machines

MDPI

Article

Fault Diagnosis of a Switch Machine to Prevent High-Speed Railway Accidents Combining Bi-Directional Long Short-Term Memory with the Multiple Learning Classification Based on Associations Model

Haixiang Lin [1,2], Nana Hu [1], Ran Lu [3], Tengfei Yuan [4,*], Zhengxiang Zhao [1], Wansheng Bai [1] and Qi Lin [5]

[1] School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China; linhaixiang@mail.lzjtu.cn (H.L.); 17339820212@163.com (N.H.); 19193156561@163.com (Z.Z.); bws199904@163.com (W.B.)
[2] Key Laboratory of Four Power BIM Engineering and Intelligent Application Railway Industry, Lanzhou 730070, China
[3] CCCC Railway Design and Research Institute Co., Ltd., Beijing 101304, China; luran191724705@sina.com
[4] SHU-UTS SILC Business School, Shanghai University, Shanghai 201800, China
[5] School of Materials Science and Engineering, Beihang University, Beijing 100191, China; valpomair07@gmail.com
* Correspondence: yuantengfei@shu.edu.cn

Abstract: The fault diagnosis of a switch machine is vital for high-speed railway operations because switch machines play an important role in the safe operation of high-speed railways, which often have faults because of their complicated working conditions. To improve the accuracy of turnout fault diagnosis for high-speed railways and prevent accidents from occurring, a combination of bi-directional long short-term memory (BiLSTM) with the multiple learning classification based on associations (MLCBA) model using the operation and maintenance text data of switch machines is proposed in this research. Due to the small probability of faults for a switch machine, it is difficult to form a diagnosis with the small amount of sample data, and more fault text features can be extracted with feedforward in a BiLSTM model. Then, the high-quality rules of the text data can be acquired by replacing the SoftMax classification with MLCBA in the output of the BiLSTM model. In this way, the identification of switch machine faults in a high-speed railway can be realized, and the experimental results show that the *Accuracy* and *Recall* of the fault diagnosis can reach 95.66% and 96.29%, respectively, as shown in the analysis of the ZYJ7 turnout fault text data of a Chinese railway bureau from five recent years. Therefore, the combined BiLSTM and MLCBA model can not only realize the accurate diagnosis of small-probability turnout faults but can also prevent high-speed railway accidents from occurring and ensure the safe operation of high-speed railways.

Keywords: high-speed railway; switch machine; fault diagnosis; text data; BiLSTM and MLCBA

Citation: Lin, H.; Hu, N.; Lu, R.; Yuan, T.; Zhao, Z.; Bai, W.; Lin, Q. Fault Diagnosis of a Switch Machine to Prevent High-Speed Railway Accidents Combining Bi-Directional Long Short-Term Memory with the Multiple Learning Classification Based on Associations Model. *Machines* **2023**, *11*, 1027. https://doi.org/10.3390/machines11111027

Academic Editor: Ahmed Abu-Siada

Received: 14 October 2023
Revised: 9 November 2023
Accepted: 14 November 2023
Published: 17 November 2023

1. Introduction

Switch machines play a crucial role in the automatic control of high-speed railways, which often have various faults due to their complicated working conditions. According to the China high-speed railway signal system fault statistics data from 2019 to 2022, the proportion of turnout faults is the highest, accounting for 40.5% of the total [1–5]. China's high-speed railway signal system fault statistics data are presented in Figure 1.

As complex mechanical equipment, switch machines consist of different types of components, so the faults are caused by many different factors. Therefore, it is difficult to diagnose the fault of a switch machine efficiently [6]. Additionally, the probability of some faults of a switch machine is too small, so the recorded text data are also sparse. Due to the phenomenon of uneven fault distribution, it is difficult to diagnose the small-probability faults of a switch machine accurately [7].

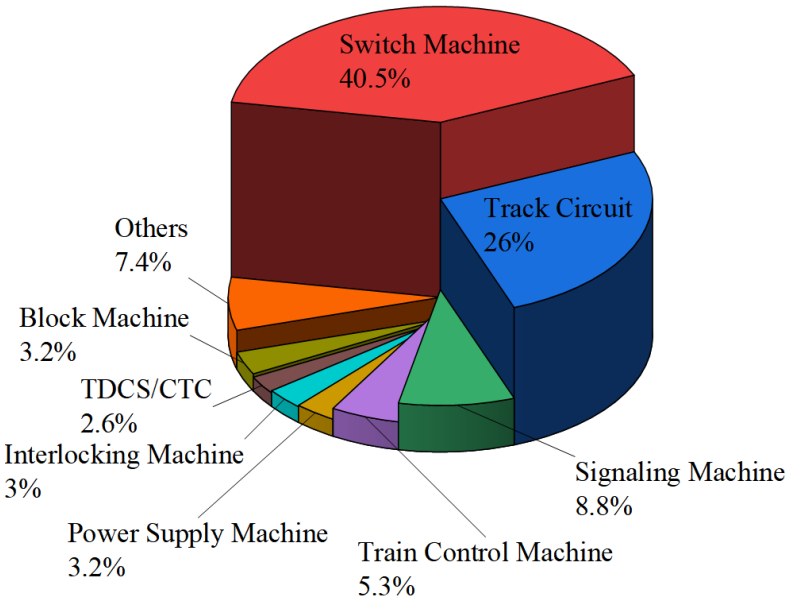


Figure 1. China's high-speed railway signal system fault statistics data from 2019 to 2022.

In view of the difficulty of fault diagnosis and the uneven fault distribution for China's high-speed railway switch machines, a novel combined BiLSTM-MLCBA model is introduced to realize the various fault diagnoses of a switch machine with text data efficiently and accurately. This model can not only improve the fault diagnosis efficiency of a switch machine but can also guarantee the safe and efficient operation of a high-speed railway.

2. Literature Review

2.1. Fault Diagnosis of High-Speed Railway Switch Machines

Railway device fault diagnosis has developed into a systematic subject that is mainly dependent on the support of practical demands and multi-disciplines [8]. Specifically, switch machines have gradually shifted from relying on expert experience to text data. Therefore, the fault diagnosis models of a high-speed rail switch machine can be classified into three types, namely mathematics models, knowledge models, and data-driven models.

The main purpose of a mathematical model for fault diagnosis is to achieve quantitative and qualitative analysis, as well as determine the location of faults. For example, in ref [9], Dai established a hidden semi-Markov model with a dynamic particle swarm algorithm that could predict the fault of a switch machine effectively. In ref [10], Eker proposed a railway switch machine fault prediction method that could predict railway turnout electromechanical system faults. In fact, a mathematical model cannot meet the demands of high-speed rail switch machine fault diagnosis because it has randomness and low probability. Additionally, a high-speed rail switch machine has a complicated relationship with other devices in the working state, and expert experience and knowledge cannot be neglected. In this context, an expert system of railway signal equipment fault pattern recognition was established by Zhang [11]. Furthermore, an expert system for rail turnout fault diagnosis was built by Bian [12], and the system applied a time dynamic programming model to achieve precise recognition of rail turnout for three running states. Although the rail switch machine fault diagnosis expert system based on knowledge is efficient in the process of fault detection, it still has some subjectivity. Given the weakness of the knowledge model, the data-driven model became popular for the promotion of the automation level of rail transit. Currently, many scholars are applying working text data to explore railway signal fault diagnosis for the popularity of various sensors. For instance, Zhang [13] established the BP neural network model to recognize a switch machine's fault pattern with the electrical current data of an S700K switch machine. To make up for the

defects of complicated fault diagnosis for the improved S700K-C switch machine, Lei [14] combined WPT (Wavelet Packet Technology) and EMD (Empirical Mode Analysis) models to study fault diagnosis. Based on the particle swarm optimization support vector machine model, Liu [15] predicted the faults for an S700K switch machine. Moreover, Wu [16] established a ZYJ7 switch machine fault diagnosis model based on a neural network with recording data. Considering the change law of the switch electric current curve, Zhang [17] put forward an intelligent detection algorithm. In summary, the data-driven fault diagnosis method can overcome the limitations of mathematics and knowledge models, and it needs a certain amount of fault data for a rail switch machine. High-speed railway switch machine equipment is composed of multiple modules, each with distinct failure mechanisms, and the probabilities of failures vary across different modules, resulting in certain components exhibiting a relative sparsity in their failure data. Therefore, the main goal of this study is to achieve the small-probability fault diagnosis of a high-speed rail switch machine.

