IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 7, 2021, accepted April 18, 2021, date of publication April 22, 2021, date of current version May 3, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3074929

An Efficient Anomaly Detection for High-Speed Train Braking System Using Broad Learning System

CHONG WANG AND JIE LIU

School of Reliability and System Engineering, Beihang University, Beijing 100083, China

Corresponding author: Jie Liu (liujie805@buaa.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 52005027.

ABSTRACT For a high-speed train, the braking system plays an essential role in safe transportation. Efficient state monitoring and anomaly detection may provide useful information for real-time decisions. With large amount of monitoring data, data-driven methods, especially deep learning methods are widely adopted for anomaly detection in various industrial applications. Although deep learning methods have advantages in discovering non-linear relations among complex and high-dimensional data, the large amount of hyperparameters can be hardly well tuned with a high computational burden for onboard computers. Therefore, in this work an efficient online anomaly detection model based on Broad Learning System (BLS) is established for detecting anomalies in the braking system. Furthermore, considering the intrinsic imbalanced data size on anomaly and normal states, the cost-sensitive learning method is integrated in the BLS model, for the first time. The proposed model is evaluated on real data collected from a high-speed train operating for one year, with respect to two performance metrics, i.e. G-mean and F1-score. Comparisons with benchmark neural networks and the combinations of sampling methods and BLS are also considered in this work.

INDEX TERMS Anomaly detection, braking system, broad learning system, cost-sensitive learning, deep learning, high-speed train, highly imbalanced data.

I. INTRODUCTION

The braking system ensures the effective deceleration of high-speed trains during emergencies and regular stops. With the development of transportation technology, the train speed increases continuously, and the safety of the braking system is becoming an ultimate important issue [1]. Fault diagnosis, which includes anomaly detection, fault location and fault identification, is a key technology for ensuring the system safety. Furthermore, anomaly detection is considered as a primary task in fault diagnosis [2].

According to the braking mechanism, the braking system can be divided into 3 categories: air braking, vacuum braking, electro-pneumatic braking, etc. The braking system adopted in most high-speed trains is electro-pneumatic, which integrates the air braking and the electric braking, for routine and emergent braking (a simplified schematic diagram is shown in Figure 1). The air brake converts the digital commands into the air pressure in the brake cylinder, and the electric brake (E-P brake in DK-1 electro-pneumatic braking system) converts the traction motor into a generator. In an electro-pneumatic braking, electric brake is given priority producing most braking force.

There are many factors affecting the safety of a braking system, such as equipment aging, miss-operation, alternative exchange of environment, etc. By analyzing these factors and assessing the health status, proper maintenance can reduce the life-cycle cost by improving the safety and availability of a braking system. However, accurate and efficient anomaly detection is progressively difficult with the large amount of data increasing.

In the published work, various methods have been proposed for anomaly detection, and these methods can be categorized into simulation-aided methods [3], [4], expert systems [5], [6] physics-of-failure models [7], [8], and data-driven models [9], [10]. The braking system is designed based on the fault-oriented safety principle, resulting in a complex structure [11]. It is quite difficult to analyze the

The associate editor coordinating the review of this manuscript and approving it for publication was Baoping Cai.

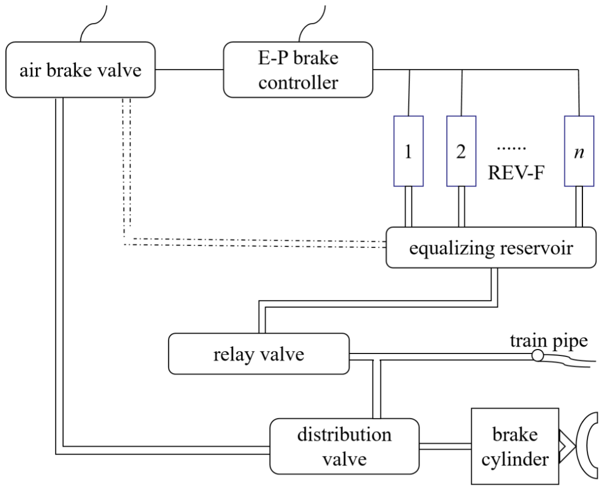


FIGURE 1. The structure of a DK-1 electro-pneumatic braking system (single line represents circuits, double line represents pipeline and double dotted line represents air level control pipeline).

operating mechanism using traditional algorithms, such as simulation-aided methods, expert systems and physics-of-failure models for online anomaly detection. In practice, numerous sensors are integrated in the braking system. The implemented sensors make it possible to collect enough data related to the system working conditions and, thus, the data-driven models that are independent of prior knowledge on the target system can be adopted.

Recently, data-driven methods, especially the Multi-Variant Analysis (MVA) and machine learning methods, receive intensive attention in high-speed train anomaly detection [12]. The MVA-based methods, such as PCA [13] and multi-mode kernel PCA [14], are used for extracting the hidden information. Machine learning methods, especially supervised learning are widely adopted for fault detection. These methods include random forest [15], Support Vector Machine (SVM) [16], [17], Bayesian-Networks [18]–[20], deep learning methods [21], etc.

