




Article

# Defect Texts Mining of Secondary Device in Smart Substation with GloVe and Attention-Based Bidirectional LSTM

Kai Chen <sup>1</sup>, Rabea Jamil Mahfoud <sup>1</sup>, Yonghui Sun <sup>1,\*</sup>, Dongliang Nan <sup>2,3</sup>, Kaike Wang <sup>3</sup>, Hassan Haes Alhelou <sup>4</sup> and Pierluigi Siano <sup>5,\*</sup>

<sup>1</sup> College of Energy and Electrical Engineering, Hohai University, Nanjing 210098, China; 18305178588@163.com (K.C.); rabea7mahfoud@hotmail.com (R.J.M.)

<sup>2</sup> School of Electrical Engineering, Xinjiang University, Urumqi 830047, China; ndl\_hhu@126.com

<sup>3</sup> Electric Power Research Institute, State Grid Electric Power Co., Ltd., Urumqi 830011, China; dlxjdky@126.com

<sup>4</sup> Department of Electrical Power Engineering, Faculty of Mechanical and Electrical Engineering, Tishreen University, Lattakia 2230, Syria; alhelou@tishreen.edu.sy or h.haesalhelou@gmail.com

<sup>5</sup> Department of Management & Innovation Systems, University of Salerno, 84084 Salerno, Italy

\* Correspondence: sunyonghui168@gmail.com (Y.S.); psiano@unisa.it (P.S.); Tel.: +86-139-0516-9126 (Y.S.)

Received: 17 July 2020; Accepted: 27 August 2020; Published: 1 September 2020



**Abstract:** In the process of the operation and maintenance of secondary devices in smart substation, a wealth of defect texts containing the state information of the equipment is generated. Aiming to overcome the low efficiency and low accuracy problems of artificial power text classification and mining, combined with the characteristics of power equipment defect texts, a defect texts mining method for a secondary device in a smart substation is proposed, which integrates global vectors for word representation (GloVe) method and attention-based bidirectional long short-term memory (BiLSTM-Attention) method in one model. First, the characteristics of the defect texts are analyzed and preprocessed to improve the quality of the defect texts. Then, defect texts are segmented into words, and the words are mapped to the high-dimensional feature space based on the global vectors for word representation (GloVe) model to form distributed word vectors. Finally, a text classification model based on BiLSTM-Attention was proposed to classify the defect texts of a secondary device. Precision, Recall and F1-score are selected as evaluation indicators, and compared with traditional machine learning and deep learning models. The analysis of a case study shows that the BiLSTM-Attention model has better performance and can achieve the intelligent, accurate and efficient classification of secondary device defect texts. It can assist the operation and maintenance personnel to make scientific maintenance decisions on a secondary device and improve the level of intelligent management of equipment.

**Keywords:** secondary device; defect classification; GloVe; attention mechanism; text mining

## 1. Introduction

A secondary device in smart substation refers to the low-voltage electrical device that monitors, controls and protects the operating state of the primary equipment such as transformers and generators. As one of the key components of the equipment of power transmission system, the health of the secondary device in smart substation directly affects the secure and stable operation of the whole system [1–3].

During the long-term operation of relay protection devices, fault recorders and other secondary devices, a large number of defects, fault reports, maintenance and defect elimination documents

accumulate through inspection, test and other ways [4,5]. This kind of data contains a huge amount of equipment online operation information, historical action information, defect information, overhaul information and other full-service operation information, which have an important guiding significance for the objective evaluation of equipment health status [6]. However, due to the features of multi ambiguity, difficult segmentation and fuzziness, the above information has not been fully mined [7–9]. When the equipment fails to operate properly, it will send out a large number of alarm signals. If the maintenance personnel cannot make a decision in a short time, the fault range might be expanded [10]. For the real-time state evaluation and defect degree research of a secondary device, many researchers at home and abroad have presented relevant research work, which can be classified into the two following categories: machine learning method based on statistics and deep learning method based on network [11,12].

Machine learning methods based on statistics mainly include support vector machine (SVM), K-means clustering, decision tree, Apriori and so on [13]. For example, in [14], a secondary device state evaluation model based on fuzzy comprehensive SVM was proposed, which improved the traditional SVM and corrected the errors caused by subjective factors. In [15], based on the historical defect data of a secondary device in smart substation, a method of mining and analyzing the defect data of secondary device based on the Apriori algorithm was proposed, which provided reference for equipment operation and control. Machine learning methods can process structured data efficiently. However, as a traditional shallow learning model, its ability to express complex functions is limited. For complex classification problems, its generalization ability is restricted to some extent. They are not good at processing unstructured data such as text and image.

Furthermore, in recent years, with the development of deep learning and natural language processing (NLP) technology, more experts and scholars have applied it to the field of safety assessment for power equipment [16]. According to the characteristics of defect texts in power devices, a defect text classification model based on convolutional neural network (CNN) was established in [17], and the model's accuracy was obviously higher than that of the traditional machine learning method. In [18], considering that it is difficult to fully mine and use the text data of a power device, an operation and maintenance text information mining method based on deep semantic learning was proposed, and the effect of text classification was significant. However, it should be pointed out that in these mentioned results, the proposed models had some shortcomings in text representation and semantic information extraction, which is easy to cause the loss of text semantic information. Therefore, in the process of feature extraction and mining classification, how to obtain better text representation and keep the semantic information of text is the key to enhance the accuracy of device defect classification.

Based on the above discussion, in order to enhance the ability of text semantic information representation and feature extraction, and improve the accuracy of secondary device defect classification, a defect texts information mining method for secondary devices in smart substation is proposed. By combining global vectors for word representation (GloVe) and attention-based bidirectional long short-term memory (BiLSTM-Attention), the integrity of semantic information in the text representation stage and feature extraction stage is effectively guaranteed, and distributes weight reasonably according to the importance of semantic information, so as to realize the accurate classification and evaluation of secondary device in smart substation. It will greatly reduce the workload of operation and maintenance personnel and equipment maintenance costs, improve the level of the intelligent management of equipment, and provide a scientific and reasonable judgment basis for equipment maintenance and decision-making. The main contributions in this paper are:

- o Considering the incompleteness of the existing word segmentation thesaurus, a professional dictionary in the field of secondary equipment is constructed to achieve the efficient and accurate word segmentation of defect texts, which effectively ensures the integrity of the semantic information of the text in the word segmentation stage.
- o In order to solve the problems of low training efficiency and an easy loss of semantic information of word2vec, FastText and other models, the GloVe model based on global corpus statistics is

used to vectorize defect texts. It takes global information into full consideration and ensures the integrity of semantic information.