2.2. Fault Diagnosis of Rail Transit with Text Data

As mentioned above, analyzing a significant quantity of railway switch machine fault data collected in textual format can provide a greater possibility of fault diagnosis with improvement in the automation degree of rail transit. The main steps of fault diagnosis with text data include feature extraction and rule classification, and each step can affect the accuracy of fault diagnosis.

Term frequency–inverse document frequency (TF-IDF) is a representative method of traditional text feature extraction that is generally used to extract the features of onboard equipment faults [18,19]. Additionally, Li [20] and Zhou [21] used a vector space model (VSM) to transform a safety log into vectors for subway safety risk system fault diagnosis. Then, Zhao [22] and Zhong [23] applied a probabilistic subject model to extract the features of fault text data for on-board equipment and turnout fault diagnosis. In the context of widespread artificial intelligence (AI), bi-directional long short-term memory (BiLSTM) networks are used for feature extraction [24]. The research results show that this can overcome the disadvantages of TF-IDF, VSM, and subject models. The main reason for this is that BiLSTM can automatically extract semantic feature vectors in lower dimensions, as well as avoid the correlation between words not being considered and a large amount of training causing overfitting. After extracting text features, Zhu [25] used BiLSTM with Support Vector Machine (SVM) to achieve a good classification result. In addition, Naive Bayesian (NB) algorithms [26] combined with association rules [27] have been widely used in the rail fault diagnosis domain. For example, Xie [28] used an NB classifier to achieve urban rail ground device fault classification with a fault log. However, the integrated deep learning methods proposed in the aforementioned studies fail to meet the accuracy requirements for diagnosing rare failures. To mitigate the issue of misdiagnosis caused by the limited number of rare samples, experts from various domains have employed diverse approaches. In the agricultural domain, Ding [29] disregarded minor crops or uncommon diseases and exclusively concentrated on prevalent crop diseases, aiming for precise diagnosis. In the medical domain, Xie [30] incorporated samples with a substantial number of missing values and a limited number of records into the category of abnormal data, employing the approach of removing such samples directly to predict antibiotic resistance. These studies effectively alleviate the issue of misdiagnosis stemming from limited samples by disregarding and eliminating rare samples. However, railway equipment and train safety are closely interconnected, and neglecting the infrequent failures of high-speed railway switch and turnout machines could result in survivorship bias [31], posing a threat to train safety and potentially causing severe accidents [4]. Consequently, Yang [5] utilized the SMOTE sample synthesis method to create synthetic samples of rare failures. Nonetheless, employing traditional SoftMax as a classifier renders it sensitive to imbalanced samples and unsuited for capturing intricate feature relationships. Accordingly, meticulous consideration should be given when selecting the fault text feature extraction model and rule classifier.

In view of the small-probability faults of a high-speed rail switch machine, this research combines BiLSTM with MLCBA to realize the feature extraction and rule classification of unstructured short Chinese text data, which can not only complete the fault diagnosis task better but can also prevent accident occurrence for a high-speed railway.

3. Data Description

The ZYJ7 switch machine is widely used in the Chinese high-speed railway network, so its fault text data are recorded in unstructured short Chinese text records, which are mainly composed of numbers, letters, special symbols, and professional vocabulary words. Examples of the fault text data for the high-speed railway ZYJ7 switch machine are shown in Table 1. The main topics include fault descriptions and fault types.

Table 1. Examples of fault text data for high-speed railway ZYJ7 switch machine.

No.	Description of ZYJ7 Switch Machine Fault	Title 3
1	On 24 August 2017, from xx to xx, the poor sealing of the purple copper gasket in the main host 121177# air cylinder resulted in oil leakage.	Fault of "air cylinder"
2	On 6 July 2017, from xx to xx, there was oil seepage at the pressure sensor of the startup oil cylinder in "11992#", making it impossible to disassemble.	Fault of "hydraulic cylinder assembly"
3	On 18 December 2018, from xx to xx, the roller inside the switch machine of "173046#" failed to unlock, causing the turnout to be unable to move.	Fault of "contact assembly"
4	On 17 June 2019, from xx to xx, it was reported on-site that the unlocking pressure of the operating lever in "174602#" was excessive, resulting in a high unlocking curve.	Fault of "operating lever"

Because the proportion of ZYJ7 switch machines is the highest in China's high-speed railway network, the ZYJ7 switch machine is taken as the research object in this paper. According to the recording data of a railway bureau from 2017 to 2019, there are 12 main fault types for a ZYJ7 switch machine. As described in Table 2, the distribution of high-speed railway ZYJ7 switch machine fault types is uneven. Among the high-speed railway ZYJ7 switch machine fault types, "the assembly of the motor oil pump", "the assembly of the contacts", and "the assembly of the hydraulic cylinder" are the most common fault types, while "bottom case", "the rod indicating the locking status", and "circuit breaker" are the less common fault types. Additionally, the ratio of maximum and minimum fault types is as high as 28:1, so the unevenly distributed fault data make diagnosis difficult, especially of the small-probability faults.

Table 2. Statistical analysis of faults observed in the ZYJ7 switch machine used in high-speed railways.

Title	Fault Type	Number of Fault Cases	Fault Occurrence Rate (%)
C1	The assembly of the motor oil pump	1366	40.37
C2	The assembly of the hydraulic cylinder	253	8.12
C3	The assembly of the contacts	545	24.09
C4	The joint of the oil pipe	136	3.67
C5	Bottom case	48	0.98
C6	Air cylinder	89	2.23
C7	Operating lever	122	2.98
C8	The rod indicating the locking status	56	1.02
C9	Defects in the railway track	173	3.81
C10	Relay	136	3.51
C11	Circuit breaker	58	1.14
C12	Cable circuit	247	8.08

4. Method Description

With the goal of improving the diagnosis accuracy of ZYJ7 switch machine faults, a deep learning integration method combining BiLSTM with MLCBA is introduced in this research. Therefore, data processing, feature extraction, and rule classification are the three essential steps discussed in this section.

As a whole, the specific process of the fault diagnosis of a ZYJ7 switch machine based on the BiLSTM-MLCBA is designed as shown in Figure 2. The process mainly includes three parts: fault text data processing for the ZYJ7 switch machine, text feature extraction based on the BiLSTM networks, and fault diagnosis with the MLCBA classifier.

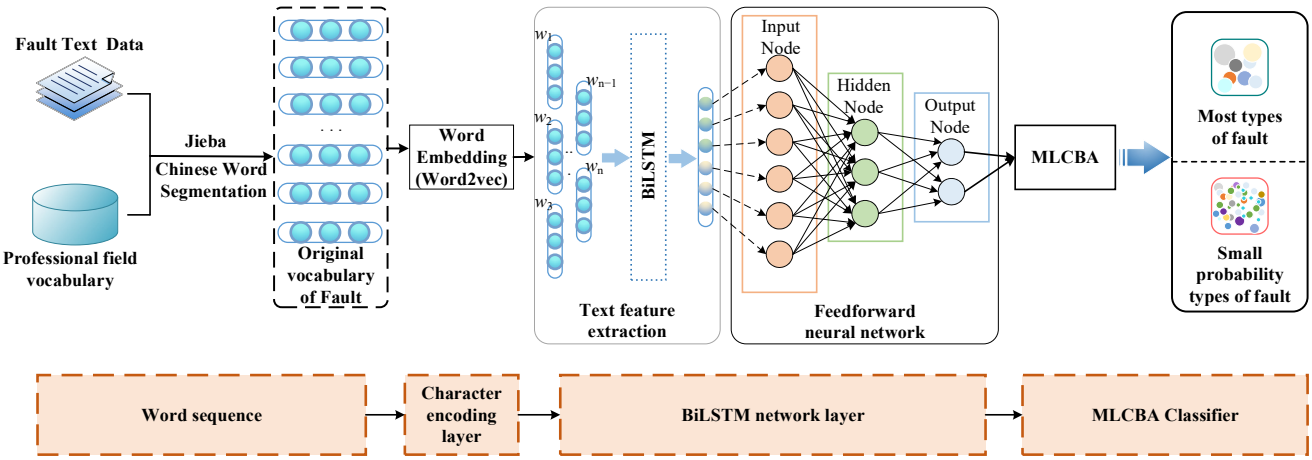


Figure 2. Fault diagnosis model for a high-speed railway with a ZYJ7 switch machine combining BiLSTM with MLCBA.

