



Article Research on a Fault Diagnosis Method for the Braking Control System of an Electric Multiple Unit Based on Deep Learning Integration

Yueheng Wang ^{1,2}, Haixiang Lin ^{1,2,*}, Dong Li ², Jijin Bao ² and Nana Hu ²

- Key Laboratory of Railway Industry of BIM Engineering and Intelligent for Electric Power, Traction Power Supply, Communication and Signaling, Lanzhou Jiaotong University, Lanzhou 730070, China; wangyh699@mail.lzjtu.cn
- ² School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China; 12231702@stu.lzjtu.edu.cn (D.L.); 12231688@stu.lzjtu.edu.cn (J.B.); 12221602@stu.lzjtu.edu.cn (N.H.)
- * Correspondence: linhaixiang@mail.lzjtu.cn

Abstract: A fault diagnosis method based on deep learning integration is proposed focusing on fault text data to effectively improve the efficiency of fault repair and the accuracy of fault localization in the braking control system of an electric multiple unit (EMU). First, the Borderline-SMOTE algorithm is employed to synthesize minority class samples at the boundary, addressing the data imbalance and optimizing the distribution of data within the fault text. Then, a multi-dimensional word representation is generated using the multi-layer bidirectional transformer architecture from the pre-training model, BERT. Next, BiLSTM captures bidirectional context semantics and, in combination with the attention mechanism, highlights key fault information. Finally, the LightGBM classifier is employed to reduce model complexity, enhance analysis efficiency, and increase the practicality of the method in engineering applications. An experimental analysis of fault data from the braking control system of the EMU indicates that the deep learning integration method can further improve diagnostic performance.

Keywords: the braking control system of EMU; fault diagnosis; deep learning integration; LightGBM; Borderline-SMOTE

1. Introduction

With the rapid development of high-speed railways, the reliability and safety of electric multiple unit (EMU) operations have garnered increasingly widespread attention [1]. Among the nine key technologies of EMUs, the stability and reliability of the braking system are directly related to their safe and stable operation [2]. The braking control system, serving as the brain and control core of an EMU braking system, is responsible for the operation and specific execution of the braking control system is a complex system with multiple potential points of failure. Therefore, during the operation of EMUs, if small or potential faults are not timely diagnosed and effectively addressed during brake control calculations, they may trigger a chain reaction resulting in accidents and even catastrophic consequences. The accurate and reliable early-stage diagnosis and detection of faults can facilitate timely repairs and, to some extent, serve as accident prevention measures.

The diagnosis of faults is crucial for ensuring the reliability and safety of system or equipment operation because it involves making precise judgments and analyzing the operational status and abnormal conditions. This process forms the basis for recovering the faults in the system [5]. Rapid advancements in sensing technology and artificial intelligence have facilitated the emergence of knowledge and data-driven intelligent fault diagnosis techniques, attracting considerable attention [6,7]. An area of research focus and difficulty lies in effectively utilizing historical monitoring data to enhance the accuracy



Citation: Wang, Y.; Lin, H.; Li, D.; Bao, J.; Hu, N. Research on a Fault Diagnosis Method for the Braking Control System of an Electric Multiple Unit Based on Deep Learning Integration. *Machines* **2024**, *12*, 70. https://doi.org/10.3390/ machines12010070

Academic Editor: Xiang Li

Received: 20 November 2023 Revised: 7 January 2024 Accepted: 8 January 2024 Published: 17 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). and stability of fault diagnosis outcomes for the braking control system of an EMU. This endeavor carries significant academic value and addresses practical demands.

Control system faults can be quickly and directly identified using knowledge-based fault diagnosis methods. Zhang Y [8] designed an expert system for diagnosing train brake faults, utilizing the acquisition technology for information on the fault status of brake systems of trains. Zhang Y C [9] and Atamuradov V [10] developed an expert system for diagnosing faults in the brake system, which is based on fault trees. Li W X [11] conducted a study on fault diagnosis techniques for the braking system of an EMU, incorporating fuzzy theory and expert systems. This study considered the characteristics of multiple cross faults, complex fault hierarchies, and the challenges in accurately matching types of fault knowledge. This approach helps us to mitigate the challenge of knowledge-based fault diagnosis methods, they encounter challenges in acquiring knowledge. Moreover, the accuracy of diagnosis relies on experiential knowledge, and incomplete knowledge databases can lead to misjudgments.