Among them, deep learning methods are becoming quite popular in anomaly detection and have achieved breakthrough success in many applications. In [21], a Deep Neural Network (DNN) for bogie fault diagnosis of high-speed train based on vibration signal was proposed. Moreover, a Convolutional Neural Network (CNN)-based cascade model was proposed in [22] for defect detection and location in insulators. Reference [13] verified the effectiveness of a fault diagnosis deep belief network (DBN) for high-speed train onboard equipment. In [23], a Neuro-adaptive fault-tolerant control model under traction-braking failures using self-structuring neural networks was established. Reference [24] designed an intelligent fault diagnosis model based on deep neural network (DNN) for high-speed train bogies. Most deep learning methods are powerful in tackling complex structures of high-dimensional data [25], [26]. However, for reaching high accuracy with a deep learning model, a large number of hyperparameters need to be tuned in the training process, making the training process time-consuming [31] This makes it difficult to implement real-time training anomaly detection based on deep learning. In fact, most of the current anomaly detection systems are in the form of discrete training and online diagnosis, which are difficult to update in an online manner.

Therefore, some variations in hierarchical structure [27], [28] or ensembles [29], [30] are proposed to improve the model training efficiency. Among them, Broad Learning System (BLS) is proposed based on the conception of Random Vector Functional Link Neural Network (RVFLNN), and it can be updated effectively and efficiently in many applications [31], [32]. A fault diagnosis model based on BLS and principal component analysis was established in [33] for a rotor system. For the fault diagnosis of aeroengine wear, an ensemble of BLS models was considered in [34]. In [35], a fault diagnosis method was designed for rolling bearings by integrating variational mode decomposition and Hilbert transform in the BLS. Similar strategy was considered in [36] with the feature incremental broad learning and singular value decomposition.

Taking into account the requirements on timely response to the environment change and on the reduction of computation cost, BLS is adopted for anomaly detection in a braking system in this work. The high reliability of the high-speed train brings the fact that most of the monitoring data concern the normal condition, i.e. the collected dataset is highly imbalanced. The decision hyperplanes of data-driven methods may be biased to the majority class, causing a low anomaly detection rate. Thus, class-imbalance problem should be properly tackled in the BLS framework. Different from assigning the same cost to different classification errors in state-of-art references on BLS which lead to the overfitting on the normal state, this work, for the first time, combines the cost-sensitive learning with BLS to establish an effective and efficient anomaly detection model. In combination with BLS, the other strategies for tackling imbalanced data are also considered as benchmark methods, along with different neural networks.

The remainder of this paper is organized as follows. The proposed anomaly detection model for high-speed train braking system is described in Section 2. And, Section 3 is the application results in real monitoring data of a high-speed train. Conclusions and some further research directions are drawn in Section 4.

II. BLS FOR ANOMALY DETECTION WITH IMBLANCED DATA

BLS is a single-layer incremental neural network based on RVFLNN and Single-Layer Feedforward Neural Network (SLFNN). BLS maps the input data to a series of random feature spaces and determines the output weights through an optimized least squares method. Moreover, the model is optimized through incremental learning without iterative calculations, which greatly reduces the computation time [31].

A. BLS WITHOUT INCREMENTAL LEARNING

BLS without incremental learning is very similar to a RVFLNN, and its structure is shown in Figure 2.

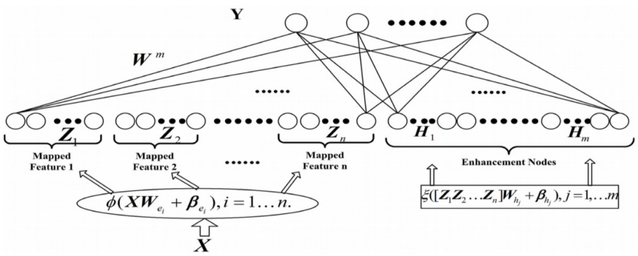


FIGURE 2. **The basic structure of BLS [31].**

Assume that training data is $(X, Y) \in R^{K\times(P+Q)}$, where K is the number of data rows, P and Q are the dimensions of X and Y respectively. According to (1), a series of random feature spaces $\mathrm{Z}_1, \mathrm{Z}_2, \cdots, \mathrm{Z}_\mathrm{n}(Z^n \triangleq \{\mathrm{Z}_1, \mathrm{Z}_2, \cdots, \mathrm{Z}_\mathrm{n}\})$ can be calculated with the input data X.

$$Z_i = \phi_i(XW_{ei} + \beta_{ei}), \quad i = 1, 2, \ldots, n \tag{1}$$

In (1), Z_i is the i th component in Z^n, i.e. the i th mapped feature, and W_{ei} and β_{ei} are separately the optimal input weights matrix calculated by sparse self-encoding and the corresponding biases matrix. The initial values of weights and the biases matrix are randomly generated, and the number of nodes in Z_i is coherent with the dimensions of the weights and biases, i.e. the mapped feature Z_i has ν nodes for $X \in R^{K\times P}$ and $W_{ei} \in R^{P\times\nu}$.

Furthermore, in order to improve the training speed, the enhancement nodes are generated by Z^n in a manner of group by group. Among them, the j th enhancement nodes group is:

$$E_j = \xi_j(Z^n W_{hj} + \beta_{hj}), \quad j = 1, 2, \ldots, m \tag{2}$$

with ξ_j being the non-linear activation function, W_{hj} and β_{hj} being the random matrix and the deviation matrix of the j th enhancement nodes group, respectively. The initial values of these two matrices are also generated randomly.