4.1. Data Processing

Because the ZYJ7 switch machine fault text data are recorded in unstructured short Chinese vocabulary, the Jieba word segmentation tool is selected to establish a professional field thesaurus to store the words, as shown in Table 3. Then, the redundant words are eliminated, such as time and place of faults, and the original high-speed railway switch machine fault vocabulary is eventually obtained.

Table 3. Professional field vocabulary of switch machine.

Professional Field Vocabulary of Switch Machine	Number of Term
Normal position of switch	234
Reverse position of switch	128
Switch blocked	341
Electro-hydraulic switch machine	230
Electro-pneumatic switch machine	412
Switch restored	145
Switch connecting rod	189
Switch point closure	367
Switch closure adjustment	120
Loss of indication of a switch	156
Switch locking	78
.....	.....

After the original high-speed railway switch machine fault vocabulary is completed, it is necessary to transform the words of the fault text data into distributed word vectors, also referred to as Word2vec. Then, the Continuous Bag of Words (CBOW) model with the hierarchical SoftMax classifier is applied to construct the language model and retain the semantic features of fault text data as much as possible. The output layer of the CBOW model is a Huffman tree, the words in the corpus are taken as leaf nodes, and the number of each word is taken as the weight. The optimization function of the CBOW model can be expressed as follows:

$$L = \sum_{w \in C} \log p(w|\text{Context}(w)) \quad (1)$$

where C is the corpus, p is the probability function, w is the current word, and $Context(w)$ is the context of the current word w. With a series of Word2vec processing with the CBOW model, the fixed-length vectors containing the contextual semantic information of words can be obtained, and the semantic features of the fault text data can be preserved as much as possible.

4.2. Feature Extraction

The reason for the choice of BiLSTM for feature extraction is that it consists of two LSTM layers. In this double-layer structure, the fault text data can be transmitted from the input layer to the output layer and then to feedback. Therefore, the fault text data can be trained repeatedly between these two LSTM models, which can obviously make up for the disadvantage of the insufficient data of a small-probability fault for a high-speed railway switch. Figure 3 shows the BiLSTM network structure diagram. Each layer of the BiLSTM is simultaneously an input layer and an output layer.

In the BiLSTM model, the fault text data for a ZYJ7 switch machine can be used fully, and the features of the fault text data can be extracted. The steps are as follows.

(1) The word vectors obtained with fault data processing are input to the embedding layer. The length of a statement in the input fault text data is supposed to be m. x_i represents the word vector of the i^{th} word, $x_i \in R_n$, where n is the word vector dimension and R is the set of word vectors. Hence, all the statements can be expressed as follows:

$$X_{1:m} = x_1 \oplus x_2 \oplus \cdots \oplus x_m \text{ (where " } \oplus \text{ " is connection operation)} \quad (2)$$

(2) BiLSTM is composed of two LSTM neural networks. Figure 4 shows the basic structure of the LSTM neural network model. The forget gate, input gate, output gate, and cell state are the components of the LSTM, and they are given by Equations (3)–(7).

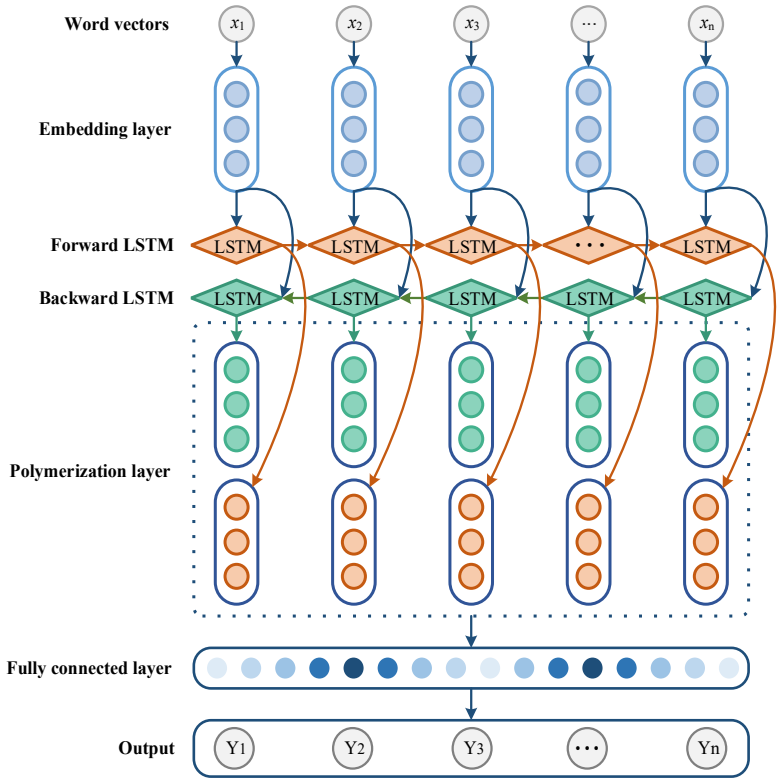


Figure 3. Network structure of BiLSTM model.

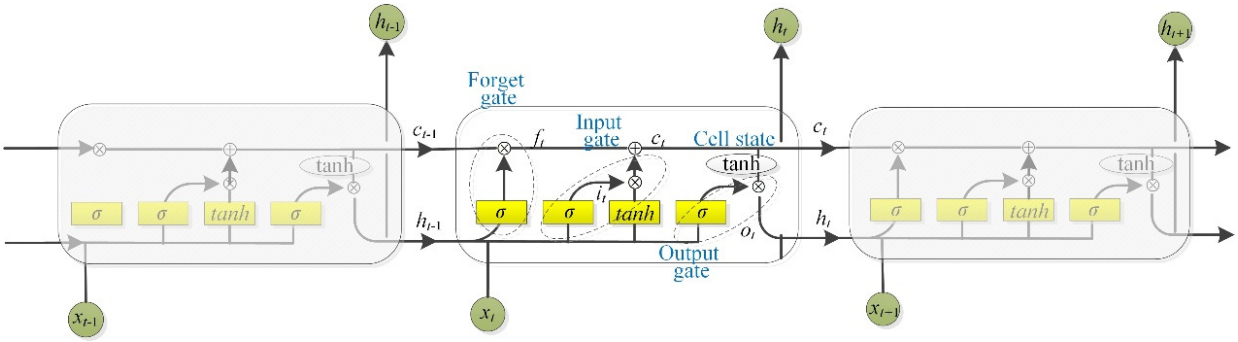


Figure 4. Structure of LSTM neural network model.

Forget gate:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{3}$$

Input gate:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{4}$$

Cell state:

$$\widetilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b) C_t \tag{5}$$

Output gate:

$$o_t = \sigma(W_O \cdot [h_{t-1}, x_t] + b) \tag{6}$$

$$h_t = o_t \cdot \tanh(C_t) \tag{7}$$

where x_t and h_t are the input vector and the value of the hidden layer state at time t, respectively. W and b represent the weight coefficient matrix and offset term, respectively. σ is the activation function of the sigmoid, and $\tanh(C_t)$ is the activation function of the hyperbolic tangent. $\widetilde{C}_t$ represents the state of the LSTM cell.

In LSTM, these three gates can not only control the proportion of forgotten information but can also determine the step size of the transmission. In addition, they can also be helpful for learning the long-distance dependence of sentences and solving the relationship of word orders so that the deep semantic expression of text data can be captured.