Data-driven techniques, noted for their intelligent and automated fault diagnosis capabilities, have found extensive use in detecting faults in railway braking systems. These techniques encompass statistical analysis, signal processing, and artificial intelligence. Statistical analysis and signal processing entail scrutinizing historical texts, images, waveforms, and other data to ascertain the operational status of samples. Zhang T [12] combined fuzzy theory with the analytic hierarchy process to diagnose faults and perform a comprehensive evaluation of the CCBII braking system. Soares N [13] introduced a feature extraction method that employs scalable hypothesis testing to select, extract, and cluster features using unsupervised models with principal component analysis. Zhou D H [14] proposed a fault detection and isolation method for brake cylinder systems that utilizes inter-variable variance and reconstruction contribution plots. Seo B [15] suggested a fault diagnosis method for the electromagnetic valves in subway braking systems, which relies on sensor signals and physical behavior models. While these techniques can swiftly identify fault types, their diagnostic accuracy diminishes when dealing with diagnostic objects possessing intricate data structures. Moreover, acquiring fault data for high-speed train braking systems and other rail vehicles poses a significant challenge. Considering the scarcity of fault data in practical scenarios relative to regular data, Liu J [16,17] introduced an enhanced method that employs support vector machines (SVMs) to tackle classification challenges arising from data imbalance. Conversely, Zuo J [18] diagnosed pneumatic unit leakage faults in the braking systems of high-speed trains by extracting distinctive features from input and output signals and employing a support vector classification algorithm.

These aforementioned methods employ traditional fault diagnosis approaches to achieve fault diagnosis in brake systems. Nonetheless, the diagnostic results obtained from these traditional methods display notable discrepancies, which undermine the robustness of the diagnostic models. As deep learning has matured, it has lessened the biases in traditional fault diagnosis methods, consequently enhancing the accuracy and stability of fault diagnosis. Liu H [19] employed a bidirectional long short-term memory network to diagnose faults in railway signal equipment, effectively tackling the high dimensionality and sparsity of fault texts. Lu R J [20] used convolutional neural networks for diagnosing faults in the onboard equipment of railway train control systems. Moreover, they substituted the fully connected classification component with PSO-SVM, leading to enhanced accuracy in fault classification. The integration of the word vector generation tool Word2vec with neural networks has led to substantial advancements in fault diagnosis [21,22]. Despite the comprehensive extraction of text feature information, the softmax classifier employed in the model exhibits sensitivity toward imbalanced samples and is inadequate for capturing intricate feature relationships, resulting in diminished diagnostic accuracy. Various methods, including the SMOTE algorithm [23], GC-SMOTE [24], and data augmentation [25], have been employed in fault diagnosis research to address data imbalance issues. This has led to the mitigation of impacts caused by the imbalanced distribution of text features on fault diagnosis outcomes.

However, these methods encounter challenges, such as the imprecise identification of fault types in the presence of complex structures as well as imbalanced distributions of fault data, incomplete feature extraction, gradual feature assignment degradation during model training, and slow training speed rates, thereby leading to suboptimal fault diagnosis results. Constructing an optimal ensemble learning model and classifier based on the data structure is crucial for enhancing the accuracy of fault diagnosis, thereby effectively improving the classification accuracy and training efficiency of the model. Building upon this, we present a fault diagnosis methodology for the train brake control system of a multiple unit train using deep learning ensembles, offering the following distinct contributions:

- (1) This study proposes an imbalance optimization method for fault data in the braking control system of EMUs, utilizing the B-SMOTE algorithm. The B-SMOTE algorithm is employed to generate minority class samples at the boundaries. Subsequently, the data distribution is optimized, effectively reducing the probability of misdiagnosis in the minority class samples.
- (2) This study presents a deep learning-based ensemble model for fault diagnosis in the braking control system of EMUs. The model incorporates deep-level feature extraction of the fault data, enabling precise fault classification even with imbalanced samples. Consequently, the fault diagnosis of EMU braking control systems can be effectively achieved. In comparison with conventional fault diagnosis models, the proposed approach significantly enhances the accuracy and robustness of fault diagnosis for the braking control system of EMUs.
- (3) This study proposes a fault classifier for the braking control system of EMUs based on LightGBM. LightGBM significantly reduces the time and computational complexity of fault classification through the utilization of the histogram algorithm and the gradientbased one-side sampling algorithm. Moreover, LightGBM accelerates model training by employing optimized feature and data parallelism methods, thereby effectively improving the fault diagnosis accuracy of the braking control system.

2. Data Source and Data Characteristics

A total of 10,170 incidents of train braking control system failures that occurred within a specific vehicle section in the past two years were utilized as the sample data and classified based on the fault location, as illustrated in Table 1.