- o The attention mechanism is innovatively introduced into the field of power equipment defect text mining, and a defect text classification model based on BiLSTM-Attention is proposed, which improves the ability of feature extraction and the mining of text information. The classification accuracy of the model was up to 94.9%, which is 5%~37% higher than the traditional classification models such as TextCNN and Decision Tree.

The rest of this paper is arranged as follows: Section 2 introduces the natural language characteristics of the defect texts of the secondary device in smart substation. Section 3 describes text segmentation, where a distributed text representation method based on the GloVe model is proposed. In Section 4, a semantic information extraction and text classification model based on BiLSTM-Attention is proposed. The classification effect of the model is evaluated in Section 5 by using the defect texts data for a secondary device of a power plant in China, and the model performance is also compared with other algorithms. Section 6 concludes the paper.

## 2. Defect Texts of Secondary Device

### 2.1. Description and Grading of Defect Texts

The defect of the secondary device in smart substation mainly refers to the state which may or has affected the secure and stable operation of relay protection and automatic safety devices, related equipment and their secondary circuits. In order to strengthen the statistical analysis and control of secondary device defects, accurately grasp its operation conditions, and improve the health level of equipment, equipment operation and maintenance personnel constantly summarize and find the problems in the installation, commissioning, operation and maintenance link from the aspects of commissioning, the operation and regular inspection of equipment, and finally summarize to form the defect record text.

Defects must be entered into the production management system (PMS) or operations management system (OMS) within 72 h after detection. The professionals of dispatch and control centers at all levels shall approve the defects, analyze the causes of defects, grade and adjust the defects, and assign the responsible units for defect handling.

The defect records of the secondary device in smart substation include the following key information: device classification, voltage level, device model, manufacturer, defect classification, defect position, specific defect situation, defect reason, defect elimination time, treatment result, defect times, and the defect rate.

According to the “Measures of State Grid Corporation of China for defect management of relay protection and automatic safety devices”, the defects of relay protection and safety automatic device put into operation (including trial operation) are basically divided into three levels: critical defects, serious defects and general defects. The specific status divisions and their corresponding maintenance decision are listed in Table 1.

**Table 1.** Definition of secondary device status.

Level of Equipment Defects	Maintenance Decision
General Defect	No maintenance required.
Serious Defect	Real time monitoring of equipment operation status, priority to arrange maintenance.
Critical Defect	Immediately cut off the power for maintenance.

### 2.2. Natural Language Characteristic of Defect Texts

The form of the defect record of secondary devices in the smart substation is mainly Chinese short text [17], which consists of four parts: defect description, device panel signal, background signal and

remote monitoring signal. A typical defect record of a secondary device in smart substation is depicted in Figure 1.

Defect description	Device panel signal	Background signal	Remote monitoring signal
The operation indicator of 220kv line protection in Hefeng Substation does not light up.	The device program has an error, and the running indicator does not light up.	Abnormal operation of protection device.	The device RCS-931 of Fengtie line had abnormal operation.

**Figure 1.** The example of typical defect record.

Compared with the general text, secondary device defect texts have the following characteristics:

- (1) The content of defect texts is unstructured Chinese short text, which is mixed with many numbers and symbols;
- (2) The text involves the field of electric power, including a large number of secondary device professional vocabulary;
- (3) The content of the text record is slightly different, and the length of a text varies from a few words to a hundred words;
- (4) The content and format of records vary from person to person, and different operation and maintenance personnel have different descriptions of the same phenomenon.

If we do not consider the natural language characteristics of secondary device defect texts, it is easy to cause text segmentation errors, which result in the loss of semantic information and greatly reduce the model classification performance. In this paper, when building the text classification model of secondary device defects, the above characteristics are fully considered, and the targeted text data processing method is adopted, so that the model is more suitable for the task of a secondary device in smart substation defect text classification.

### 3. Text Representation

#### 3.1. Cleaning and Pretreatment

At present, the defect texts of a secondary device in smart substation are mainly recorded by operation and the maintenance personnel on site. In the process of analysis, summary and upload, it is easy to make mistakes, such as inconsistent record format, text content loss and so on, which result in greatly reduced text quality. Therefore, it is necessary to implement data cleaning before text mining [19,20].

The cleaning of secondary device defect texts basically includes the following steps:

**Filter useless characters.** Defect texts usually contain a lot of spaces, punctuation and other characters that are not related to the text content, so they need to be filtered.

**Unified lowercase of English characters.** There are many English characters in the defect texts of the secondary device, and there is a lack of standardization in the record format. The most common are descriptions of transformer levels, such as "220 KV", "220 kV" and "220 Kv". They all represent the same voltage level. However, if the format is not unified, the vector representation and semantic information of the text trained by GloVe will be different.

**Detect and remove duplicate records and incomplete texts.** In the process of uploading defect records, operation and maintenance personnel are prone to data loss, data re-entry and other problems due to improper operation. This kind of data is not conducive to text classification and information mining, and needs to be processed in advance to ensure the quality of the text.

### 3.2. Texts Segmentation

Word segmentation is usually the first step of Chinese text information mining. In this paper, the open source word segmentation tool based on Python is used, Jieba, to segment the defect text.

There is a large number of professional words in the fields of power system, place names, equipment names and other proper terms in the secondary device defect texts of smart substation. Therefore, only based on the existing word segmentation tool and its own thesaurus, it will be difficult to achieve the accurate segmentation of the defect text.

As shown in Table 2, the defect texts are segmented directly using the Jieba segmentation tool, and there are two obvious segmentation errors in the segmentation results. One is the incorrect division of the protection line name. The proper name of the protection line “wumi line” is directly classified as “wu” and “mi line”. The second is the improper division of proper nouns in the field of power systems, “high frequency protection” is incorrectly divided into “high frequency” and “protection.” The incorrect segmentation of proper nouns will change its original meaning and greatly affect the effect of text representation.

**Table 2.** The example of text segmentation.

Text to be Segmented	Segmentation Result
(Abnormal high frequency protection channel of 220 kv wumi line)	(220 kv/wu/miline/high frequency/protection/channel/abnormal)

In order to further improve the accuracy of defect text segmentation, a secondary device professional dictionary was constructed. The dictionary is based on the existing secondary device-related literature, guidelines and national standard documents, combined with the experience of experts and equipment operation as well as maintenance personnel. Some examples of dictionaries are given in Table 3. Based on the construction of a secondary device professional dictionary, the defect text is segmented more accurately.