(3) The overall feature vectors of the text data can be obtained with the forward and backward bi-directional processing between the two-way LSTM layers. Additionally, the fault text features can be generated through aggregation and used as input for the classifier.

(4) The classification result can be converted into a probability value between 0 and 1 by the SoftMax classifier, and the fault type values can be finally output as the classification results.

4.3. Rule Classification

The critical step of rule classification after feature extraction is to use multiple learning classification based on associations (MLCBA) to replace the SoftMax classifier. MLCBA is integrated to extract more high-quality rules, and the concept of the association degree [32] is introduced to improve the classification accuracy of unbalanced fault data, which can achieve full coverage of all fault types. Initially, Ji, H.P. [33] proposed the association rule classification algorithm CBA and applied association rules to achieve data classification in 2016. However, the CBA algorithm never considered the imbalance of all types of data samples, which led to some rules being ignored and all the training examples not being able to be covered. Given the characteristics of ZYJ7 switch machine fault data, the MLCBA algorithm was further designed to achieve more high-quality rules. The basic definition and the specific classification process of the MLCBA algorithm are introduced in the following sections.

4.3.1. Basic Definition of Correlation Rule Classification

According to the literature, correlation rule classification originates from class association rules (CARs), and all CARs are generally mined with the Apriori algorithm [34,35]. In CARs, B is restricted to a type attribute in a classification task, and A is a set of several feature attributes. For transaction set I, where the CAR can be expressed as $A \Rightarrow B(A \in I, B \in I, A \cap B =)$, A and B are, respectively, the premise and consequence. In detail, several important indicators of correlation rule classification are as follows:

(1) The support expression is as follows:

$$\text{Support}(A) = \frac{\text{Count}(A)}{|I|} \tag{8}$$

(2) The confidence expression is as follows:

$$\text{Confident}(A \Rightarrow B) = \frac{\text{Support}(A \cup B)}{\text{Support}(A)} \tag{9}$$

where $Confident(A \Rightarrow B)$ indicates the frequency of category B occurring in the number of transactions that contain item set A. In addition to the above, it is necessary to consider frequent item sets. In common usage, the support threshold is set by the users. If the support degree of item set A is not less than the support threshold, item set A can be called a frequent item set.

(3) Furthermore, the degree of lift is expressed as follows:

$$\text{Lift}(A \Rightarrow B) = \frac{\text{Support}(A \Rightarrow B)}{\text{Support}(A) \cdot \text{Support}(B)} \tag{10}$$

In Equation (10), if the lift is bigger than 1, A is positively associated with type B. Otherwise, A is negatively associated with type B, and the negative association rules should be removed. The promotion degree is bigger, and the influence of A on category B is greater.

(4) $CL\ \sup(A \Rightarrow B)$ is the class support between sets A and B, which is expressed as shown in Equation (11). Additionally, the class support of the association rule $A \Rightarrow B$ may be higher when there is a small amount of data in a certain type B.

$$\mathrm{CLsup}(A \Rightarrow B) = \frac{\mathrm{Support}(A \cup B)}{\mathrm{Support}(B)} \tag{11}$$

For class frequent item sets, it is defined that if the class support degree of item set A in type B is not less than the class support threshold of type B, then item set A is called the class frequent item set of type B.

(5) The complement class support is expressed as follows:

$$\mathrm{CCS}(A \Rightarrow B) = \frac{\mathrm{Support}(A \cup \overline{B})}{\mathrm{Support}(\overline{B})} \tag{12}$$

where B is the current category and $\overline{B}$ is a complement to B. *CCS* represents the strength of the rules in the complementary classes. If the *CCS* value of the rule is smaller, the quality of the rule is better.

(6) The expression of the Laplace rule strength is as follows:

$$\mathrm{Laplace}(A \Rightarrow B) = \frac{\mathrm{Count}(A \cup B) + 1}{\mathrm{Count}(A) + C} \tag{13}$$

where B is the current data category and C is the number of all types in the transaction set. If the intensity of the Laplace rules is higher, then the quality of the rules is higher.

(7) In correlation rule classification, the corresponding concept of the correlation degree is put forward for the unbalanced data, and it is expressed as follows:

$$\mathrm{CD}(A \Rightarrow B) = \mathrm{Confident}(A \Rightarrow B) - \mathrm{CCS}(A \Rightarrow B) \tag{14}$$

where confidence and complementary support need to be considered comprehensively. Additionally, if the confidence of the rules is higher, the support of complementary classes is lower, while the quality of the rules is higher. Because the correlation degree considers various evaluation indexes simultaneously, it can better generate high-quality rules and make small-class rules have higher quality. Based on this, the correlation degree can be more suitable for the fault classification of small-probability data for the ZYJ7 switch machine of a high-speed railway.

(8) The rule strength is expressed as follows:

$$RS(A \Rightarrow B) = Laplace(A \Rightarrow B)\frac{Lift(A \Rightarrow B)(CCS(A \Rightarrow B) + 1)}{MAX(CCS(A \Rightarrow B), t)} \tag{15}$$

where t is set to a small value to prevent the complementary support degree from being zero and the denominator from making no sense. Because the strength of the RS rules roundly considers the Laplace strength, complement class support, and lift degree, the strength of the RS rules can consider the correlation between item sets and categories of rules. Additionally, it can also make high-quality subclass rules process a higher rule strength compared with the single confidence and promotion degree.

4.3.2. Fault Classification Process of High-Speed Railway Switch Machines

This subsection describes how more high-quality rules and correlation degrees can be generated through multiple learning with a training set with the proposed MLCBA

algorithm. Hence, the feature extraction of small-probability faults can be carried out several times to obtain more complete rules, and the specific steps are as follows.

(1) First, the number of multiple learning times and the extraction ratio of each learning instance are set, and the thresholds of the support and correlation degrees are set simultaneously.
(2) Second, the frequent item sets of the new training sets are mined by the support threshold, and the new training sets are randomly selected from the original training sets.
(3) Next, the correlation degree of the frequent item sets mined from each new training set to each type is calculated, and the appropriate rules with the threshold of correlation degree are explored.
(4) Then, all the rules learned each time are merged, the repeated rules are eliminated, and the rules are pruned at the same time.
(5) Finally, the training examples that cannot be judged in the training set are learned again, and the new rules are extracted and added to the rule set. Therefore, the full rule of the training example can be covered completely, and all rules can be sorted by the strength of the RS rule.

As mentioned above, in the process of fault diagnosis of a high-speed railway ZYJ7 switch machine, the text data for the ZYJ7 switch machine fault are first segmented. Then, the original fault thesaurus is generated, and the vectorization of the original fault thesaurus is achieved using Word2vec. Afterward, the bi-directional semantic fault feature extraction is achieved with BiLSTM, which can obtain the feature matrix of the fault. Finally, the fault diagnosis layer sends the fault feature matrix to MLCBA instead of the SoftMax classifier. Fault classification and identification for the high-speed railway ZYJ7 switch machine can be implemented effectively by following the above steps, especially for the diagnosis of small-probability high-speed railway switch machine faults.

5. Empirical Results and Discussion

5.1. Empirical Results

5.1.1. Environment and Configuration Parameters for Simulation

In this research, 3229 high-speed railways with ZYJ7 switch machine fault text data in China were chosen to verify the reliability and validity of the proposed model. The detailed parameters of the environment and configuration for the simulation are presented in Table 4. The parameters include the configuration parameters of the hardware and software, the programming language, and the fault text data processing tools.

Table 4. Experimental environment and configuration.