Primary Fault Secondary Fault Railway carriage (A1) Wall panels (A2) Floor (A3) Vehicle body and chassis Beam (A4) Reservoir (A5) Release valve (A6) Chassis structure (A7) Brake regulator (B1) Tube system (B2) Braking Brake cylinder (B3) Drawbar (B4) Bolster (C1) The fault classification of the braking Side frame (C2) control system of EMU Bogie Crossbar (C3) Spring-loaded pallet (C4) Vehicle hook (D1) Hooked tongue (D2) Coupler yoke (D3) Coupler buffer device Uncoupling lever bracket (D4) Coupling lever (D5) Buffers (D6) Rim-spoke plate (E1) Wheel axle Âxles (E2)

Table 1. Model volume comparison.

By classifying and statistically analyzing the text data of train braking control system failures, as depicted in Figure 1, an issue is evident through the imbalanced distribution among fault categories. This implies that a significant proportion of faults in certain categories can overshadow a smaller proportion of faults, thereby affecting the accuracy of a fault diagnosis and leading to significant deviations between diagnosis results and the actual issues. Specifically, the imbalance coefficient between the vehicle hook and uncoupling lever bracket faults reaches as high as 61. Such imbalanced data distribution can result in a tendency to diagnose minority class samples as majority class samples, which is a critical issue in fault diagnosis research.



Figure 1. Distribution of original text data.

3. Research Framework

The integrated fault diagnosis model for the braking control system of EMUs, based on deep learning, is depicted in Figure 2. Initially, the B-SMOTE algorithm is applied to rectify the issue of imbalanced textual data and optimize the distribution of fault-related textual data. Subsequently, the BERT pre-trained model's multi-layered bidirectional transformer architecture is employed to generate multidimensional word representations. BiLSTM is then utilized to effectively capture contextual information and fused bidirectional semantics while integrating an attention mechanism to emphasize critical fault-related details. Finally, the LightGBM model is employed to establish the association between features and fault location labels.



Figure 2. Framework for fault diagnosis of EMU braking control system.

3.1. Optimization of Text Data Distribution

The imbalanced distribution of data in the fault texts of the EMU braking control system is evident from Figure 1, which poses challenges for fault diagnosis. In this study, we introduce the B-SMOTE algorithm to enhance the SMOTE data generation mechanism. Instead of oversampling all minority class samples, B-SMOTE selectively oversamples

samples at the boundaries, thereby mitigating the risk of overfitting. Additionally, the B-SMOTE algorithm incorporates density distribution and distance between samples during synthetic sample generation, preserving the dataset's distribution characteristics and ensuring the authenticity of the fault text dataset in the EMU braking control system. The algorithm categorizes minority class samples into safe class (a), dangerous class (b), and noise class (c), with oversampling specifically applied to the dangerous class comprising minority samples with more than half of their neighbors being majority class samples. Figure 3 illustrates the principle of synthetic sample generation.





The B-SMOTE algorithm employs various parameters. *S* represents the sample set and distinguishes between the majority class (S_{maj}) and minority class (S_{min}) sample sets. x_n represents the neighboring samples, with the quantity determined by the value of *K*, while x_i and x_{ij} represent the attributes of the sample itself and the neighboring sample, respectively. The term R_{ij} is typically set to 0.5 or 1. The specific steps of the B-SMOTE algorithm are outlined below:

Step 1: determine the nearest sample set of $S' \in S$ for each minority class sample of x_i . Step 2: determine the count of nearest neighbors from the majority sample set for each x_i in the minority class and assign it as $|S' \cap S_{mai}|$.

Step 3: choose the sample of x_i that meets the criteria of the dangerous class, specifically the condition $\frac{K}{2} < |S' \cap S_{maj}| < K$.

Step 4: calculate the distance between samples $d_{ij} = x_i - x_{ij}$ between attributes of *j* that correspond to x_i and x_n using the synthetic minority class samples from the dangerous class. Obtain the newly synthesized minority class sample of $h_{ij} = x_i + d_{ij} \times rand(0, R_{ij})$.

3.2. BERT

BERT is a language representation model in the field of deep learning that is built upon a multi-layer bidirectional transformer architecture. Unlike traditional unidirectional language models, such as Word2Vec, BERT leverages a multi-layer bidirectional transformer architecture and integrates new multi-task training objectives. This enables BERT to generate context-based multidimensional word representations, effectively increasing the expressiveness of word vectors. In order to address the challenges posed by the limited size and complexity of texts related to brake control system failures, the BERT model is used to obtain multidimensional word vector representations. The training process of the BERT model involves using annotated data from control system failure texts. The obtained training results and deep multi-level text features are then utilized in downstream feature extraction tasks.

BERT considers the relationship features of token embedding, segment embedding, and position embedding during the process of obtaining word representations for control

system failure texts. This is performed to achieve more comprehensive semantic expression vectors. In the pretraining process of BERT, which is based on control system failure texts and displayed in Figure 4, an input sequence of length *n* undergoes multiple layers of feature embedding and transformer layers to generate output vectors. The labels "[CLS]" and "[SEP]" are utilized to identify the starting and ending positions of sentences in the text, respectively.