**Table 3.** Examples of secondary device professional dictionary.

Category Name	Proper Noun Set
Transformer Name	(duolang substation) (wusu substation) (bachu substation) (wucaiwan substation)
Name of Protection Line	(chuba line) (yaningzhou line) (qichai line)
Terminology of Equipment Protection	(high-frequency protection) (intelligent terminal) (automatic safety device) (high resistance protection) (stability control device)

### 3.3. Text Representation Based on GloVe

The primary problem of NLP is to find an appropriate text representation. A good text representation is the basis of further mining text information. At present, the mainstream text representation methods include word2vec [21–23], FastText [24] and GloVe [25].

GloVe is a new word vector model based on the theory of the word co-occurrence matrix. As an improvement of the word2vec model, it can synthesize the global and local statistical information of words to generate word vectors. GloVe model can properly capture the semantic information of words and improve the accuracy of the expression of word vector semantic information.

### 3.3.1. Construct Co-Occurrence Matrix

The co-occurrence matrix  $X$  is constructed according to the defect text corpus. Each element  $X_{ij}$  in the matrix represents the number of times that word  $i$  and its context word  $j$  appear together in a specific size context window. Taking the defect text “The cpu plug-in of 220 kv rainbow line merging unit is abnormal” as an example, a sliding window with a width of 3 is adopted, and the statistics of window contents are explained in Table 4.

**Table 4.** Sliding window content statistics of the sample corpus.

Window Label	Central Word	Window Contents
0	220 kv	220 kv—(rainbow line)
1	rainbow line	220 kv—(rainbow line)—(merging unit)
2	merging unit	(rainbow line)—(merging unit)—cpu
3	cpu	(merging unit)—cpu—(plug-in is abnormal)
4	plug-in is abnormal	cpu—(plug-in is abnormal)

When the central word is “cpu” and its context words are “merging unit” and “plug-in is abnormal”, execute:

$$X_{\{\text{cpu,merging unit}\}^+} = 1 \quad (1)$$

$$X_{\{\text{cpu,plug-in is abnormal}\}^+} = 1 \quad (2)$$

Through the window traversing the whole defect text corpus composed of defect reports, fault reports, maintenance and elimination files related to the running status of secondary equipment, the co-occurrence matrix  $X$  can be obtained.

### 3.3.2. Word Vector Training

GloVe constructs the relationship with the word vector through a co-occurrence matrix, which is presented by

$$v_i^T v_j + B_i + B_j = \log(X_{ij}) \quad (3)$$

where  $v_i$  is the word vector of the target word,  $v_j$  is the word vector of the context word.  $B_i$  and  $B_j$  are the bias of the two-word vectors, respectively.

Based on the principle that the higher the frequency of the occurrence is, the greater the weight of words are, the weight term is added to the cost function  $J$ , as given by

$$J = \sum_{i,j=1}^N f(X_{ij}) (w_i^T w_j + b_i + b_j - \log(X_{ij}))^2 \quad (4)$$

where  $N$  is the size of the vocabulary, and the weight function  $f(x)$  is shown as

$$f(x) = \begin{cases} (x/X_{\max})^{0.75}, & \text{if } x < X_{\max} \\ 1, & \text{if } x \geq X_{\max} \end{cases} \quad (5)$$

## 4. Text Classification Model

### 4.1. Network Architecture of LSTM

In the field of NLP, most of the mainstream text classification algorithms are based on convolutional neural network (CNN) and recurrent neural network (RNN) [26–29]. As a variant of RNN, long short-term memory (LSTM) solves the problem of the long-term dependence of the RNN model and the problem of gradient disappearance and explosion caused by a too long sequence.

The network structure of the LSTM time  $t$  is demonstrated in Figure 2. Among them,  $X(t)$  and  $h(t)$  are the input and output of LSTM at time  $t$ .  $h(t-1)$  and  $c(t-1)$  are the output of the hidden layer and historical information at time  $t-1$ .  $f_t$ ,  $i_t$  and  $o_t$  are the forget gate, input gate and output gate at time  $t$ .  $\tilde{c}_t$  is the new information after transformation at time  $t$ , and  $c(t)$  is the updated historical information at time  $t$ . The specific calculation process is as follows:

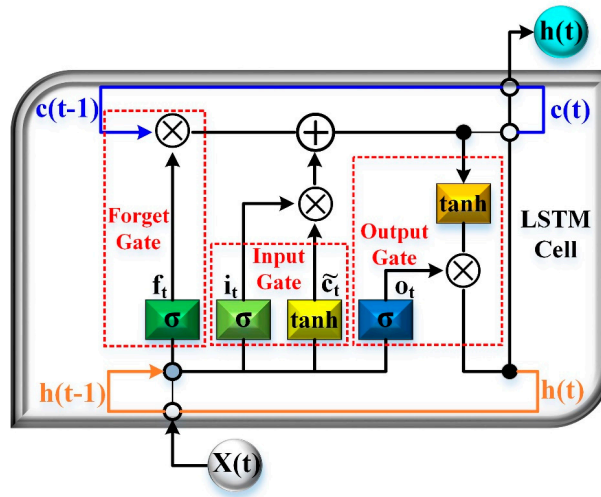


Figure 2. Internal network structure of long short-term memory (LSTM).

- (1) The forget gate screens the information of  $X(t)$  and  $h(t-1)$ :

$$f_t = \sigma(W_f X(t) + U_f h(t-1) + b_f) \quad (6)$$

where  $W_f$ ,  $U_f$ , and  $b_f$  are the cycle weight, input weight, and bias of the forget gate, respectively.

- (2) The input gate determines the values that need to be updated at this time and the new information to be added:

$$i_t = \sigma(W_i X(t) + U_i h(t-1) + b_i) \quad (7)$$

$$\tilde{c}_t = \tanh(W_c X(t) + U_c h(t-1) + b_c) \quad (8)$$

where  $W_i$ ,  $U_i$  and  $b_i$  represent the cycle weight, input weight, and bias of the input gate,  $W_c$ ,  $U_c$  and  $b_c$  represent the cycle weight, input weight, and bias of the cell, respectively.

- (3) Integrate the information of the input gate and the forget gate to update internal memory cell state  $c(t)$ :

$$c(t) = f_t \odot c(t-1) + i_t \odot \tilde{c}_t \quad (9)$$

- (4) Use the output gate  $o_t$  to control the information output of the internal memory cell:

$$o_t = \sigma(W_o X(t) + U_o h(t-1) + b_o) \quad (10)$$

where  $W_o$ ,  $U_o$ , and  $b_o$  are the cycle weight, input weight, and bias of the output gate, respectively.