Experimental Environment	Environment Configuration
Operating system	Linux (manufacturer: IBM, Armonk, NY, USA)
CPU	Intel (R) Core (TM) (manufacturer: Intel, Santa Clara, CA, USA)
GPU	NVIDIA GeForce RTX3090Ti (manufacturer: NVIDIA Corporation, Santa Clara, CA, USA)
CUDA	Version No.: 11.2.162
Memory	64 GB
Programming language	Python 3.7
Word segmentation tool	Jieba
Word Vector Training Toolkit	Gensim (Version No. 4.1.0)
Deep learning framework	TensorFlow-GPU (Version No. 1.14.0)
C10	136
C11	58
C12	247

5.1.2. Evaluation Indicators of the BiLSTM-MLCBA Fault Diagnosis Model

Once the specific steps of the BiLSTM-MLCBA fault diagnosis model have been decided, *Precision*, *Recall*, and *F*1 are selected as evaluation indicators. Then, the 3229 high-speed railways with the ZYJ7 switch machine fault text data are divided into training and testing samples in the ratio of 7:3. Moreover, the model training adopts the five-fold cross-validation method [36]. The concrete parameter matrix of the evaluation indicators for the BiLSTM-MLCBA fault diagnosis model is shown in Table 5.

Table 5. Parameters of evaluation indicators of BiLSTM-MLCBA fault diagnosis model.

Data Category	Positive Example of Projection	Negative Example of Projection
Positive example of reality	TP	FN
Negative example of reality	FP	TN

For the three evaluation indicators, *Precision* is usually used to test the accuracy of a model and *Recall* is commonly used to check the integrity of a model. These can be calculated as follows:

$$\begin{aligned} Precision &= TP/(TP+FP) \\ Recall &= TP/(TP+FN) \end{aligned} \tag{16}$$

In addition, *F*1 is the total mean of *Precision* and *Recall*, which can be calculated as follows:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{17}$$

5.1.3. Fault Text Data Preprocessing for a High-Speed Railway with a ZYJ7 Switch Machine

As described in Section 4.1, fault text data preprocessing for a high-speed railway with a ZYJ7 switch machine can also be executed as shown in Figure 5.

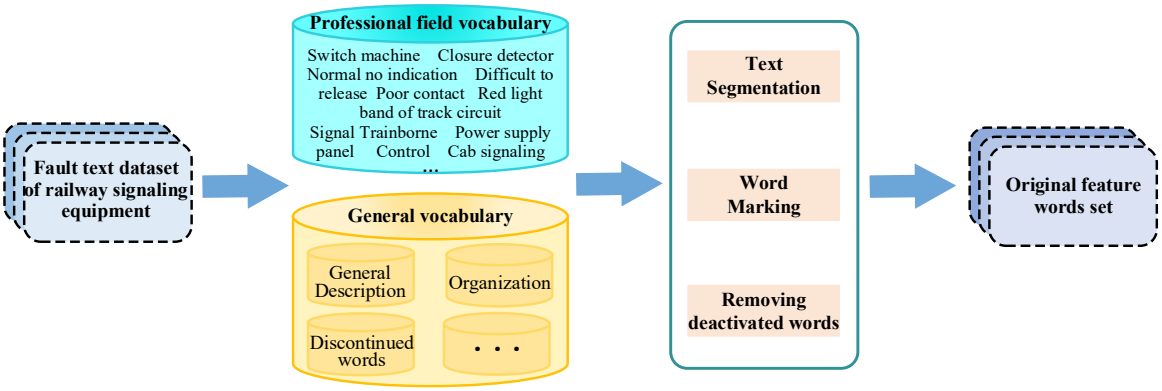


Figure 5. Fault text data preprocessing for a high-speed railway with a ZYJ7 switch machine.

First, the feature word set of the fault text data information can be obtained by constructing the professional field vocabulary based on general vocabulary. Because the professional field vocabulary can accurately segment fault feature words so that they contain the critical fault type information, the word item should be created in the word segmentation processing. Second, the precise pattern of the Jieba library is applied to accomplish the automatic word segmentation in this experiment, which can not only segment the words in sentences more accurately but can also preserve the most features of the text data.

5.1.4. Analysis of Different Hyper-Parameters in the BiLSTM Model

For testing the performance of the BiLSTM of the combination model, it is necessary to determine the related optimal hyper-parameters, especially those of the BiLSTM model. The different hyper-parameters have different impacts on the model's performance, and the dropout value is a critical means to enhance the generalization ability of the model as well as improve the over-fitting problem and efficiency of the training mode. Since the length of the fault text of the ZYJ7 switch machine is short, the window length of the CBOW model is set to five, and the word vector dimension is set to 150. Figure 6 demonstrates the fluctuation of evaluation metrics with varying dropout values. Specifically, the *Precision*, *Recall*, and *F1* score exhibit changes as the dropout values vary. Selecting an appropriate dropout value not only mitigates the issue of overfitting but also improves the training efficiency of the model. Figure 6 clearly shows that the model achieves its best performance when the dropout value is set to 0.5.

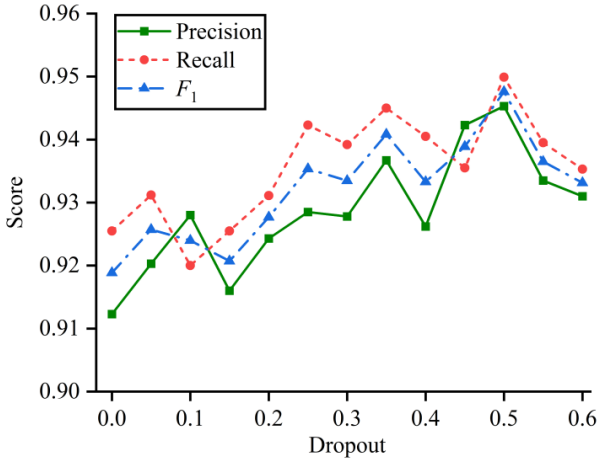
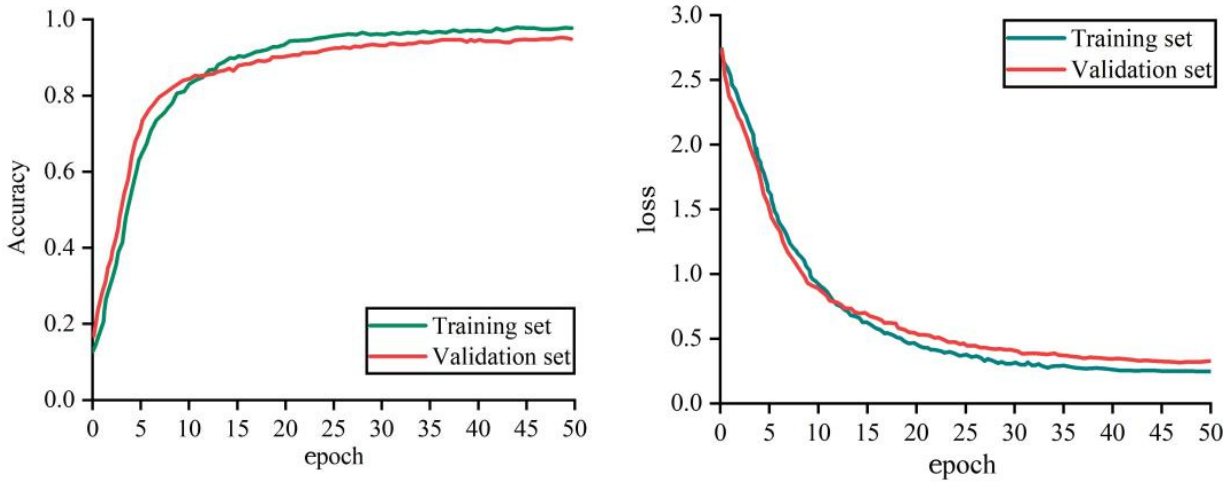


Figure 6. Scores of the three evaluation indicators with different dropout values.

As shown in Figure 7, 50 is selected as the final epoch in this research because the values of the accuracy and the loss function tend to be stable when the epoch reaches 40–50, which indicates that the model has been fully trained and that it has achieved good results.



(**a**) Accuracy of model training (**b**) Loss of model training

Figure 7. Model training process.

In summary, the optimal hyper-parameters of the BiLSTM model can be configured so that the epoch is 50, the number of LSTM hidden layer nodes is 256, the LSTM layer is 1, the batch size is 20, the dropout value is 0.5, and the learning rate is 0.001.

