IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 10 June 2024, accepted 7 August 2024, date of publication 16 August 2024, date of current version 27 September 2024.

*Digital Object Identifier 10.1109/ACCESS.2024.3444852*

APPLIED RESEARCH

# BERT-Based Dual-Channel Power Equipment Defect Text Assessment Model

**ZHENAN ZHOU[1], CHUYAN ZHANG[1], (Member, IEEE), XINYI LIANG[1], HUIFANG LIU[1], MINGGUANG DIAO[1], AND YU DENG[2], (Member, IEEE)**

[1]School of Information Engineering, China University of Geosciences (Beijing), Beijing 100083, China
[2]Tibet Yangbajing High Altitude Electrical Safety and Electromagnetic Environment National Observation and Research Station, China Electric Power Research Institute, Yangbajain, Beijing 100192, China

Corresponding author: Chuyan Zhang (zcy@cugb.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 52377162, and in part by the 2023 Annual Open Funds of Tibet Yangbajing High Altitude Electrical Safety and Electromagnetic Environment National Observation and Research Station under Grant GY80-23-006.

**ABSTRACT** Accumulating a substantial amount of textual data on power equipment defects during maintenance and inspection stages presents a valuable problem of assessing and grading these text-based information. This paper proposes a dual-channel text feature extraction model based on the pre-trained BERT model, applied to the evaluation of power equipment defect levels in textual data. Firstly, a dataset of power equipment defect levels is established, followed by data augmentation and preprocessing. Then, a neural network model is constructed, utilizing the pre-trained BERT model for initial semantic information extraction from the text, further extracting features through two modules, Bi-LSTM and CNN, on top of BERT's output. Finally, the obtained feature vectors are concatenated to generate the output. Comparative experiments with other algorithms demonstrate that the proposed method out-performs others in multiple metrics, achieving an F1 score of 96%. The research findings can serve as a reference for achieving intelligent processing of power textual information.

**INDEX TERMS** Power defect text, natural language processing, deep learning, BERT, dual channel model.

## ACRONYMS USED

| | |
|---|---|
| NLP | -Natural Language Processing. |
| RNN | -Recurrent Neural Network. |
| CNN | -Convolutional Neural Network. |
| LSTM | -Long Short-Term Memory. |
| BERT | -Bidirectional Encoder Representations from Transformers. |
| ALBERT | -A Lite BERT. |
| RoBERTa | -Robustly Optimized BERT Pretraining Approach. |
| NSP | -Next Sentence Prediction. |
| Bi-LSTM | -Bidirectional Long Short-Term Memory. |
| CRF | -Conditional Random Fields. |
| EPAT-BERT | -Electric Power Audit Text BERT. |
| MLM | -Masked Language Model. |
| FC | -Fully Connected. |

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

## I. INTRODUCTION

With the rapid development of China's power grid, the extensive and diverse array of power equipment generates a vast amount of defect texts during daily operations. These texts include maintenance test records, inspection and defect elimination records, defect fault description reports, and event sequence records [1]. These defect texts not only reflect the operational state of the power grid but also serve as crucial references for defect analysis and the maintenance and replacement of equipment within the grid. As the process of grid intelligence accelerates, leveraging NLP techniques for the intelligent processing of power equipment defect texts has emerged as an important approach [2]. However, due to the sensitivity and specialization inherent in power texts, the development of NLP in the field of power text analysis remains relatively slow.

Natural language texts are characterized by their ambiguity and complexity, making it challenging for computers to accurately interpret their semantics. Nevertheless, the advancement of deep learning technology provides a potential solution through the collection of natural language texts to train deep neural networks, which can then be used to extract textual features. In 2011, Socher et al. first applied RNN to NLP tasks, demonstrating the feasibility and effectiveness of deep learning models in processing natural language [3]. Subsequently, Kim Y. drew from the success of CNN in the computer vision domain and proposed the Text-CNN, which treats textual data as one-dimensional time series signals and utilizes convolutional operations to extract textual features. This approach proved that CNNs could effectively handle textual data and achieve good performance [4]. LSTM, a special RNN structure, has also been shown to perform well in tasks such as text generation and sequence prediction [5], [6], [7]. In 2017, the Google team introduced the Transformer model architecture, which relies entirely on attention mechanisms for NLP tasks, addressing issues of long-term dependency and parallel processing in sequential data [8]. Despite achieving considerable success in NLP tasks, these models are limited by their network structures and parameter counts, which hinder their ability to extract deep semantic information from long texts.