Figure 4. BERT pretraining model.

3.3. BiLSTM

The word representations of failure texts acquired from the BERT layer are fed into the BiLSTM layer, which comprises both forward and backward LSTM units capable of capturing long-range information and contextual semantic features. This aids in a comprehensive understanding and analysis of the causes of failures. Figure 5 illustrates the structure of the LSTM network unit.



Figure 5. LSTM network unit structure diagram.

The LSTM gating unit is composed of an input gate i_t , a forget gate f_t , an output gate o_t , and a memory cell c_t . These components are expressed as follows:

$$f_t = \sigma \Big(W_f \cdot [h_{t-1}, x_t] \Big) + b_f \tag{1}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t]) + b_i \tag{2}$$

$$c_t = \tanh(W_c \cdot [h_{t-1}, x_t]) + b_c \tag{3}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t]) + b_o \tag{4}$$

$$h_t = o_t \cdot \tanh(c_t) \tag{5}$$

In the equation, *W* represents the weight coefficients and *b* represents the bias terms. The variables of h_{t-1} and x_t represent the hidden state value at time t - 1 and the input

value at time is represented by *t*, respectively. The variables of σ and tanh represent the sigmoid and hyperbolic tangent activation functions, respectively. The output sequence of *Y* is subsequently passed into the BiLSTM by passing the input sequence through both the forward and backward LSTM layers, whereby the BiLSTM network generates the output, which is denoted as $H = \begin{bmatrix} \overrightarrow{h}, \overrightarrow{h} \end{bmatrix}$.

3.4. Attention Mechanism

While BiLSTM effectively integrates the contextual information of the fault text in the train braking control system, it does not effectively highlight the key fault information within the text. The attention mechanism utilizes attention aggregation to assign distinct weights to characters within the text, thus facilitating the continuous learning and updating of these weights. This mechanism empowers the model to effectively filter essential information from intricate fault features, eliminating or diminishing redundant details, thereby enhancing the efficiency of extracting critical features and improving the accuracy of fault diagnosis. As a result, this allows the model to selectively focus on specific fault information, thereby enhancing the efficiency and accuracy of the model. In this study, we adopt a static attention mechanism to compute the importance of words by utilizing the

BiLSTM output vector of $H = \left[\overrightarrow{h}, \overleftarrow{h}\right]$, which is expressed as:

$$e_i = \tanh\left(W^T H_i + b\right) \tag{6}$$

$$\alpha_i = \frac{\exp(e_i)}{\sum\limits_i \exp(e_i)} \tag{7}$$

$$H_i' = H_i \cdot \alpha_i \tag{8}$$

In the equation, W^T represents the weight vector and *b* represents the bias component. The variable of e_i quantifies the significance of the *i*th character within the sentence. α_t represents the weight assigned to the input and H'_i represents the weight of the sentence.

3.5. LightGBM

GBDT, a model obtained through the iterative training of weak classifiers, represents an optimal solution. LightGBM incorporates additional techniques on top of GBDT, including one-sided gradient-based sampling, depth-limited histograms, and leaf growth strategies. The combination of one-sided sampling based on gradients and depth-limited histograms effectively determines optimal split points, reduces information loss, saves memory, and improves processing speed, effectively addressing the time consumption issue for large sample datasets. Additionally, LightGBM adopts a leaf-wise growth strategy whereby it identifies the leaf among all current leaves with the highest split gain and performs another split, thus creating a cyclic process, as demonstrated in Figure 6. Nevertheless, leaf-wise growth can result in overfitting due to the development of excessively deep decision trees. Consequently, we implement a leaf-wise growth strategy with depth limitations to prevent overfitting low split gain, unnecessary searching, and splitting.



Figure 6. The learning process of leaf-wise growth strategy.

To locate an appropriate tree, f_K , that minimizes the function, we employ the objective function specified in Equation (9). Here, y_i represents the fault label value, \hat{y}^{K-1} signifies the outcome of the (K - 1)th learning, and c^{K-1} denotes the regularization term from the preceding K - 1 trees.

$$O^{K} = \sum_{i} L(y_{i}, \hat{y}_{i}^{K}) + \Omega(f_{K}) + c^{K-1}$$

=
$$\sum_{i} L(y_{i}, \hat{y}_{i}^{K-1} + f_{K}(x_{i})) + \Omega(f_{K}) + c^{K-1}$$
(9)