- (5) The output of the LSTM network node is determined by the activation function and output gate:

$$h(t) = o_t \odot \tanh(c(t)) \quad (11)$$

#### 4.2. Extracting Semantic Information Based on BiLSTM

Considering that LSTM can only extract single direction text semantic information, leading to some information loss. The BiLSTM network is used for the feature extraction of text semantic information.

BiLSTM is a new network structure improved on the basis of LSTM [30–32]. It consists of the following four layers: input layer, forward LSTM layer, backward LSTM layer and the output layer. BiLSTM consists of two LSTM layers: forward flow and reverse flow to ensure that it can effectively extract the text's context information. The calculation flow of each unit is the same as that of LSTM, and the final output is determined by the LSTM state in the front and back directions:

$$\vec{h}_t = LSTM(X_t, \vec{h}_{t-1}) \quad (12)$$

$$\overleftarrow{h}_t = LSTM(X_t, \overleftarrow{h}_{t-1}) \quad (13)$$

$$h_t = w_t \vec{h}_t + v_t \overleftarrow{h}_t + b_t \quad (14)$$

where  $\vec{h}_t$  is the forward output of the LSTM at time  $t$ ,  $\overleftarrow{h}_t$  is the reverse output of LSTM at time  $t$ ,  $h_t$  is the output of the BiLSTM at time  $t$ ,  $X_t$  is the input at time  $t$ ,  $w_t$  is the weight matrix of forward output,  $v_t$  denotes the weight matrix of reverse output, and  $b_t$  is the offset at time  $t$ .

#### 4.3. Model Optimization Method Based on Attention Mechanism

In the text, different words have different semantic contributions to the sentence. Based on the attention mechanism [33–35], this paper allocates the weight according to the contribution degree of words to the text semantics, so that the classifier pays more attention to the semantic information related to the defect degree and improves the classification performance of the model.

By inputting the output  $h_t$  of BiLSTM into a single-layer perceptron, the result  $u_t$  is taken as the implicit representation of  $h_t$ :

$$u_t = \tanh(W_w h_t + b_w) \quad (15)$$

where  $W_w$  and  $b_w$  denote the weight and offset of attention, respectively.

The similarity between  $u_t$  and a randomly initialized context vector  $u_w$  is used to measure the importance of words, and the attention weight matrix  $\alpha_t$  is obtained by normalization:

$$\alpha_t = \frac{\exp(u_t^\top u_w)}{\sum_t \exp(u_t^\top u_w)} \quad (16)$$

Based on the attention weight matrix, the word vector of a sentence is weighted and summed to get the semantic vector  $S$  at the sentence level:

$$S = \sum_t \alpha_t h_t \quad (17)$$

Finally, the BiLSTM-Attention defect texts classification model integrating GloVe is depicted in Figure 3.



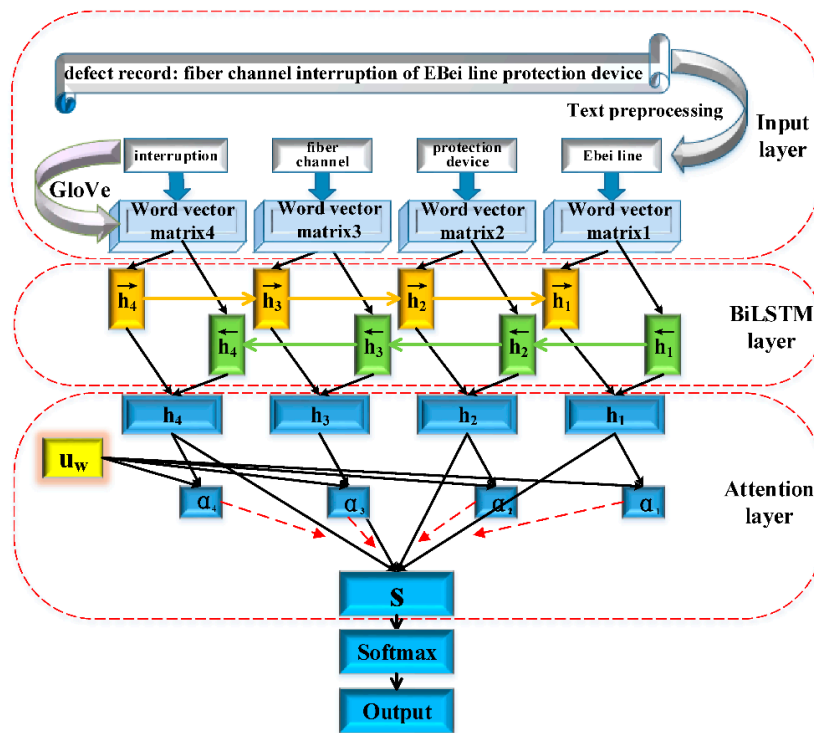


Figure 3. Architecture of global vectors for word representation (GloVe)-BiLSTM-Attention.

### 5. Case Study and Analysis

In order to validate the classification performance of the proposed model for secondary device defect texts, a power plant in China was selected as the sample data for secondary device defect records from 2015 to 2019, where a total of 1800 typical defect data were collected. The distribution of each category is illustrated in Figure 4.

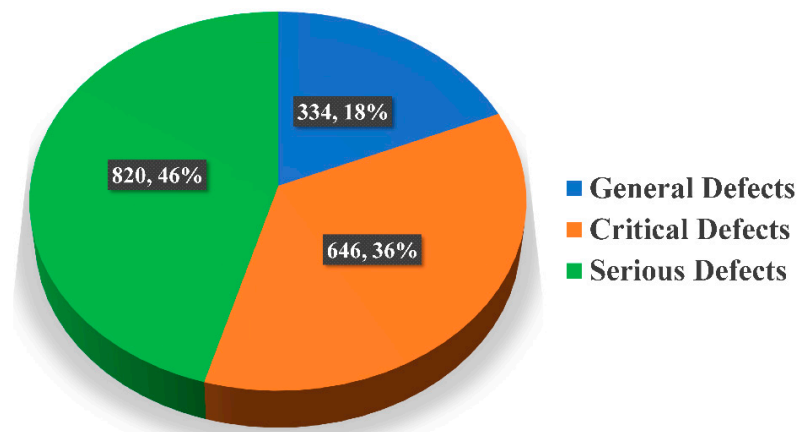


Figure 4. Distribution of three types of defect texts.

After data cleaning, word segmentation and text vectorization, the defect texts were randomly divided into three sets: training, verification, and test set, using the ratio of 3:1:1, as listed in Table 5.