Represent the outputs of the enhancement layer as $E^m \triangleq \{E_1, E_2, \ldots, E_m\}$. Then the matrix $H = [Z^n \,|E^m]$ with the splicing of matrix Z^n and E^m becomes the actual input of the system. Consequently, the final output of a BLS can be expressed as:

$$Y = HW = [Z^n \left|E^m\right.]W \tag{3}$$

where $Y \in R^{K\times Q}$ is the output of the model. In a binary classification task, $Y \in R^K$. The matrix W is the output weights that connects $H = [Z^n\,|E^m]$ to the output layer. Therefore, it can be optimized by the least square method.

However, the generalization error of the model optimized with least square estimation is normally large. In order to control the structural complexity of the network and to improve its generalization capability, the l_2-norm regularization term is added in the loss function. Then the output weights matrix W can be optimized with the following ridge regression problem:

$$\min_W f(W) = \min \|HW - Y\|_F^2 + c\,\|W\|_F^2 \tag{4}$$

where c is a trade-off regularization parameter on W.

The closed-form solution of (4) is given with Moore-Penrose pseudo inverse, as shown in (5) [37]

$$W = \begin{cases} H^T(cI + HH^T)^{-1}Y, & K < L \\ (cI + H^T H)^{-1}H^T Y, & K \geq L \end{cases} \tag{5}$$

where, K is the number of data rows, and $L = n\nu + m\eta$ is the number of hidden-layer nodes.

In a BLS model without incremental learning, after the training process is completed, the randomly weights W_{ei}, W_{hj} and the output weights matrix W are fixed. And the prediction result of a test sample x_s is:

$$\hat{y}_s = H_s W \tag{6}$$

where Hs is the splicing feature matrix of x_s.

B. INCREMENTAL LEARNING FOR BLS

For BLS, if the network cannot achieve the required accuracy, additional enhancement nodes can be inserted to improve the network, following the idea of incremental learning.

Assuming the existent hidden layer is $H^{\mathrm{m}} = [Z^n\,|E^m]$, When p additional enhancement nodes are inserted in the network (as illustrated in Figure 3), p columns will be added to the matrix H^{m} and the new hidden layer is defined as:

$$H^{\mathrm{m}+1} \triangleq [H^m \left|\xi(Z^n\ W_{h_{m+1}} + \beta_{h_{m+1}})\right.] \tag{7}$$

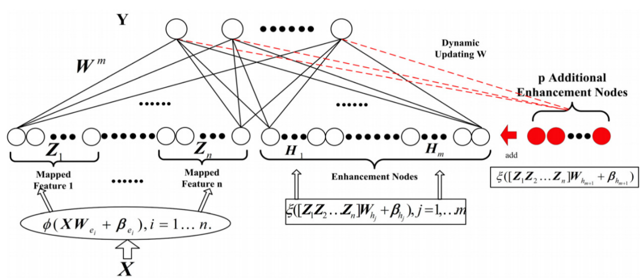


FIGURE 3. **Illustration of incremental learning algorithm [31].**

According to the incremental learning strategy of BLS, the new output weights matrix can be calculated as follow:

$$W^{m+1} = \begin{bmatrix} W^m - DB^T Y \\ B^T Y \end{bmatrix} \tag{8}$$

with $D = (H^m)^+\xi(Z^n W_{h_{m+1}} + \beta_{h_{m+1}})$,

$$B^T = \begin{cases} (C)^+, & C \neq 0 \\ (1 + D^T D)^{-1} D^T (H^m)^+, & C = 0 \end{cases}$$

and $C = \xi(Z^n W_{h_{m+1}} + \beta_{h_{m+1}}) - H^m D$.

During the incremental learning, if additional enhancement nodes are needed, the new output weights matrix

W^{m+1} can be obtained by calculating the pseudo inverse of the additional enhancement nodes instead of calculations from scratch. It is unnecessary to retrain the entire network from the beginning, thus, the training process is accelerated tremendously [31].

C. COST SENSITIVE LEARNING FOR BLS

The classification hyperplane of the previous BLS model may be biased to the majority class for imbalanced data, causing a low classification accuracy on minority class. However, in most situations, the minority class is the one of interest to the practitioners. Thus, it is quite important to tackle properly imbalanced data. Currently, there are three main methods to deal with imbalanced data [38]:

1) Convert the imbalanced data to balanced data by adjusting data distribution or select features that better represent the unbalanced dataset. Under-sampling majority classes and over-sampling minority classes are two popular directions.

2) Optimize the structure or parameters of the algorithm. For example: modify the classification threshold, add cost sensitive learning, or use fuzzy function to optimize the algorithm [39].

3) Establish a combination classifier by integrated learning. Among them, cost-sensitive learning is a common method, which avoids adding virtual samples which may destroy the prior distributions of variables. It has been successfully integrated in may many methods, such as support vector machines [17], neural networks [40], Bayesian methods [41], etc.

In this paper, based on an in-depth analysis of BLS, cost-sensitive learning is integrated in the training process of BLS and the new method is named as CS-BLS. Cost-sensitive learning makes abnormal data receive more attention during model training, i.e. the misclassification cost of an abnormal sample is greater than that of a normal one. Moreover, the weight matrix is adjustable according to different applications.