We assume that the prediction, denoted as \hat{y}_i^K , for the *i*th sample at the *K*th iteration and the objective function O^K for the *K*th iteration can be approximated through the use of a second-order Taylor expansion, as demonstrated below:

$$\sum_{i} L(y_{i}, \hat{y}_{i}^{K-1} + f_{K}(x_{i})) = \sum_{i} \left[L(y_{i}, \hat{y}_{i}^{K-1}) + L'(y_{i}, \hat{y}_{i}^{K-1}) f_{K}(x_{i}) + \frac{1}{2} L''(y_{i}, \hat{y}_{i}^{K-1}) f_{K}^{2}(x_{i}) \right]$$
(10)

To minimize O^{K} in each training iteration, it is necessary to obtain M non-overlapping regions $\{Rmk\}_{m=1}^{M}$ for each tree and the optimal leaf node scores $\{\beta mk\}_{m=1}^{M}$. As a result, with α and λ serving as regularization factors for their respective parameters, the objective function can be represented as follows:

$$O^{K} = \sum_{i=1}^{M} \left(\sum_{xi \in RmK} L'(y_{i}, \hat{y}_{i}^{K-1}) \beta_{mk} + \frac{1}{2} \left(\sum_{xi \in RmK} L''(y_{i}, \hat{y}_{i}^{K-1} + \lambda) \beta_{mk}^{2} \right) + \alpha L$$
(11)

LightGBM selects the leaf with the highest split gain for splitting. However, calculating the information gain for split points requires traversing all data points for each feature, resulting in substantial memory consumption. To address this issue, a leaf growth strategy with depth limitation is implemented, whereby the maximum depth is specified during the training process. Splitting halts at a node when it either reaches the depth limit or when no further splitting is possible. This strategy ensures efficiency, optimizes memory usage, and prevents overfitting.

4. Experimental Analysis

4.1. Experimental Dataset and Evaluation Index

To assess the effectiveness of the model, the experimental data was divided into a 70% training set and a 30% test set. The five-fold cross-validation method [26] was applied to randomly partition the training set into smaller subsets for validation. Evaluation metrics in terms of accuracy, precision, recall, and F1 score [27] were employed.

4.2. Setup of the Experimental Environment

The experimental setup involved a CPU (Intel i9 12900K) and GPU (NVIDIA GeForce RTX3080Ti) running on the PyCharm 2021.2 integrated development environment (IDE), with the code implemented in Python 3.8. Before executing the fault diagnosis model for the train unit's braking control system, it was necessary to configure the model parameters. Among these parameters, the number of iterations played a critical role as a key hyperparameter. Setting it too low might result in underfitting, while setting it too high would increase training time and hinder generalization. Figure 7 illustrates the performance metric variations during epochs 0 to 30. The graph demonstrates that the model's accuracy and loss function stabilize, achieving the desired effect around epoch 25. Consequently, for this study, 25 epochs were chosen as the final iteration count.



Figure 7. The impact of epochs on model performance.

In the B-SMOTE algorithm, the selection of boundary samples in the minority class is determined through the value of R_{ij} . Only those minority class samples identified as boundary samples are chosen for synthesis in order to mitigate the potential adverse impact on the overall distribution of the minority class. Figure 8 illustrates the variations in evaluation metrics as the value of R_{ij} changes. Notably, the model performs optimally when R_{ij} is set to 0.5.



Figure 8. Change in evaluation index with *R*_{*ij*} value.

Furthermore, Figure 9 illustrates the length statistics of fault texts in the EMU braking control system. The average sentence length is 60, with the sentence padding size set to 72.



Figure 9. Fault text length statistics histogram.

Based on the findings from multiple experiments, the optimal combination of key parameters for diagnosing faults in the train unit's braking control system is determined and presented in Table 2.

Table 2. Model volume comparison.

Hyperparameter Name	Hyperparameter Value
Epoch	25
Learning rate	0.001
BERT embedding dimension	300
Maximum sequence length	128
Number of nodes in the hidden layer of LSTM	256
Number of LSTM layers	2
Optimization function	Adam
Batch size	32
Loss	0.5
Depth	14
$\hat{R_{ii}}$	0.5
Padding size	72

4.3. B-SMOTE-Generated Minority Class Samples for Experimentation

Figure 10 displays the results obtained from utilizing B-SMOTE for the automatic generation of minority class samples. In comparison to Figure 1, there is a substantial decrease in the imbalance of the different data classes. The imbalance coefficients for the vehicle hook and uncoupling lever bracket decrease from 61 to 2.8, effectively optimizing the imbalanced distribution of fault text data for the train brake control system.



Figure 10. Sample distribution after B-SMOTE is utilized.