5.1.5. Study of the Threshold of Correlation Degree for the MLCBA Algorithm

After the BiLSTM model is configured, the threshold of the correlation degree for the MLCBA algorithm should be studied to improve the classification accuracy of the fault text data. Because the number of optimal learning times is determined to be three, the extraction ratios of the training data are set to 75%, 85%, and 100%. Since the correlation threshold ranges from 0.2 to 0.8, the optimal threshold can be obtained through comparison, as shown in Figure 8. The results show that the accuracy of the MLCBA algorithm can reach approximately 94% when the threshold ranges from 0.45 to 0.6. Therefore, the threshold of the correlation degree for the MLCBA algorithm is set to 0.5 in this research.

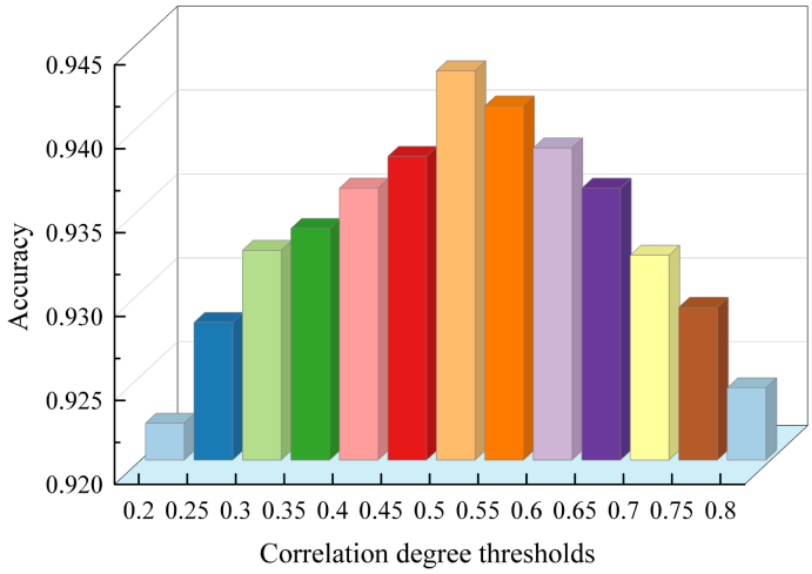


Figure 8. Different correlation thresholds with different classification accuracies.

5.2. Results and Discussions

5.2.1. Study of the Threshold of Correlation Degree for the MLCBA Algorithm

Figure 9a,b show the various fault diagnosis result confusion matrixes for BiLSTM-CBA and BiLSTM-MLCBA. In the figures, the x and y coordinates represent C1–C12 fault types for a ZYJ7 switch machine of a high-speed railway. Additionally, the optimal support threshold and the confidence threshold of the BiLSTM-CBA model are set as 0.1 and 0.5, respectively, according to the requirements of the experiment.

From the diagnosis results, the small-probability fault "The joint of the oil pipe" (C4) is more easily diagnosed than the large-probability fault "The assembly of the motor oil pump" (C1). The main reason for this is that some fault text data features of "The joint of the oil pipe" are mixed with some relevant fault text data features of "The assembly of the motor oil pump", and the fault "The joint of the oil pipe" is described as "Oil leakage at external tubing joint". The obvious issue is that BiLSTM-MLCBA has a higher diagnosis rate for the switch machine of a high-speed railway than BiLSTM-CBA, especially in the small-probability fault data sample. The discriminant analysis results are as follows.

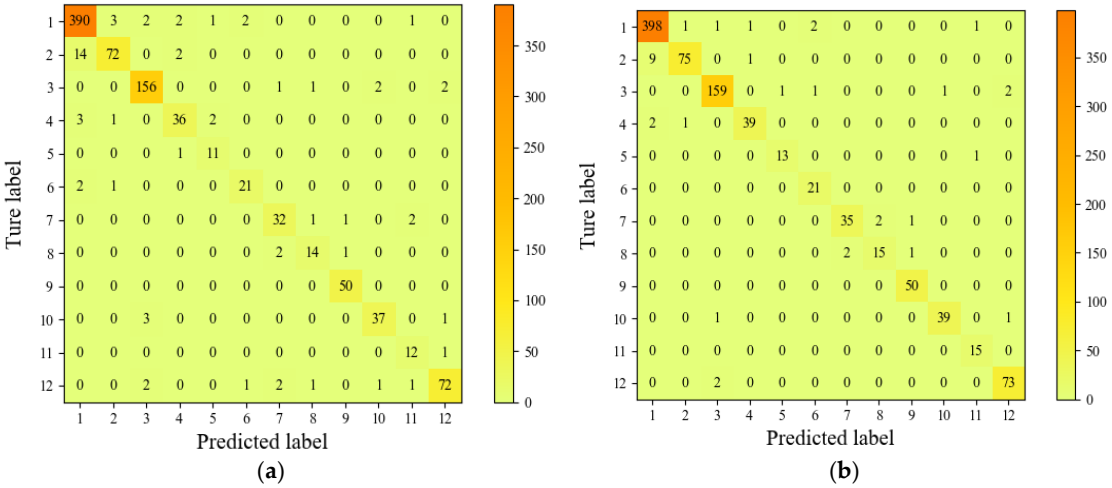


Figure 9. (**a**) Confusion matrix of BiLSTM-CBA model. (**b**) Confusion matrix of BiLSTM-MLCBA model.

The MLCBA algorithm adopts a new strategy in pruning. The correlation degree of $A_j \Rightarrow B$ is less than that of $A_i \Rightarrow B$. If $A_i \Rightarrow B$, $A_j \Rightarrow B$, and $A_j \subseteq A_i$, then $A_j \Rightarrow B$ needs to be pruned. For instance, the correlation degree of {"screw loosening, Seeping of oil"} $\Rightarrow$ {C2 Fault of "hydraulic cylinder assembly"} is greater than the two rules of {"Seeping of oil"} $\Rightarrow$ {C2 Fault of "hydraulic cylinder assembly"} and {"screw loosening"} $\Rightarrow$ {C2 Fault of "hydraulic cylinder assembly"}. According to the above principles, the latter two rules are pruned to reduce the low-quality rules and improve the diagnosis effect.

However, the MLCBA algorithm can avoid generating rules with high confidence, but there is a negative correlation between attributes and categories. For instance, if the traditional CBA algorithm is adopted, the confidence of {"Leakage of oil"} $\Rightarrow$ {C1 Fault of "The assembly of the motor oil pump"} is high, and the relevant fault data are directly classified into the category of a motor oil pump unit fault. The C1 faults of the motor oil pump unit account for more than 40% of the total, which is the large-probability fault in the unbalanced training dataset. Because the attribute of "Leakage of oil" is negatively correlated with C1, most training examples of C2, C4, and C6 have the common attribute of "Leakage of oil".

Figure 10 shows that the *Recall* of BiLSTM-MLCBA is improved by 4.08% compared with that of BiLSTM-CBA, and the diagnosis efficiency of BiLSTM-MLCBA is obviously improved because MLCBA can avoid generating poor-quality rules and misclassifying unknown data examples.

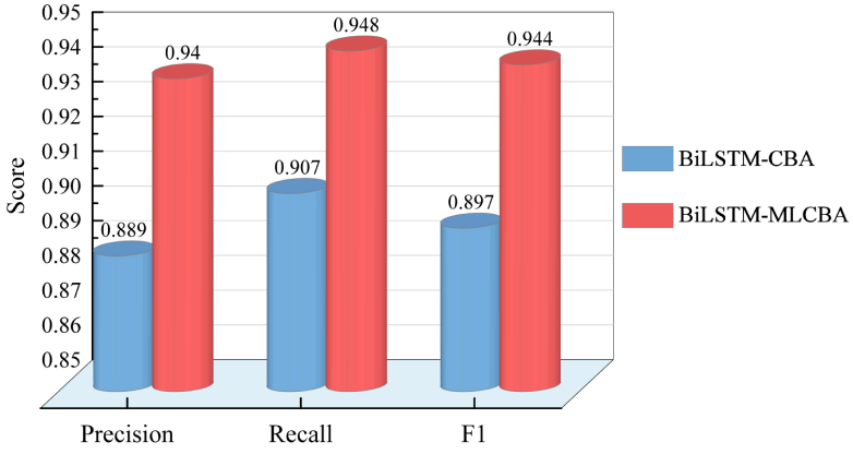


Figure 10. Comparison between BiLSTM-CBA and BiLSTM-MLCBA models.