When the training data is $(X, Y) \in R^{K\times(P+Q)}$, according to cost-sensitive learning, a weights matrix Λ that represents the importance of different classes can be established. Here Λ is a K-dimensional diagonal matrix taking a general form as follows:

$$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_K \end{bmatrix} \tag{9}$$

where K is still the number of data rows, λ_i is the weight for the i th row, and same weight is set for the same category.

When cost-sensitive learning is integrated in BLS, the n groups of mapped features Z'^n and m groups of enhancement nodes E'^m are:

$$Z'_i = \phi_i(\Lambda X W_{ei} + \beta_{ei}), \quad i = 1, 2, \ldots, n \tag{10}$$

$$E'_j = \xi_j(Z'^n W_{hj} + \beta_{hj}), \quad j = 1, 2, \ldots, m \tag{11}$$

The actual input of the system is converted to $H' = [Z'^n \,|E'^m]$, and the final output is:

$$Y = H'W^m = [Z'^n \,|E'^m]W^m \tag{12}$$

The output weights matrix W^m can be obtained by ridge regression, and the result is as follows:

$$W^m = \begin{cases} H'^T(cI + H'H'^T)^{-1}Y, & K < L \\ (cI + H'^T H')^{-1}H'^T Y, & K \geq L \end{cases} \tag{13}$$

here K is the number of data rows, $L = n\nu + m\eta$ is the number of hidden-layer nodes.

Similarly, when additional enhancement W^{m+1} nodes are inserted to the system, the new weights matrix can also be obtained:

$$W^{m+1} = \begin{bmatrix} W^m - D'B'^T Y \\ B'^T Y \end{bmatrix} \tag{14}$$

where $D' = (H'^m)^+ \xi(Z'^n W_{h_{m+1}} + \beta_{h_{m+1}})$,

$$B'^T = \begin{cases} (C')^+, & C' \neq 0 \\ (1 + D'^T D')^{-1} D'^T (H'^m)^+, & C' = 0 \end{cases}$$

and $C' = \xi(Z'^n W_{h_{m+1}} + \beta_{h_{m+1}}) - H'^m D$.

III. APPLICATION RESULTS

Based on the monitoring dataset of a high-speed train braking system within one year, a comparative experiment is carried out in this section to verify the effectiveness of the proposed method. The key steps of the experiment are as demonstrated in Figure 4. Conventional neural networks i.e. ANN and CNN are considered as benchmark methods. All the experiments are carried out using PYTHON (3.6) on a 1.60 GHz intel(R) Core(TM) i5-8250U CPU with 7.86 GB RAM.

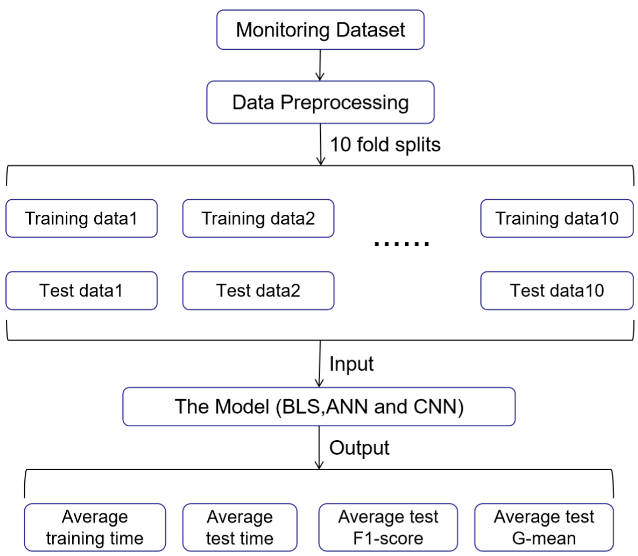


FIGURE 4. Flow chart of comparative experiment.

The braking system considered in this experiment is an electronically controlled electro-pneumatic brake system.

The brake control system adopts composite braking and has a variety of braking modes and states. In order to comprehensively monitor the factors affecting the health status of braking system, sensors are arranged at the key positions of the system, and the environment factors are also recorded. The collected raw data contains 43 variables related to abnormal conditions of the braking system. These 43 variables are obtained by combining practical experience and fault mechanism, and all variables are equally considered when establishing anomaly detection model. These variables include train-level conditions, braking system-level conditions and operating environment-level conditions, such as GPS position, travel speed, operation time, operation mode, line current, line voltage, battery voltage, external power voltage, brake state, achieved braking force, internal temperature, external temperature, etc. Bound by confidentiality agreements, these variables cannot be exhaustively listed here.

First of all, data cleaning is necessary, for example, the reconstruction of missing values, data normalization, etc. Data normalization is performed to balance the dimensional differences among different numeric variables. Furthermore, according to the data requirements of the models, some variables in the raw data need to be converted to numerical values. After data processing, all 43 variables are converted to values between [0,1]. The status label represents whether the braking system is in abnormal state. While the abnormal state is recorded as 1, the normal state is recorded as 0.

There are 28837 normal state data points and 159 abnormal state data points. The data is severely imbalanced, and we focus on the detection accuracy on abnormal state. In this case, if the overall accuracy is used as the model evaluation index, the model effect will inevitably depend on the normal category and is rarely affected by the fault category. In fact, we only pay attention to whether the prediction result of the fault category is accurate and complete. Therefore, in this paper, instead of using the overall accuracy as the model performance metrics, F1-score and G-mean are adopted.