This experimental comparison compares the model proposed in this article with the model that does not utilize the B-SMOTE algorithm. The confusion matrix of the model without the B-SMOTE algorithm is illustrated in Figure 11. The imbalanced distribution of fault texts makes minority class fault samples in the diagnostic model of the train brake control system more susceptible to being misclassified as majority class fault samples. For example, faults in the minority class belonging to the chassis structure (A7) are erroneously identified as buffer faults (D6), resulting in decreased accuracy in fault diagnosis. A comparison between Figures 11 and 12 reveals that the proposed model demonstrates enhanced diagnostic accuracy for minority class samples following the application of the B-SMOTE algorithm while preserving the diagnostic accuracy for majority class samples. This underscores the superior performance of the model in dealing with imbalanced datasets.



Figure 11. Confusion matrix without B-SMOTE algorithm.



Figure 12. Confusion matrix of the proposed model.

4.4. Experimental Analysis

To further substantiate the excellence of the fault diagnosis model developed for the braking control system of the EMU in this study, the control group encompassed the BiLSTM [28], BiLSTM-Attention [29], Word2vec-CNN [30], Word2vec-BiLSTM-Attention [31], BERT [32], and BERT-BiLSTM-Attention models [33]. Figure 13 displays the experimental results.



Figure 13. Fault diagnosis results of different models.

The experimental results demonstrate that the model proposed in this study achieves an accuracy of 94.34% and an F1 score of 92.72%, outperforming the other six models. The performance of the entity extraction model and the control models will be analyzed from the perspectives of feature extraction and embedding methods as well as classifiers.

- (1) The BiLSTM-Attention model, which is based on the BiLSTM model, utilizes the attention mechanism to enable the model to focus on identifying input information that is highly relevant to the current classification. Consequently, when compared with the BiLSTM model, the BiLSTM-Attention model demonstrates improvements in both accuracy and F1 score of 1.87% and 4.19%, respectively. These results indicate that the attention mechanism effectively enhances the efficiency of feature extraction for key vectors.
- (2) Then, the BiLSTM-Attention and CNN models are separately combined with Word2vec, and the word embedding BERT is introduced for comparison. The results demonstrate that both word embedding methods improve the model's performance. However, as a dynamic word embedding method, BERT specifically enhances the representation of text features, enabling it to achieve optimal results when dealing with complex texts in the railway field. Consequently, compared to Word2vec, the BERT model exhibits a classification effect that is 1.73% and 0.74% higher than the Word2vec-CNN and Word2vec-BiLSTM-Attention models, respectively. Furthermore, the classification accuracy of the BERT-BiLSTM-Attention model surpasses that of the BERT model by 1.14%, with a corresponding 0.69% increase in the F1 value. This indicates that the combined model of BERT and BiLSTM-Attention overcomes the limitations of a single BERT model in text feature extraction and weight distribution.
- (3) In order to improve the practicality of the method for engineering applications, we incorporate LightGBM as the classifier. The experimental results indicate that under the same experimental conditions, the proposed model outperforms the BERT-BiLSTM-Attention model, achieving a 4.73% improvement in comprehensive evaluation metrics. Furthermore, to examine the influence of LightGBM on model performance, a comparison is conducted among different control models based on their training and testing times, as depicted in Table 3. In contrast, the BERT-BiLSTM-Attention model, configured using the parameters specified in Table 2, requires 24.8 min to train the dataset, whereas our model requires only 16.4 min. These results suggest that the LightGBM model enables parallel analysis, effectively reducing model complexity and thus enhancing efficiency of analysis. Therefore, the proposed model enhances the efficiency of fault diagnosis in the braking control system of the EMU to a certain degree, thereby increasing the practicality of the method in engineering applications.

Model	Training Time/min	Testing Time/s
BiLSTM	10.7	1.41
BiLSTM-Attention	12.6	1.53
Word2vec-CNN	10.8	1.43
Word2vec-BiLSTM-Attention	14.5	1.78
BERT	15.6	1.95
BERT-BiLSTM-Attention	24.8	2.12
Proposed method	16.4	2.01

 Table 3. Comparison of fault diagnosis efficiencies among different models.

5. Conclusions

In this paper, we presented a fault diagnosis method for an EMU's braking control system using deep learning integration. The method utilized fault data from the EMU braking control system to conduct experiments and analysis, which yielded the following conclusions:

(1) To mitigate the problem of inadequate model generalization and diagnostic accuracy, which are caused by imbalanced sample distribution, in diagnosing EMU braking control system faults, we employed the B-SMOTE algorithm to generate minority class samples and optimize the distribution of fault textual data. The experimental results demonstrate that the application of the B-SMOTE algorithm improved the diagnostic accuracy of minority class samples in our model while maintaining the diagnostic accuracy of majority class samples. This model effectively handles imbalanced datasets. However, its performance may be affected by hyperparameter

settings, and it may not be suitable for all types of datasets. In future research, we will investigate adaptive parameter adjustment methods based on dataset characteristics to enhance the flexibility and adaptability of B-SMOTE, thus reducing reliance on hyperparameter settings.