**Table 5.** Division of sample data.

Defect Classification	Training Set	Verification Set	Test Set
General Defect	200	67	67
Serious Defect	492	164	164
Critical Defect	388	129	129

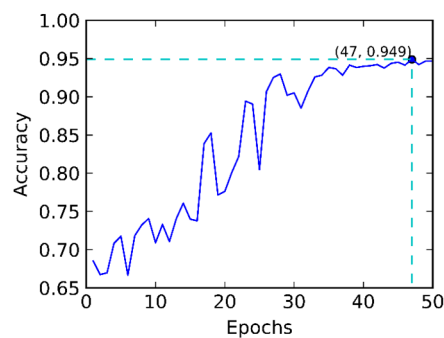
Among them, the training set is used to fit the model as well as train the classification model by setting the classifier's parameters. The verification set is used to adjust the parameters of the model and optimize the performance of the model, while the test set is utilized to measure the model's performance and classification ability.

Based on the Keras deep learning framework, a BiLSTM-Attention model is constructed for defect texts classification. Considering the experience of model super parameter setting in [36,37], based on the grid search method [38–40], the parameters of the model are optimized. The model's optimal parameter settings are obtained as shown in Table 6.

**Table 6.** Optimal parameter setting of the classification model (BiLSTM-Attention).

Parameter	Value
Learning Rate	0.01
Dropout Rate	0.5
L2RegLambda	0.5
Number of Neurons in LSTM Hidden Layer	64
Batch Size	64
Word Vector Dimension	200

The defect texts are classified by the optimal classification model obtained from the super parameter optimization of the model. As illustrated in Figure 5, with the increase in iterations, the model classification accuracy tends to converge. When the iteration number is 47, the classification accuracy reaches 94.9%.

**Figure 5.** Trend of model classification accuracy.

The performance of the classification model is evaluated by a confusion matrix, which is depicted in Figure 6. Based on the real category and prediction category, it is divided into TP (true positive), TN (true negative), FP (false positive) and FN (false negative). Precision, Recall and F1-score were utilized to test the classification accuracy as follows:

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (18)$$

$$Recall = \frac{TP}{(TP + FN)} \times 100\% \quad (19)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\% \quad (20)$$

		Actual Value (as confirmed by experiment)	
		Positive	Negative
Predicted Value (predicted by the test)	Positive	TP True Positive	FP False Positive
	Negative	FN False Negative	TN True Negative

Figure 6. Confusion matrix.

Precision is the ratio of the number of samples predicted to be positive to all the actual positive samples. Recall is the ratio of the number of samples predicted to be positive to the actual number of such samples.

F1-score can be regarded as a weighted average of model Precision and Recall. It takes into account both the Precision and Recall of the classification model. The higher the F1-score, the better the classification performance of the model.

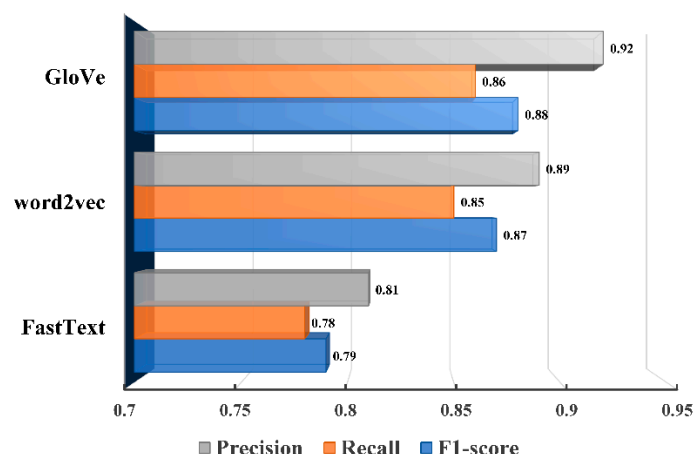
### 5.1. Performance Comparison of Word Vector Representation

The first step of text classification is text representation. The accuracy and integrity of text representation directly determine the quality of the text classification results. In order to compare the current text representation models for defect texts representation, we choose FastText and word2vec to compare with GloVe, and used the BiLSTM-Attention model with the same parameter settings for text classification. The optimal parameter settings of GloVe, word2vec and FastText are obtained as listed in Table 7.

Table 7. Parameter settings of the word vector training models.

GloVe		word2vec		FastText	
Parameters	Value	Parameters	Value	Parameters	Value
vector size	200	vector size	200	vector size	200
window size	8	window size	6	window size	5
min count	2	min count	2	min count	2
batch_size	512	batch_size	512	batch_size	512
xmax	100	sg	1(skip-gram)	sg	1(skip-gram)
$\alpha$	0.75	hs	1(negative sampling)	loss function	hierarchical softmax

As shown in Figure 7, under the condition that the classification model and its parameter settings remain unchanged, each performance index of text classification based on the GloVe–BiLSTM-Attention is significantly higher than that of FastText–BiLSTM-Attention and slightly better than that of word2vec–BiLSTM-Attention.



**Figure 7.** Model performance comparison based on the different word vector representation models.

The results show that in terms of text semantic representation performance, GloVe is better than word2vec and FastText in the defect text data set of the secondary device in smart substation.

FastText is actually an extension of word2vec, the core of which is n-gram. On the basis of word2vec, the author adds n-gram information to transform the vector of the central word in word2vec into the vector representation of the central word n-gram [41]. Because of the n-gram information, FastText can solve the OOV (out-of-vocabulary) problem to some extent. However, research shows that FastText is more suitable for languages with rich forms, such as Russian, Turkish, French, etc. The effect of Chinese may not be as good as the aforementioned language, and this experiment just shows this point. As an improvement of the word2vec model, GloVe takes more consideration of the global information of the text. For some polysemy words, it will give more detailed and different vector representations, and at the same time, it will take into account the global information of the text. Therefore, in the task of vectorizing the text sentences, it can stand out and obtain better results.