With TP, TN, FP and FN as the number of true positive, true negative, fault positive and fault negative, F1-score and G-mean can be calculated as follows [42]:

$$F1 - score = 2 * \frac{precision * recall}{precision + recall} \tag{15}$$

$$G - mean = \sqrt{TPR * TNR} \tag{16}$$

where TP is the number of samples correctly predicted to be positive; TN is the number of samples correctly predicted to be negative; FP is the number of samples wrongly predicted to be positive; and FN is the number of samples wrongly predicted to be negative, then $precision = TP/(TP + FP)$, $recall = TPR = TP/(TP + FN)$, $TNR = TN/(TN + FP)$.

In order to reduce the influence of random factors, the average accuracy and running time of 10-fold cross-validation are listed in this paper. And, to get the approximate local optimal solution, the network depth, the number of nodes and the cost-sensitive weight are gradually increased until the accuracy on the verification dataset is basically unchanged or decreased. With 10-fold cross-validation, the outputs of a model include the average training time, the average test time, the average test F1-score and the average test G-mean. Here, the training time refers to the total time required to get the final model under training and optimization, and the test time refers to the time needed to use the trained model for prediction.

Figure 5 is a two-dimensional projection of the dataset based on t-distributed Stochastic Neighbor Embedding (t-SNE), and the red 1.0 represents the projection of abnormal samples, the gray 0.0 represents the projection of normal ones. The t-SNE algorithm is one of the most commonly used and effective techniques in the exploratory analysis of high-dimensional data. This method, which converts the similarity between data points into probabilities, could achieve the visualization of high-dimensional data by projecting them into 2-D or 3-D space [43]. When using all the data, the anomalous samples are totally covered by the normal ones. Therefore, only 10% of the collected normal samples are displayed in the figure. After comparisons, the t-SNE projection distribution of the 10% normal samples used in the figure is roughly consistent with that of the whole dataset.

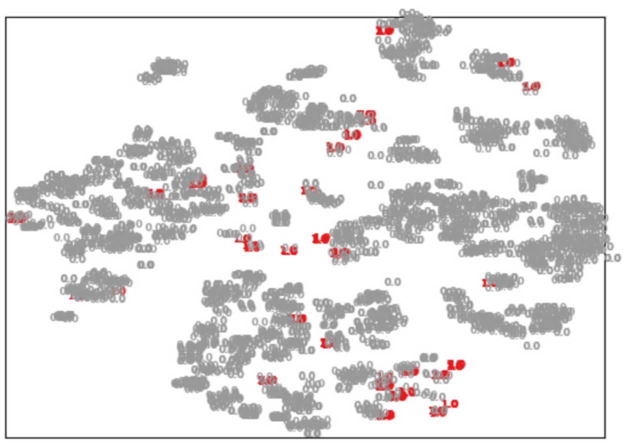


FIGURE 5. T-SNE of the dataset.

From Figure 5, one may see that the data is seriously imbalanced. Moreover, the data of normal state and abnormal state are largely overlapped, which makes it difficult to achieve both high precision and recall values on the abnormal state at the same time. Consequently, F1 score and G-mean are incapable to reach the maximum value at the same time. In this case, by simulating the form of F-score [42], the weighted harmonic mean (denoted as F-G) of the F1-score and the G-mean is selected as a comprehensive indicator to evaluate the generalization accuracy of the algorithm.

$$F - score = \frac{(1 + \beta^2)precision * recall}{(\beta^2 * precision) + recall} \tag{17}$$

where β is the adjustment weight. When $\beta = 1$, the precision and the recall are almost equally important. And, when $\beta > 1$, the weight of recall is higher.

In this anomaly detection problem, the precision and the recall of abnormal state is much more important than those of normal state, thus we set the weight of F1-score higher than that of G-mean.

$$F-G = 5 * \frac{F1 - score * G - means}{F1 - score + 4 * G - means} \tag{18}$$

First, we compare the influence of under-sampling, over-sampling and cost-sensitive learning on BLS (as shown in Table 1).

TABLE 1. Comparison with different data processing methods.

	Cost-sensitive learning	Under-sampling	Over-sampling	Untreated
F1-score	0.89524	0.47619	0.54545	0.60465
G-mean	0.94894	0.6447	0.8628	0.78429
F-G	0.90549	0.50246	0.58876	0.63368
average training time(s)	1.917	0.396	2.169	1.847
average test time(s)	0.085	0.085	0.086	0.081

Here, under-sampling is achieved by random-choice, i.e. randomly delete some normal state samples in the training data to make the number of samples of different categories close. Over-sampling in the experiment refers to the regular SMOTE-fit [44] which can create synthetic minority class samples. The size of sampling and interpolation samples are adjusted adaptively in the training data i $(1 \leq i \leq 10)$. Figure 6 and 7 are respectively two-dimensional t-SNE projections of the under-sampling dataset and the over-sampling dataset. The dataset is balanced by these two technologies. When the under-sampling technique is used, the training data is insufficient to obtain a high-precision model. And, the over-sampling technology may lead to more serious class crossing. Therefore, their generalization effect is worse than the original BLS.

