 707–12.
7. Ladenson JH. Patients as their own controls: use of the computer to identify “laboratory error”. *Clin Chem* 1975;21:1648–53.
8. Randell EW, Yenice S. Delta checks in the clinical laboratory. *Crit Rev Clin Lab Sci* 2019;56:75–97.
9. Tan RZ, Markus C, Choy KW, Doery JCG, Loh TP. Optimized delta check rules for detecting misidentified specimens in children. *Am J Clin Pathol* 2020;153:605–12.
10. Tan RZ, Markus C, Loh TP. An approach to optimize delta checks in test panels – the effect of the number of rules included. *Ann Clin Biochem* 2020;57:215–22.
11. Mitchell TM. Does machine learning really work? *AI Mag* 1997;18: 11–20.
12. Zhao BW, You ZH, Hu L, Guo ZH, Wang L, Chen ZH, et al. A novel method to predict drug-target interactions based on large-scale graph representation learning. *Cancers (Basel)* 2021;13:2111.
13. Tan RZ, Markus C, Loh TP. Relationship between biological variation and delta check rules performance. *Clin Biochem* 2020; 80:42–7.
14. Ricós C, Alvarez V, Cava F, García-Lario JV, Hernández A, Jiménez CV, et al. Current databases on biological variation: pros, cons and progress. *Scand J Clin Lab Invest* 1999;59:491–500.
15. Ichihara K, Kawai T. Determination of reference intervals for 13 plasma proteins based on IFCC international reference preparation (CRM470) and NCCLS proposed guideline (C28-P, 1992): a strategy for partitioning reference individuals with validation based on multivariate analysis. *J Clin Lab Anal* 1997;11: 117–24.
16. Yamashita T, Ichihara K, Miyamoto A. A novel weighted cumulative delta-check method for highly sensitive detection of specimen mix-up in the clinical laboratory. *Clin Chem Lab Med* 2013;51:781–9.
17. Lippi G, Blanckaert N, Bonini P, Green S, Kitchen S, Palicka V, et al. Causes, consequences, detection, and prevention of identification errors in laboratory diagnostics. *Clin Chem Lab Med* 2009;47:143–53.
18. Hong J, Cho EJ, Kim HK, Lee W, Chun S, Min WK. Application and optimization of reference change values for delta checks in clinical laboratory. *J Clin Lab Anal* 2020;34: e23550.
19. Clinical and Laboratory Standards Institute (CLSI). Use of Delta Checks in the Medical Laboratory, CLSI guideline EP33. 1st ed. Clinical and Laboratory Standards Institute, 950 West Valley Road, Suite 2500, Wayne, Pennsylvania 19087 USA, 2016.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/cclm-2021-1171>).