5.2.2. Comparative Analysis of BiLSTM-CBA and Other Models

In this section, the BiLSTM-CBA model is compared with the TF-IDF algorithm [7], SVM model [10], PLSA topic model [23], and LSTM algorithm [36] to verify performance. The comparison results are shown in Figure 11.

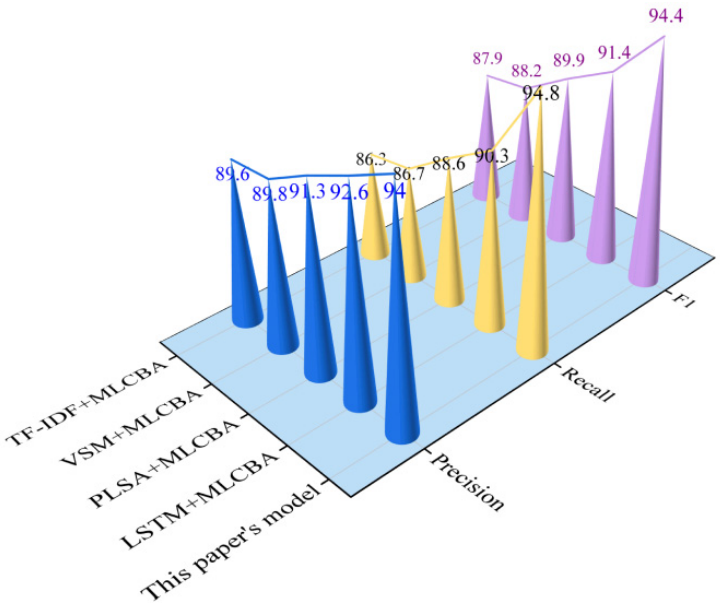


Figure 11. Test results of different text representation models.

Figure 11 shows that the feature extraction based on BiLSTM has better results for the three evaluation indicators than that based on LSTM. In addition, the accuracy of the MLCBA classifier is higher than that of the common SoftMax classifier, especially when the number of training samples is less than 1000, as shown in Figure 12.

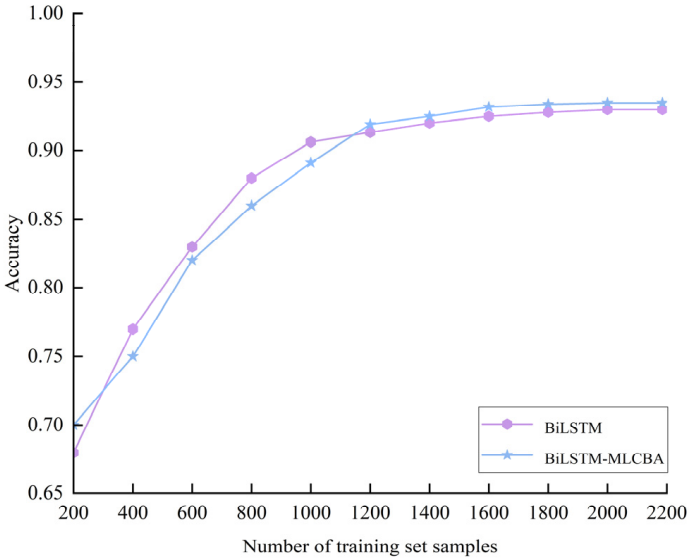


Figure 12. Comparison of classification results of MLCBA and SoftMax.

Furthermore, BiLSTM has other text classifiers, which are NB [37], KNN [38], and C4.5 [39]. These classifiers are used for comparison with the BiLSTM-MLCBA model. The comparative results are shown in Table 6.

Table 6. Comparative results of the four combined models.

Models	*Precision*	*Recall*	*F*1	Processing Speed/s
BiLSTM-MLCBA	0.9404	0.9478	0.9441	1.12
BiLSTM-NB	0.9235	0.8911	0.9070	0.98
BiLSTM-KNN	0.8989	0.8466	0.8719	1.38
BiLSTM-C4.5	0.8543	0.8481	0.8512	1.47

Compared with the BiLSTM-NB model, the *Precision*, *Recall*, and *F*1 of the BiLSTM-MLCBA model are increased by 1.69%, 5.67%, and 3.71%, respectively. The main reasons for this are that the NB algorithm needs to set prior probabilities in the classification process, and the prior probabilities depend on the assumptions. Hence, the BiLSTM-NB model with pre-assumptions can lead to a poor diagnosis rate. Next, the *Recall* of the BiLSTM-KNN model is the lowest among the four models. Therefore, its classification effect on unbalanced data is not obvious, and it needs to be recalculated with the training dataset and test data in each classification. Therefore, the efficiency of the BiLSTM-KNN model is too low, and the accuracy of classification is not high. As an algorithm for generating a decision tree, the C4.5 classifier does not consider the correlation between attributes, and the accuracy is low. Therefore, the C4.5 classifier is not suitable for the fault classification of a high-speed railway with ZYJ7 switch machine fault text data because the faults are interrelated.

For the above reasons, the proposed BiLSTM-MLCBA model is obviously superior to the other four models. The correlation algorithm of the association rules can perform better because the faults of a high-speed railway switch machine are caused by various fault modes. In consequence, the BiLSTM-MLCBA combined model can extract high-quality rules by learning with the training dataset multiple times and further applying the correlation degree to improve the quality of the rules. At the same time, despite the relative complexity of the proposed BiLSTM-MLCBA model structure, its processing speed is only slightly slower (0.14 s) than the fastest BiLSTM-NB. For the task of diagnosing faults in high-speed railway switches, it is already able to meet the requirements of high-speed processing. In this way, the BiLSTM-MLCBA combined model can not only improve the ability to diagnose a small-probability fault but can also guarantee the safe operation of a high-speed railway.

6. Conclusions

To improve the diagnosis accuracy of a small-probability fault for a high-speed railway switch machine, a combined BiLSTM and MLCBA model is proposed to deal with the uneven distribution of fault data for a high-speed railway ZYJ7 switch machine. The structured text data can be achieved with the feature extraction of fault text data with the BiLSTM model, which can use the two-layer LSTM model for feature extraction and to make preparations for fault feature classification. To obtain more high-quality association rules, the MLCBA algorithm is applied to deal with the small-probability fault data many times so that the rules of full fault types can be covered. Through the experiment of the discriminant analysis of the small-probability fault data and comparative analysis of the BiLSTM-CBA model and other models, the results show that the proposed high-speed railway switch machine fault diagnosis method based on BiLSTM and MLCBA is superior to other models, especially in terms of the evaluation indicators, *Precision*, *Recall*, and *F*1. In consequence, the high-speed railway switch machine fault diagnosis of the combined BiLSTM and MLCBA model can not only effectively improve the diagnosis accuracy of high-speed railway switch machine faults but can also ensure the safe and timely operation of a high-speed railway.

Author Contributions: Conceptualization, H.L. and N.H.; methodology, R.L. and T.Y.; data curation, Z.Z. and W.B.; writing—original draft preparation, N.H.; writing—review and editing, H.L.; visualization, Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Open Fund of the Key Laboratory of Four Power BIM Engineering and Intelligent Application Railway Industry (No. BIMKF-2022-02), the key research and development program of Gansu Province-Industry (No. 23YFGA0046), the State Key Laboratory of Rail Transit Engineering Informatization (No. SKLKZ22-06), and the Science and Technology Commission of Shanghai Municipality (No. 20DZ2251900, No. 21ZR1423800).

Data Availability Statement: The data that have been used are confidential.