- (2) In order to provide additional evidence regarding the efficacy of the deep learning integration model in diagnosing faults within the EMU braking control system, we conducted comparative experiments utilizing authentic fault data and alternative models. The outcomes clearly demonstrate that the proposed model surpasses other models in terms of accuracy, recall rate, and F1 score, achieving values of 94.34%, 92.32%, and 92.72%, respectively.
- (3) In order to enhance the practicality of the method in engineering applications, we employed the LightGBM classifier to classify the extracted semantic features. A comparative analysis between the proposed model and the original BERT-BiLSTM-Attention model illustrates a 4.73% improvement in overall evaluation metrics, accompanied by a reduction in training time of 8.4 min. These findings indicate that LightGBM has the ability to decrease model complexity and expedite runtime, effectively addressing the time-consuming challenges encountered by traditional algorithms when dealing with large-scale sample data. Ultimately, this approach significantly enhances the accuracy and robustness of the model.

In conclusion, the proposed deep learning-based integrated fault diagnosis method enables intelligent fault diagnosis for EMU braking control systems. It demonstrates the potential to assist maintenance personnel in identifying malfunctioning components, thereby facilitating faster and more precise repair decision making.

Author Contributions: Conceptualization, Y.W. and H.L.; data curation, Y.W. and N.H.; formal analysis, H.L. and D.L.; methodology, D.L. and J.B.; writing—original draft, J.B.; writing—review and editing, Y.W. and N.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Young Scholars Science Foundation of Lanzhou Jiaotong University (2021016).

Data Availability Statement: The data used are confidential.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Awad, F.A.; Graham, D.J.; Singh, R.; AitBihiOuali, L. Predicting urban rail transit safety via artificial neural networks. *Saf. Sci.* 2023, 167, 106282. [CrossRef]
- Shrestha, S.; Spiryagin, M.; Wu, Q. Friction condition characterization for rail vehicle advanced braking system. *Mech. Syst. Signal Process.* 2019, 134, 106324. [CrossRef]
- 3. Wang, Q.; Wang, Z.; Mo, J.; Zhou, Z. Modelling and stability analysis of a high-speed train braking system. *Int. J. Mech. Sci.* 2023, 250, 108315. [CrossRef]
- 4. Gültekin, Ö.; Cinar, E.; Özkan, K.; Yazici, A. Multisensory data fusion-based deep learning approach for fault diagnosis of an industrial autonomous transfer vehicle. *Expert Syst. Appl.* **2022**, 200, 117055. [CrossRef]
- 5. Velasco-Gallego, C.; De Maya, B.N.; Molina, C.M.; Lazakis, I.; Mateo, N.C. Recent advancements in data-driven methodologies for the fault diagnosis and prognosis of marine systems: A systematic review. *Ocean. Eng.* 2023, 284, 115277. [CrossRef]
- 6. Zuo, B.; Zhang, Z.; Cheng, J.; Huo, W.; Zhong, Z.; Wang, M. Data-driven flooding fault diagnosis method for proton-exchange membrane fuel cells using deep learning technologies. *Energy Convers. Manag.* **2022**, 251, 115004. [CrossRef]
- Yang, D.; Karimi, H.R.; Gelman, L. An explainable intelligence fault diagnosis framework for rotating machinery. *Neurocomputing* 2023, 541, 126257. [CrossRef]
- 8. Zhang, Y.; Pan, H.; Zhou, Y.; Li, M.; Sun, G. Efficient visual fault detection for freight train braking system via heterogeneous self distillation in the wild. *Adv. Eng. Inform.* 2023, *57*, 102091. [CrossRef]
- 9. Zhang, Y.C. Real-time monitoring and fault diagnosis expert system for locomotive braking system. *Comput. Meas. Control* 2013, 21, 2615–2617, 2620.
- Atamuradov, V.; Camci, F.; Baskan, S.; Sevkli, M. Failure diagnostics for railway point machines using expert systems. In Proceedings of the 2009 IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, Cargèse, France, 31 August–3 September 2009; IEEE: New York, NY, USA, 2009; pp. 1–5.