Considering both accuracy and training time, the results in Table 1 implied that for BLS, cost-sensitive learning has a great advantage in reducing the influence of between-class imbalance problem in the case study.

Then, the results achieved by different algorithms are shown in Table 2.

It can be inferred from Table 2 that the generalization accuracies of different methods are comparable, while the training time of CS-BLS is much shorter than that of the benchmark methods. It takes only 1.917 seconds to retrain the BLS model. In practical application, the model may be updated promptly to ensure the accuracy of prediction all the time, and the operation and maintenance cost will be greatly reduced. Therefore, CS-BLS is suitable for complex systems that may often encounter unexpected situations and require

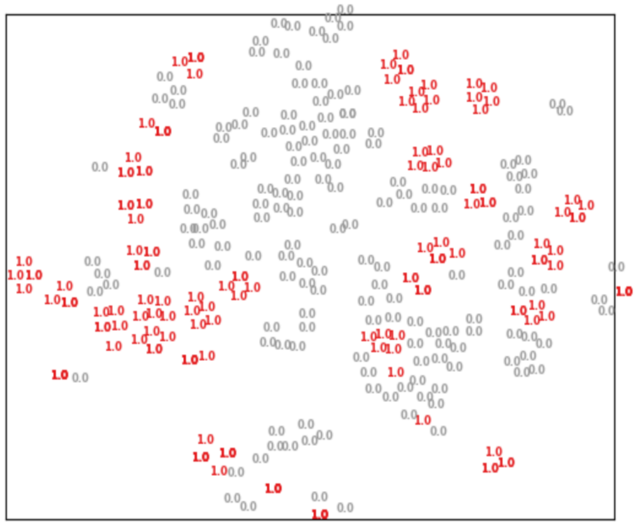


FIGURE 6. T-SNE of the under-sampled dataset.

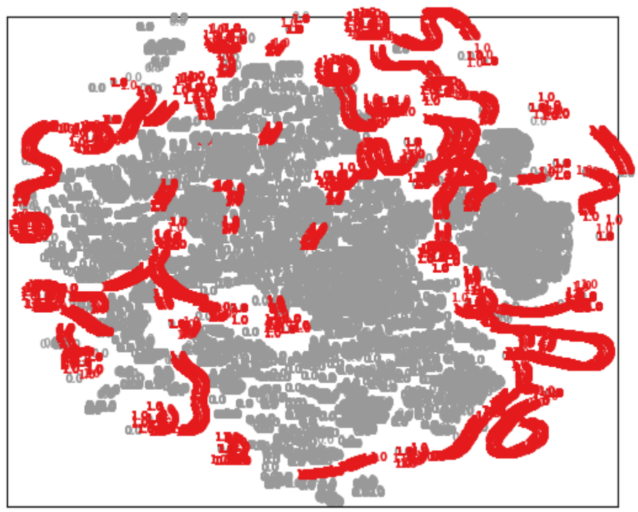


FIGURE 7. T-SNE of the over-sampled dataset.

high reliability, such as high-speed rail braking systems in this paper. When the system reliability requirements are not particularly high or the structure is relatively simple, in order to save maintenance costs, it may be better to use traditional fault diagnosis methods, such as preventive measure, regular maintenance, offline fault diagnosis, and post-repair. This experiment infers that the requirements for real-time and accurate monitoring of the high-speed train braking system can be achieved by the proposed CS-BLS. Simultaneously, we can see that, cost-sensitive can significantly optimize the effect of BLS, but has a small improvement on the ANN and CNN models.

Moreover, the experiment is only based on the monitoring data of a high-speed train operated within one year. It may be predicted that when the amount of data is larger, the high efficiency of the proposed CS-BLS model will be more important for online anomaly detection.

TABLE 2. Comparison of different models.

	BLS	CS-BLS	ANN	CNN	CS-ANN	CS-CNN
F1-score	0.60465	0.89524	0.87999	0.91667	0.92683	0.93506
G-mean	0.78429	0.94894	0.95743	0.95743	0.92932	0.94839
F-G	0.63368	0.90549	0.89446	0.92454	0.92733	0.93770
average training time(s)	1.847	1.917	232.652	76.899	241.060	78.934
average test time(s)	0.081	0.085	0.019	0.053	0.029	0.071

IV. CONCLUSION

The braking system plays a vital role in safe transportation of high-speed trains. An efficient anomaly monitoring system may greatly assist the train operation and maintenance as a health state information source. Based on the BLS, this paper establishes a real-time anomaly detection model for the high-speed train braking system. Considering the high imbalance between the data sizes on normal and abnormal conditions, the BLS is integrated with adaptive cost-sensitive learning, i.e. CS-BLS. In comparison with classical neural networks and the combinations of sampling methods and BLS, it can be inferred that the method proposed in this paper achieves comparable accuracy in a much shorter time in the case study. CS-BLS can process highly imbalanced data conveniently and effectively, and can be updated quickly and timely when necessary. This study provides a new idea for real-time anomaly detection of high reliability systems.

We also notice that online fault prediction based on BLS has certain research prospects for the time series data of high-speed train. Moreover, thorough data preprocessing, e.g. feature selection can be added to reduce the impact of highly imbalance data. These thoughts can be further studied in the future. Similarly, CS-BLS also provides an alternative method for real-time fault diagnosis and prediction of other complex systems with imbalanced data.