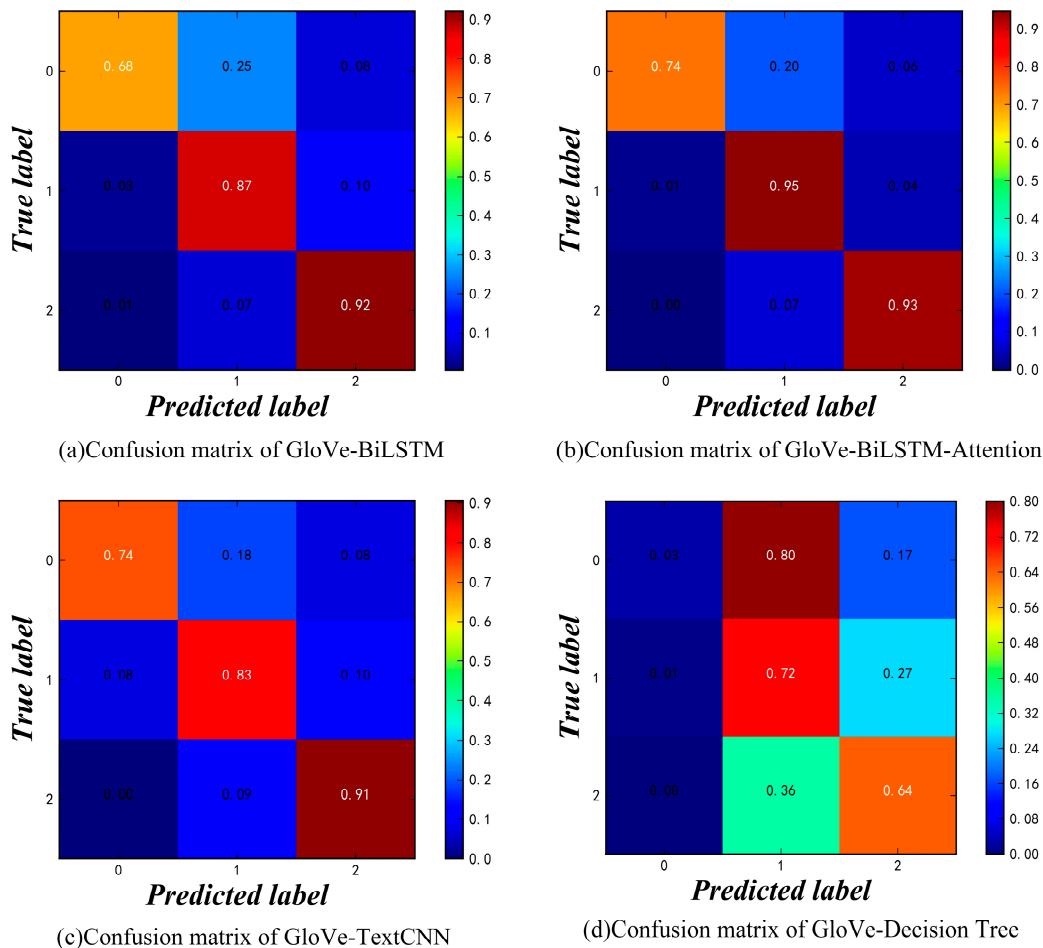
## 5.2. Performance Comparison of Different Classification Models

In order to verify the superiority of the proposed model in this paper, many typical classification models in the fields of machine learning and deep learning were selected for comparative experiments. Among them, the machine learning algorithm chose the decision tree model, and the two deep learning algorithms are TextCNN and BiLSTM. Each model uses the GloVe model to quantify the defect texts in the text representation stage, and compare the classification performance of the model under the condition of ensuring the model's optimal parameters. The parameter settings of BiLSTM and BiLSTM-Attention refer to Table 6 and the optimal parameter settings of TextCNN and decision tree are obtained as listed in Table 8.

**Table 8.** Parameter setting of the classification models (TextCNN and decision tree).

Text Classification Model	Parameter	Value
TextCNN	Filter Size	(2,3,4)
	Pooling	Max_pooling
	Dropout Rate	0.2
	Iter	5
	Optimization	Gradient Descent
	Batch Size	64
Decision Tree	criterion	gini
	splitter	best
	min_samples_split	2
	min_samples_leaf	1

The confusion matrix of the classification model is shown in Figure 8. The labels 0, 1 and 2 in the figure represent the general, serious, and critical defects, respectively. Comparing the confusion matrices of the four classification models, it can be seen that the classification accuracy of BiLSTM-Attention for three types of defect text is higher than that of the other three models.



**Figure 8.** Confusion matrix of the classification models.

Because the depth of decision tree model is shallow and the ability of semantic information extraction is insufficient, the accuracy of text classification with general defects is only 3%. GloVe-BiLSTM and GloVe-TextCNN have some semantic information extraction ability, but compared with the proposed model in this paper, there are still some problems of semantic information extraction and semantic loss.

Since the number of general defect texts is small, resulting in insufficient model training, GloVe-BiLSTM, GloVe-BiLSTM-Attention and GloVe-TextCNN have low accuracy for text classification with general defect label.

Figure 9 is a comparison diagram of the performance evaluation indicators of several defect text classification models. It can be noted from Figure 9 that the performance indexes of the proposed model are higher than those from the other three models. Among them, the comprehensive performance index F1-score of GloVe-BiLSTM-Attention model is 5–6% which is higher than that of GloVe-BiLSTM and GloVe-TextCNN, and 37% which is higher than that of GloVe-decision tree. Combined with the training time of the classification models in Table 9, although the training time of BiLSTM-Attention is 22 s longer than that of BiLSTM, it improves the classification performance of the model, which is more important. However, the training time of TextCNN is longer than those of BiLSTM and BiLSTM-Attention, and the classification performance is poor. The Decision Tree model can achieve fast text classification, but its

classification accuracy is too low compared with the other three models. It can be seen that the proposed model in this paper can classify and evaluate the defect texts accurately. Furthermore, it is feasible as well as practical in the actual defect texts' classification of device in power system.

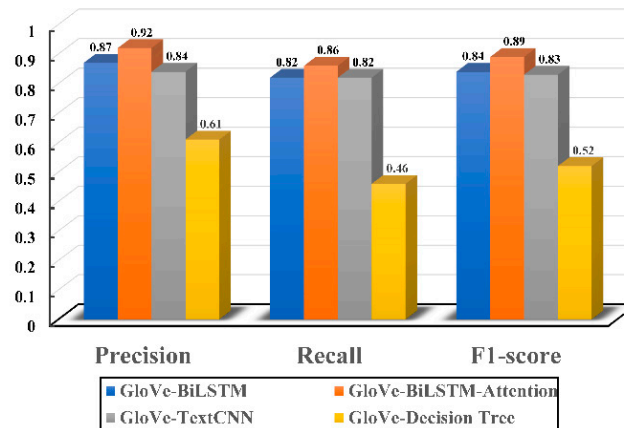


Figure 9. Performance index of the classification models.

Table 9. Training time of the classification models.

Classification Model	Training Time
BiLSTM	8 min 50 s
BiLSTM-Attention	9 min 12 s
TextCNN	10 min 20 s
Decision Tree	5 min 30 s

## 6. Conclusions

In the view of the defect texts of power equipment represented by a secondary device in smart substation, the defect texts are intelligently classified by combining NLP and the deep learning model, so as to reduce the maintenance cost and workload of maintenance personnel, and assist them to make scientific and reasonable maintenance decisions.