Conflicts of Interest: Author Ran Lu was employed by the company CCCC Railway Design Research Institute Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. China Railway Kunming Group Co., Ltd. *Principle of Turnout Switching Machine and Failure Cases*; China Railway Publishing House: Beijing, China, 2022.
2. National Railway Administration of People's Republic of China. *Research and Investigation Points on the Causes of Railway Traffic Accidents*; China Railway Publishing House: Beijing, China, 2019.
3. Cao, Y.; An, Y.T.; Su, S.; Xie, G. A statistical study of railway safety in China and Japan 1990–2020. *Accid. Anal. Prev.* **2022**, *175*, 106764. [CrossRef] [PubMed]
4. Li, X.Q.; Zhang, P.X.; Shi, T.Y.; Li, P. Research on fault diagnosis method for high-speed railway signal equipment based on deep learning integration. *J. China Railw. Soc.* **2020**, *42*, 97–105.
5. Yang, L.B.; Shen, X.; Li, X.Q.; Dong, X.Z.; Xue, R.; Xu, G.H. Classification model of high-speed railway turnout failures based on text analysis. *China Railw.* **2020**, *8*, 13–18. [CrossRef]
6. Hamadache, M.; Dutta, S.; Olaby, O.; Ambur, R.; Stewart, E.; Dixon, R. On the fault detection and diagnosis of railway switch and crossing systems: An overview. *Appl. Sci.* **2019**, *9*, 5129. [CrossRef]
7. Liu, R.; Yang, B.; Zio, E.; Chen, X. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mech. Syst. Signal Process.* **2018**, *108*, 33–47. [CrossRef]
8. Zhang, T.; Chen, J.; Li, F.; Zhang, K.; Lv, H.; He, S.; Xu, E. Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions. *ISA Trans.* **2022**, *119*, 152–171. [CrossRef] [PubMed]
9. Dai, Q.J.; Chen, Y.G.; Tao, R.J. Research on PHM model of switch machine based on dynamic particle swarm optimization. *Railw. Stand. Des.* **2018**, *62*, 174–178. [CrossRef]
10. Eker, O.F.; Camci, F.; Kumar, U. SVM based diagnostics on railway turnouts. *Int. J. Perform. Eng.* **2012**, *8*, 289–298.
11. Zhang, Q.Q. Research on the design of expert system for fault diagnosis of railway signal equipment. *Technol. Dev. Enterp.* **2016**, *35*, 28–29.
12. Bian, C.; Yang, S.; Huang, T.; Xu, Q.; Liu, J.; Zio, E. Degradation state mining and identification for railway point machines. *Reliab. Eng. Syst. Saf.* **2019**, *188*, 432–443. [CrossRef]
13. Zhang, T.F. Intelligent analysis of the action current curve of S700K switch machine. *J. Manu. Auto.* **2014**, *36*, 71–74.
14. Lei, Y.T. Switch machine diagnostic system based on WPT And EMD incorporated feature extraction. *Mach. Build. Autom.* **2017**, *46*, 219–222. [CrossRef]
15. Liu, Y.J.; Si, Y.B.; Chen, G.W.; Wei, Z.S. Turnout fault diagnosis based on CDET/MPSO-SVM. *J. Beijing Jiaotong Univ.* **2021**, *45*, 52–59.
16. Wu, X.C.; Chu, X. Research on Division of Degradation Stage of Turnout Equipment Based on Wavelet Packet Decomposition and GG Fuzzy Clustering. *J. China Rail. Soc.* **2022**, *44*, 79–85.
17. Zhang, K. The railway turnout fault diagnosis algorithm based on BP neural network. In Proceedings of the IEEE International Conference on Control Science and Systems Engineering IEEE, Yantai, China, 29–30 December 2014; pp. 135–138.
18. Liang, X.; Wang, H.F.; Guo, J.; Xu, T.H. Bayesian network based fault diagnosis method for on-board equipment of train control system. *J. China Railw. Soc.* **2017**, *39*, 93–100.
19. Fan, L.H.; Wu, X.C.; Guo, R.C. Working State Evaluation Method of On-board Equipment of Train Control System Based on Association Rules and Variable Weight Coefficient. *J. Rail. Stand. Desi.* **2021**, *65*, 171–176.
20. Li, J.; Wang, J.P.; Xu, N.; Zhou, Z. Analysis of safety risk factors for metro construction based on text mining method. *Tunn. Constr.* **2017**, *37*, 160–166.
21. Zhou, L.J.; Dong, Y. Research on fault diagnosis method for on-board equipment of train control system based on GA-BP neural network. *J. Railw. Sci. Eng.* **2018**, *15*, 3257–3265. [CrossRef]
22. Zhao, Y.; Xu, T.H.; Zhou, Y.P. Text mining based fault diagnosis for vehicle on-board equipment of high-speed railway signal system. *J. China Railw. Soc.* **2015**, *37*, 53–59.

23. Zhong, Z.W.; Tang, T.; Wang, F. Research on fault extraction and diagnosis of railway based on PLSA and SVM. *J. China Railw. Soc.* **2018**, *40*, 80–87.
24. Hang, X.Y.; Liu, G.F.; Liu, X.Y.; Yang, A.; Ling, Y. Sentiment classification depth model based on word2vec and bi-directional LSTM. *Appl. Res. Comput.* **2019**, *36*, 3583–3587. [CrossRef]
25. Zhu, F.P.; Wang, X.F. Text classification for ship industry news. *J. Electron. Meas. Instrum.* **2020**, *34*, 149–155. [CrossRef]
26. Gao, H.; Zeng, X.; Yao, C. Application of improved distributed Naive Bayesian algorithms in text classification. *J. Supercomput.* **2019**, *75*, 5831–5847. [CrossRef]
27. Ge, S.; Zhuang, Y.; Hu, Y.; Ai, X. Research on enterprise hidden danger association rules based on text analysis. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *252*, 032170. [CrossRef]
28. Xie, M.J.; He, J.F.; Hu, X.X. Fault diagnosis for urban rail transit track side signaling equipment based on fault logs. *J. Beijing Jiaotong Univ.* **2020**, *44*, 27–35.
29. Ding, J.Q.; Li, B.; Qiao, Y. Crop disease diagnosis method based on multi-type data fusion of plant electronic medical records. *Trans. Chin. Soc. Agric. Mach.* **2023**, *54*, 196–204+223.
30. Xie, X.J.; Gu, B. Analysis of antimicrobial drug resistance based on deep learning. *J. Hunan Univ. (Nat. Sci. Ed.)* **2021**, *48*, 113–120.
31. Yuan, H.; Zhang, Q.; Tang, Q. A modified yield-based mean-variance model with survivorship bias. *J. Asset Manag.* **2019**, *20*, 145–157.
32. Xie, H.Y. Research and case analysis of apriori algorithm based on mining frequent Item-Sets. *Open J. Soc. Sci.* **2021**, *9*, 458. [CrossRef]
33. Ji, H.P.; Wang, T.Y.; Liu, J.; Fan, S.Y.; Wang, Z.P.; Zhang, K.R. An efficient parallel association rules mining algorithm for fault diagnosis. *Key Eng. Mater.* **2016**, *693*, 1326–1330. [CrossRef]
34. Lin, H.X.; Lu, R.; Lu, R.J.; Xu, L.; Zhao, Z.X.; Bai, W.S. Fault diagnosis for turnout of high-speed railway based on LDA-CLCBA hybrid model. *J. Electron. Meas. Instrum.* **2022**, *36*, 251–259. [CrossRef]
35. 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]
36. Zhu, Y.T.; Zhang, J.; Cao, X.B. Predicion of railway passenger ticket booking quantity based on ensembles of multi-step LSTM. *J. China Railw. Soc.* **2021**, *43*, 19–25.
37. Zhang, S.R.; Zhu, Z.B.; Feng, B. New channel selection and classification algorithm based on group sparse Bayesian logistic regression motor imagery EEG signal classification model. *Chin. J. Sci. Instrum.* **2019**, *40*, 179–191. [CrossRef]
38. Chen, Z.; Zhou, L.J.; Da, L.X.; Zhang, J.N.; Huo, W.J. The Lao text classification method based on KNN. *Procedia Comput. Sci.* **2020**, *166*, 523–528. [CrossRef]
39. Zhen, J.; Lee, D.J. Implementation of fatigue identification system using C4.5 algorithm. *J. Korea Converg. Soc.* **2019**, *10*, 21–26.