REFERENCES

[1] E. Galardi, E. Meli, D. Nocciolini, L. Pugi, and A. Rindi, "Development of efficient models of magnetic braking systems of railway vehicles," *Int. J. Rail Transp.*, vol. 3, no. 2, pp. 97–118, Apr. 2015.

[2] H. Ji, X. He, J. Shang, and D. Zhou, "Incipient fault detection with smoothing techniques in statistical process monitoring," *Control Eng. Pract.*, vol. 62, pp. 11–21, May 2017.

[3] C.-G. Kang, "Analysis of the braking system of the Korean high-speed train using real-time simulations," *J. Mech. Sci. Technol.*, vol. 21, no. 7, pp. 1048–1057, Jul. 2007.

[4] C.-T. Lin, "Analyzing the effect of imperfect debugging on software fault detection and correction processes via a simulation framework," *Math. Comput. Model.*, vol. 54, nos. 11–12, pp. 3046–3064, Dec. 2011.

[5] S. Lee, G. Kim, and S. Kim, "Self-adaptive and dynamic clustering for online anomaly detection," *Expert Syst. Appl.*, vol. 38, no. 12, pp. 14891–14898, Nov. 2011.

[6] M. Nakano, A. Takahashi, and S. Takahashi, "Fuzzy logic-based portfolio selection with particle filtering and anomaly detection," *Knowl.-Based Syst.*, pp. 113–124, 2017.

[7] N. Patil, S. Menon, D. Das, and M. Pecht, "Anomaly detection of non punch through insulated gate bipolar transistors (IGBT) by robust covariance estimation techniques," in *Proc. 2nd Int. Conf. Rel., Saf. Hazard Risk-Based Technol. Phys. Failure Methods (ICRESH)*, Dec. 2010, pp. 68–72.

[8] Z. Mao, G. Tao, B. Jiang, and X.-G. Yan, "Adaptive compensation of traction system actuator failures for high-speed trains," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 11, pp. 2950–2963, Nov. 2017.

[9] M. A. Zuluaga, E. J. D. Leyton, M. H. Hoyos, and M. Orkisz, "Feature selection for SVM-based vascular anomaly detection," in *Proc. Int. MICCAI Workshop Med. Comput. Vis.*, 2018, pp. 141–152.

[10] M. Du, F. Li, G. Zheng, and V. Srikumar, "Anomaly detection and diagnosis from system logs through deep learning," in *Proc. ACM SIGSAC Conf. (ACM)*, 2017, pp. 1285–1295.

[11] H. Zhun-Qing, P. Jun-Bin, C. Wei-Dong, *EMU Braking System*. Beijing, China: Beijing Jiaotong Univ. Press, 2012.

[12] H. Chen and B. Jiang, "A review of fault detection and diagnosis for the traction system in high-speed trains," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 2, pp. 450–465, Feb. 2020.

[13] J. Yin and W. Zhao, "Fault diagnosis network design for vehicle on-board equipments of high-speed railway: A deep learning approach," *Eng. Appl. Artif. Intell.*, vol. 56, pp. 250–259, Nov. 2016.

[14] H. Chen, B. Jiang, N. Lu, and Z. Mao, "Multi-mode kernel principal component analysis-based incipient fault detection for pulse width modulated inverter of China Railway High-speed 5," *Adv. Mech. Eng.*, vol. 9, no. 10, pp. 1–12, Oct. 2017.

[15] C. Dai, Z. Liu, K. Hu, and K. Huang, "Fault diagnosis approach of traction transformers in high-speed railway combining kernel principal component analysis with random forest," *IET Electr. Syst. Transp.*, vol. 6, no. 3, pp. 202–206, Sep. 2016.

[16] J. Liu, Y. Hu, and S. Yang, "A SVM-based framework for fault detection in high-speed trains," *Measurement*, vol. 172, Feb. 2021, Art. no. 108779.

[17] J. Liu, Y.-F. Li, and E. Zio, "A SVM framework for fault detection of the braking system in a high speed train," *Mech. Syst. Signal Process.*, vol. 87, pp. 401–409, Mar. 2017.

[18] B. Cai, H. Liu, and M. Xie, "A real-time fault diagnosis methodology of complex systems using object-oriented Bayesian networks," *Mech. Syst. Signal Process.*, vol. 80, pp. 31–44, Dec. 2016.

[19] B. Cai, Y. Liu, and M. Xie, "A dynamic-Bayesian-network-based fault diagnosis methodology considering transient and intermittent faults," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 1, pp. 276–285, Jan. 2017.

[20] B. Cai, H. Fan, X. Shao, Y. Liu, G. Liu, Z. Liu, and R. Ji, "Remaining useful life re-prediction methodology based on Wiener process: Subsea christmas tree system as a case study," *Comput. Ind. Eng.*, vol. 151, Jan. 2021, Art. no. 106983.

[21] J. M. Yan, Y. Zhao, and Z. H. Guo, "Vibration signal analysis and fault diagnosis of bogies of the high-speed train based on deep neural networks," *J. Vibroeng.*, vol. 19, no. 4, pp. 2456–2474, Jun. 2017.