In this paper, a defect text classification model based on GloVe and BiLSTM-Attention was proposed, which was evaluated on the basis of Precision, Recall and F1-score, and compared with decision tree, TextCNN and BiLSTM. The analysis of a case study shows that compared with word2vec and FastText, GloVe can better ensure the integrity of the semantic information in the text representation stage. Compared with the traditional machine learning model (decision tree), the deep learning model can extract the deep features of the text semantic information. The BiLSTM model with attention mechanism can automatically assign weight according to the importance of semantic information, weaken invalid information, and effectively improve the feature extraction capability of the model. Therefore, compared with TextCNN and BiLSTM, BiLSTM-Attention has better classification performance.

The quality of the text determines the classification performance of the model, and how to build a high-quality defect text corpus is the next research direction. Studying the matching and the screening of similar defects in power equipment and constructing the knowledge graph of the power system is the direction that can be considered in future research.

**Author Contributions:** K.C. conceived and designed the experiment and wrote the original manuscript. Y.S. performed the experiments and evaluated the data. D.N., K.W., R.J.M., H.H.A. and P.S. reviewed and proofread the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Science and Technology Project of State Grid Xinjiang Electric Power Co., Ltd. 5230DK20000D (Key Technologies for Evaluation of State and Operation of Secondary Device Based on Multi-source Information Fusion).

**Conflicts of Interest:** The authors declare no conflict of interest.

## Nomenclature

$\alpha_t$	the attention weight matrix after normalization
$B_i, B_j$	the bias of word vectors $i$ and $j$
$b_f, b_i, b_o, b_c$	the bias of forget gate, input gate, output gate and cell
$b_t, b_w$	the bias of BiLSTM and attention model
$c(t-1), c(t)$	the cell state of the LSTM at time $t-1$ and $t$
$\tilde{c}_t$	the new information after transformation at time $t$
$f(x)$	the weight function
$f_t, i_t, o_t$	the forget gate, input gate and output gate at time $t$
$h(t-1), h(t)$	the output of LSTM at time $t-1$ and $t$
$\vec{h}_t, \overleftarrow{h}_t$	the forward and reverse output of LSTM at time $t$
$h_t$	the output of BiLSTM at time $t$
hs	hierarchical softmax
$J$	the cost function of the GloVe model
$N$	the size of the vocabulary
$\sigma$	the sigmoid function
$S$	the semantic vector at the sentence level
sg	skip-gram
$U_f, U_i, U_o, U_c$	the input weight of forget gate, input gate, output gate and cell
$u_t$	the implicit representation of the output of BiLSTM
$u_w$	a randomly initialized context vector
$W_f, W_i, W_o, W_c$	the cyclic weight of forget gate, input gate, output gate and cell
$w_t, v_t$	the weight matrix of forward output and reverse output in BiLSTM
$W_w$	the weight of attention model
$v_i, v_j$	the word vectors of word $i$ and $j$
$X$	the co-occurrence matrix
$X_{ij}$	the element of the co-occurrence matrix
$X(t)$	the input of LSTM at time $t$

## References

- Chen, X.; Xiong, X.; Qi, X.; Zheng, C.; Zhong, J. A big data simplification method for evaluation of relay protection operation state. *Proc. CSEE* **2015**, *35*, 538–548.
- Chen, G.; Wang, D.; Qiu, Y.; Wang, S.; Qi, X. Challenges and development prospects of relay protection technology. *Autom. Electr. Power Syst.* **2017**, *41*, 1–11.
- Li, G.; Zhang, B.; Zhao, W.; Liu, Y.; Gao, S. Data science issues in state evaluation of power equipment: Challenges and prospects. *Autom. Electr. Power Syst.* **2018**, *42*, 10–20.
- Miao, X.; Zhang, D.; Sun, D. The opportunity and challenge of big data's application in power distribution networks. *Power Syst. Technol.* **2015**, *39*, 3122–3127.
- Zhang, X.; Cheng, X.; Zhao, D. Rule extraction of network operation ticket for power system based on the rough sets. *Power Syst. Technol.* **2014**, *38*, 1600–1605.
- Liu, Y.; Xu, Z.; Li, G.; Xia, Y.; Gao, S. Review on applications of artificial intelligence driven data analysis technology in condition based maintenance of power transformers. *High Volt. Eng.* **2019**, *45*, 337–348.
- Qiu, J.; Wang, H.; Ying, G.; Zhang, B.; Zou, G.; He, B. Text mining technique and application of lifecycle condition assessment for circuit breaker. *Autom. Electr. Power Syst.* **2016**, *40*, 107–112.
- Du, X.; Qin, J.; Guo, S.; Yan, D. Text mining of typical defects in power equipment. *High Volt. Eng.* **2018**, *44*, 1078–1084.
- Hu, J.; Yin, L.; Li, Z.; Guo, L.; Duan, L.; Zhang, Y. Fault diagnosis method of transmission and transformation equipment based on big data mining technology. *High Volt. Eng.* **2017**, *43*, 3690–3697.
- Wang, C.; Jiang, Q.; Tang, Y.; Zhu, B.; Xiang, Z.; Tang, J. Fault diagnosis of power dispatching based on alarm signal text mining. *Electr. Power Autom. Equip.* **2019**, *39*, 126–132.