- 11. Li, W.X.; Zhang, Y.; Lin, R.W.; Wang, L. Fault Diagnosis and Safety Measures of EMU Braking System. *Railw. Locomot. Car* 2011, 31, 39–42.
- 12. Zhang, T. CCBII brake based on multi-hierarchy fuzzy evaluation. *Electr. Eng.* 2009, 2009, 61–65.
- 13. Soares, N.; de Aguiar, E.P.; Souza, A.C.; Goliatt, L. Unsupervised machine learning techniques to prevent faults in railroad switch machines. *Int. J. Crit. Infrastruct. Prot.* **2021**, *33*, 100423. [CrossRef]
- Zhou, D.H.; Ji, H.Q.; He, X.; Shang, J. Fault detection and isolation of the brake cylinder system for electric multiple units. *IEEE Trans.* Control. Syst. Technol. 2018, 26, 1744–1757. [CrossRef]
- 15. Seo, B.; Jo, S.H.; Oh, H.; Youn, B.D. Solenoid valve diagnosis for railway braking systems with embedded sensor signals and physical interpretation. In Proceedings of the Annual Conference of the PHM Society, Denver, CO, USA, 2–8 October 2016; p. 8.
- Liu, J.; Li, Y.F.; Zio, E. A SVM framework for fault detection of the braking system in a high-speed train. *Mech. Syst. Signal Process.* 2017, 87, 401–409. [CrossRef]
- 17. Liu, J.; Zio, E. A scalable fuzzy support vector machine for fault detection in transportation systems. *Expert Syst. Appl.* **2018**, *102*, 36–43. [CrossRef]
- Zuo, J.; Ding, J.; Feng, F. Latent leakage fault identification and diagnosis based on multi-source information fusion method for key pneumatic units in Chinese standard electric multiple units (EMU) braking system. *Appl. Sci.* 2019, 9, 300. [CrossRef]
- 19. Liu, H.; Wang, Y.; Zhou, X.; Ye, Y. Question answering system for deterministic fault diagnosis of intelligent railway signal equipment. *Smart Resilient Transp.* **2021**, *3*, 202–214. [CrossRef]
- Lu, R.J.; Lin, H.X.; Xu, L.; Lu, R.; Zhao, Z.X.; Bai, W.S. Fault diagnosis for on-board equipment of train control system based on CNN and PSO-SVM hybrid model. J. Meas. Sci. Instrum. 2022, 13, 430–438.
- 21. Shangguan, W.; Meng, Y.Y.; Yang, J.M.; Cai, B.G. LSTM-BP neural network based fault diagnosis for on-board equipment of Chinese train control. *J. Beijing Jiaotong Univ.* **2019**, *43*, 54–62.
- Chen, S.; Ding, Y.; Xie, Z.; Liu, S.; Ding, H. Chinese Weibo sentiment analysis based on character embedding with dual-channel convolutional neural network. In Proceedings of the 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), Chengdu, China, 20–22 April 2018; IEEE: New York, NY, USA, 2009; pp. 107–111.
- 23. Yang, L.B.; Li, P.; Xue, R.; Ma, X.; Wu, Y.H.; Zou, D.L. Intelligent classification of faults of railway signal equipment based on imbalanced text data mining. *J. China Railw. Soc.* 2018, 40, 59–66.
- 24. Georgios, D.; Fernando, B. Geometric SMOTE a geometrically enhanced drop-in replacement for SMOTE. Inf. Sci. 2019, 501, 118135.
- 25. Li, X.Y.; Yang, M.; Quan, R. Unbalanced text classification method based on deep learning. J. Jilin Univ. 2022, 52, 1889–1895.
- 26. Deng, J.F.; Cheng, L.L.; Wang, Z.W. Attention-based BiLSTM fused CNN with gating mechanism model for Chinese long text classification. *Comput. Speech Lang.* 2021, *68*, 101182. [CrossRef]
- Gao, F.; Li, F.; Wang, Z.F. Research on multilevel classification of high-speed railway signal equipment fault based on text mining. J. Electr. Comput. Eng. 2021, 2021, 7146435. [CrossRef]
- 28. Li, W.J.; Qi, F.; Yu, Z.T. Sentiment classification method based on multi-channel features and self-attention. J. Softw. 2021, 32, 2783–2800.
- Liriam, E.; Santos, A.R.; Ricardo, M.; Li, W.G.; Geraldo, P. Multi-label legal text classification with BiLSTM and attention. *Int. J. Comput. Appl. Technol.* 2022, 68, 369–378.
- 30. Lin, H.X.; Lu, R.J.; Lu, R. Automatic classification method of railway signal fault based on text mining. J. Yunnan Univ. 2022, 44, 281–289.
- 31. Han, G.; Bu, T.; Wang, M.M. Text classification of railway traffic accidents based on dual-channel bidirectional long short term memory network. *J. China Railw. Soc.* **2021**, *43*, 71–79.
- 32. Alammary, A.S. BERT Models for Arabic Text Classification: A Systematic Review. Appl. Sci. 2022, 12, 5720. [CrossRef]
- Shobana, J.; Murali, M. An Improved Self Attention Mechanism Based on Optimized BERT-BiLSTM Model for Accurate Polarity Prediction. Comput. J. 2023, 66, 1279–1294. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.