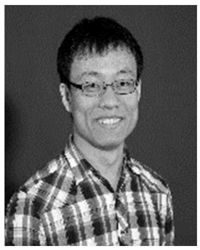
[22] X. Tao, D. Zhang, Z. Wang, X. Liu, H. Zhang, and D. Xu, "Detection of power line insulator defects using aerial images analyzed with convolutional neural networks," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 50, no. 4, pp. 1486–1498, Apr. 2020.

[23] R.-Z. Gao, Y.-J. Wang, J.-F. Lai, and H. Gao, "Neuro-adaptive faulttolerant control of high speed trains under traction-braking failures using self-structuring neural networks," *Inf. Sci.*, vols. 367–368, pp. 449–462, Nov. 2016.

[24] H. Hu, B. Tang, X. Gong, W. Wei, and H. Wang, "Intelligent fault diagnosis of the high-speed train with big data based on deep neural networks," *IEEE Trans. Ind. Informat.*, vol. 13, no. 4, pp. 2106–2116, Aug. 2017.

[25] Y. Lecun, Y. Bengio, and G. Hinton, ''Deep learning,'' *Nature*, vol. 521, no. 7553, p. 436, 2015.
[26] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
[27] J. Tang, C. Deng, and G.-B. Huang, ''Extreme learning machine for multilayer perceptron,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 4, pp. 809–821, Apr. 2016.
[28] M. Gong, J. Liu, H. Li, Q. Cai, and L. Su, ''A multiobjective sparse feature learning model for deep neural networks,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 12, pp. 3263–3277, Dec. 2015.
[29] C. L. P. Chen, C.-Y. Zhang, L. Chen, and M. Gan, ''Fuzzy restricted Boltzmann machine for the enhancement of deep learning,'' *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 6, pp. 2163–2173, Feb. 2015.
[30] Z. Yu, H. Chen, J. Liu, J. You, H. Leung, and G. Han, ''Hybrid K-nearest neighbor classifie,'' *IEEE Trans. Cybern.*, vol. 46, no. 6, pp. 1263–1275, Jun. 2016.
[31] C. L. P. Chen and Z. Liu, ''Broad learning system: An effective and efficient incremental learning system without the need for deep architecture,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 10–24, Jan. 2018.
[32] C. L. P. Chen and Z. Liu, ''Broad learning system: A new learning paradigm and system without going deep,'' in *Proc. 32nd Youth Academic Annu. Conf. Chin. Assoc. Autom. (YAC)*, May 2017, pp. 1271–1276.
[33] H. Zhao, J. Zheng, J. Xu, and W. Deng, ''Fault diagnosis method based on principal component analysis and broad learning system,'' *IEEE Access*, vol. 7, pp. 99263–99272, 2019, doi: 10.1109/ACCESS.2019.2929094.
[34] M. Wang, Q. Ge, H. Jiang, and G. Yao, ''Wear fault diagnosis of aero-engines based on broad learning system and ensemble Learning,'' *Energies*, vol. 12, no. 24, p. 4750, 2019.
[35] J. Zheng, Y. Yuan, L. Zou, W. Deng, C. Guo, and H. Zhao, ''Study on a novel fault diagnosis method based on VMD and BLM,'' *Symmetry*, vol. 11, no. 6, p. 747, 2019.
[36] S. B. Jiang, P. K. Wong, and Y. C. Liang, ''A fault diagnostic method for induction motors based on feature incremental broad learning and singular value decomposition,'' *IEEE Access*, vol. 7, pp. 157796–157806, 2019, doi: 10.1109/ACCESS.2019.2950240.
[37] A. E. Hoerl and R. W. Kennard, ''Ridge regression: Biased estimation for nonorthogonal problems,'' *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.
[38] H. He and E. A. Garcia, ''Learning from imbalanced data,'' *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.
[39] J. Liu, ''Fuzzy support vector machine for imbalanced data with borderline noise,'' *Fuzzy Sets Syst.*, vol. 413, pp. 64–73, Jun. 2021.
[40] Y. Geng and X. Luo, ''Cost-sensitive convolutional neural networks for imbalanced time series classification,'' *Intell. Data Anal.*, vol. 23, no. 2, pp. 357–370, Apr. 2019.
[41] L. Jiang, C. Li, and S. Wang, ''Cost-sensitive Bayesian network classifiers,'' *Pattern Recognit. Lett.*, vol. 45, pp. 211–216, Aug. 2014.
[42] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 2003.
[43] M. Nguyen *et al.*, ''m-TSNE: A framework for visualizing high-dimensional multivariate time series,'' in *Proc. Workshop Vis. Anal. Healthcare (VAHC)*, 2016.
[44] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, ''SMOTE: Synthetic minority over-sampling technique,'' *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.

CHONG WANG received the B.Sc. degree in applied mathematics and the M.Sc. degree in statistics from Beihang University, Beijing, China, in 2014 and 2017, respectively, where she is currently pursuing the Ph.D. degree with the School of Reliability and Systems Engineering. Her research interests include fault diagnosis and imbalance data.



JIE LIU received the B.Sc. degree in mechanical engineering and the M.Sc. degree in physics from Beihang University, Beijing, China, in 2009 and 2012, respectively, and the Ph.D. degree from CentraleSupélec, France. He is currently an Associate Professor with the School of Reliability and Systems Engineering, Beihang University. His research interests include fault detection, diagnostics, and prognostics.