11. Rudin, C.; Waltz, D.; Anderson, R.N.; Boulanger, A.; Salleb-Aouissi, A.; Chow, M.; Dutta, H.; Gross, P.N.; Huang, B.; Jerome, S.; et al. Machine learning for the New York city power grid. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 328–345. [[CrossRef](#)]
12. Zhang, D.; Han, X.; Deng, C. Review on the research and practice of deep learning and reinforcement learning in smart grids. *CSEE J. Power Energy Syst.* **2018**, *4*, 362–370. [[CrossRef](#)]
13. Deng, J.; Zhang, W.; Yang, X. Recognition and Classification of Incipient Cable Failures Based on Variational Mode Decomposition and a Convolutional Neural Network. *Energies* **2019**, *12*, 2005. [[CrossRef](#)]
14. Ma, B.; Dong, H.; Ren, W. Status assessment of UHV substation secondary equipment based on fuzzy comprehensive support vector machine method. *Comput. Syst. Appl.* **2014**, *23*, 191–197.
15. Zhang, Y.; Hu, C.; Huang, S.; Feng, S.; Lin, G. Apriori algorithm based data mining and analysis method for secondary device defects. *Autom. Electr. Power Syst.* **2017**, *41*, 147–151.
16. Liu, Z.; Wang, H.; Cao, J.; Qiu, J. A classification model of power equipment defect texts based on convolutional neural network. *Power Syst. Technol.* **2018**, *42*, 644–651.
17. Jiang, Y.; Li, L.; Li, Z.; Su, C.; Wang, G.; Xiao, X. An information mining method of power transformer operation and maintenance texts based on deep semantic learning. *Proc. CSEE* **2019**, *39*, 4162–4172.
18. Cao, J.; Chen, S.; Qiu, J.; Wang, H.; Ying, G.; Zhang, B. Semantic framework-based defect text mining technique and application in power grid. *Power Syst. Technol.* **2017**, *41*, 637–643.
19. Shao, G.; Wang, H.; He, B. Quality assessment and improvement method for power grid equipment defect text. *Power Syst. Technol.* **2019**, *43*, 1472–1479.
20. Zhao, L.; Zeng, G.; Wang, W.; Zhang, Z. Forecasting Oil Price Using Web-based Sentiment Analysis. *Energies* **2019**, *12*, 4291. [[CrossRef](#)]
21. Balli, S.; Karasoy, O. Development of content-based SMS classification application by using Word2Vec-based feature extraction. *IET Softw.* **2019**, *13*, 295–304. [[CrossRef](#)]
22. Zhou, J.; Lu, Y.; Dai, H.; Wang, H.; Xiao, H. Sentiment analysis of Chinese microblog based on stacked bidirectional LSTM. *IEEE Access* **2019**, *7*, 38856–38866. [[CrossRef](#)]
23. Zhang, D.; Xu, H.; Su, Z.; Xu, Y. Chinese comments sentiment classification based on word2vec and SVM perf. *Expert Syst.* **2015**, *42*, 1857–1863. [[CrossRef](#)]
24. Dai, L.; Jiang, K. Chinese text classification based on FastText. *Comput. Mod.* **2018**, *5*, 39–44.
25. Pennington, J.; Socher, R.; Manning, C. Glove: Global vectors for word representation. In Proceedings of the Conference Empirical Methods Natural Language Process, Doha, Qatar, 25–29 October 2014; pp. 1532–1543.
26. Ren, X.; Zhou, Y.; He, J.; Chen, K.; Yang, X.; Sun, J. A convolutional neural network-based Chinese text detection algorithm via text structure modeling. *IEEE Trans. Multimed.* **2017**, *19*, 506–518. [[CrossRef](#)]
27. Wei, D.; Wang, B.; Lin, G.; Liu, D.; Dong, Z.; Liu, H.; Liu, Y. Research on Unstructured Text Data Mining and Fault Classification Based on RNN-LSTM with Malfunction Inspection Report. *Energies* **2017**, *10*, 406. [[CrossRef](#)]
28. Bai, Z.; Sun, G.; Zang, H.; Zhang, M.; Shen, P.; Liu, Y.; Wei, Z. Identification Technology of Grid Monitoring Alarm Event Based on Natural Language Processing and Deep Learning in China. *Energies* **2019**, *12*, 3258. [[CrossRef](#)]
29. Li, W.; Li, D.; Yin, H.; Zhang, L.; Zhu, Z.; Liu, P. Lexicon-Enhanced Attention Network Based on Text Representation for Sentiment Classification. *Appl. Sci.* **2019**, *9*, 3717. [[CrossRef](#)]
30. Xu, G.; Meng, Y.; Qiu, X.; Yu, Z.; Wu, X. Sentiment Analysis of Comment Texts Based on BiLSTM. *IEEE Access* **2019**, *7*, 51522–51532. [[CrossRef](#)]
31. Dias, M.; Boné, J.; Ferreira, J.C.; Ribeiro, R.; Maia, R. Named Entity Recognition for Sensitive Data Discovery in Portuguese. *Appl. Sci.* **2020**, *10*, 2303. [[CrossRef](#)]
32. Sun, Y.; Wang, P.; Zhai, S.; Hou, D.; Wang, S.; Zhou, Y. Ultra short-term probability prediction of wind power based on LSTM network and condition normal distribution. *Wind Energy* **2020**, *23*, 63–76. [[CrossRef](#)]
33. Xu, J.; Wei, H.; Li, L.; Fu, Q.; Guo, J. Video Description Model Based on Temporal-Spatial and Channel Multi-Attention Mechanisms. *Appl. Sci.* **2020**, *10*, 4312. [[CrossRef](#)]
34. Zheng, J.; Cai, F.; Shao, T.; Chen, H. Self-Interaction Attention Mechanism-Based Text Representation for Document Classification. *Appl. Sci.* **2018**, *8*, 613. [[CrossRef](#)]
35. Li, S.; Tang, M.; Guo, Q.; Lei, J.; Zhang, J. Deep neural network with attention model for scene text recognition. *IET Comput. Vis.* **2017**, *11*, 605–612. [[CrossRef](#)]



36. Meng, W.; Wei, Y.; Liu, P.; Zhu, Z.; Yin, H. Aspect based sentiment analysis with feature enhanced attention CNN-BiLSTM. *IEEE Access* **2019**, *7*, 167240–167249. [[CrossRef](#)]
37. Guan, P.; Li, B.; Lv, X.; Zhou, J. Attention enhanced bi-directional LSTM for sentiment analysis. *J. Chin. Inf. Process.* **2019**, *33*, 105–111.
38. Xie, Y.; Wang, Y.; Zhu, P.; You, X. Grid-search-based hybrid TOA/AOA location techniques for NLOS environments. *IEEE Commun. Lett.* **2009**, *13*, 254–256. [[CrossRef](#)]
39. Blanco, M.A.; Marti-Puig, P.; Gibert, K.; Cusidó, J.; Solé-Casals, J. A Text-Mining Approach to Assess the Failure Condition of Wind Turbines Using Maintenance Service History. *Energies* **2019**, *12*, 1982. [[CrossRef](#)]
40. Shi, L.; He, Y.; Li, B.; Cheng, T.; Huang, Y.; Sui, Y. Tilt Angle Monitoring by Using Sparse Residual LSTM Network and Grid Search. *IEEE Sens. J.* **2019**, *19*, 8803–8812. [[CrossRef](#)]
41. Feng, Y.; Qu, B.; Xu, H.; Wang, R.; Zhang, Y. Chinese FastText Short Text Classification Method Integrating TF-IDF and LDA. *J. Appl. Sci.* **2019**, *37*, 378–388.



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).