In 2018, researchers proposed the BERT model based on the Transformer architecture. BERT pre-trains the model using large amounts of unlabeled text in an unsupervised manner and then fine-tunes it for downstream tasks. Compared to traditional NLP models, BERT achieved breakthroughs in both accuracy and practicality [9]. Subsequent research has focused on improving pre-trained models, such as ALBERT, which optimizes parameter sizes and introduces order prediction as a pre-training task to accelerate model training [10]. RoBERTa, which uses a larger dataset and removes the NSP task during pre-training, while employing dynamic masking to enhance model performance [11]. XLNet which integrates the autoregressive language model with permutation language modeling, combining BERT's advantages and incorporating Transformer-XL for handling long sentences, further improving performance over the BERT model [12].

In the field of power text analysis, scholars have enhanced the BERT model by adding Bi-LSTM and CRF modules for power equipment named entity recognition [13]. Improvements have also been seen in a deep analysis model for power equipment defects based on semantic framework text mining techniques, and in the EPAT-BERT model, which uses word-level and entity-level masking languages, significantly outperforming the BERT model in power audit text classification tasks [14], [15]. Additionally, some research teams have employed NLP techniques to resolve ambiguities in free-text power equipment maintenance tickets for better utilization in supervised learning for fault prediction and classification technologies [16].

Given the diversity and irregularity of defect record structures in power equipment defect texts, and the inefficiency and inaccuracy of current mainstream manual classification methods, using NLP techniques for the intelligent evaluation of defect text levels can greatly reduce the workload of maintenance personnel. Based on these considerations, this paper proposes a dual-channel text feature extraction model using a pre-trained BERT model on a Chinese corpus, applied to classify power equipment defect text levels. We first establish a sizeable power text dataset and, during training, use Bi-LSTM and Text-CNN dual channels to parallelly further extract features from BERT model outputs. Finally, features extracted from both channels are integrated to obtain the final classification result. The results demonstrate that the classification accuracy of the dual-channel model surpasses other models, achieving up to 96%, indicating the superior performance of the proposed new model.

## II. TEXT PREPROCESS

### A. DATASET

For deep learning models, the quantity and quality of the training dataset are crucial as they directly affect the model's performance. Due to the confidentiality of power grid operations and maintenance, the defect text datasets of power equipment used in existing research are typically not publicly available. Consequently, for the problem of classifying power equipment defect texts, this paper constructs its own Chinese power text dataset for model training and testing.

The data primarily comes from power equipment maintenance texts over the past five years from a regional power grid in Southwest China. These texts are annotated in accordance with current natural language processing tasks. According to the "Defect Text Guidelines" stipulated by the State Grid Corporation of China, the classification of power equipment defect levels includes three categories: "general," "severe," and "critical." "General" refers to defects that affect the equipment but still allow it to operate safely. "Severe" refers to defects that significantly impact safe and economic operations, allowing only short-term operation and potentially developing into emergency defects or causing accidents if not promptly addressed, "critical" refers to defects posing serious threats to personnel or equipment, likely causing accidents if not promptly addressed. The collected power equipment defect texts are classified into three levels: "general," "severe," and "critical," with sample counts of 715, 1123, and 1193, respectively. Table 1 shows examples of these samples

To enhance the diversity of the dataset and improve the model's generalization capabilities, data augmentation are employed to expand the current dataset. Three types of data augmentation methods are selected: text concatenation, random deletion, and synonym replacement. After applying data augmentation, the expanded dataset contains 10,607 samples, with 2,502, 3,930, and 4,175 samples for the "general," "severe," and "critical", respectively.

### B. PREPROCESSING

Unlike English sentences, where words are clearly separated by spaces, Chinese sentences have ambiguous boundaries

**TABLE 1. Dataset sample.**

| Defect Level | Example Text |
|---|---|
| General | The insulation resistance of the arrester base is low, but it does not affect the normal monitoring of leakage current. |
| Severe | The sealing of the operating mechanism box is not tight, failing to effectively prevent moisture and dust, and no measures have been taken to prevent small animals from entering. |
| Critical | Severe oil leakage of more than 5 drops per minute, or the oil level is below the specified operational value of the transformer. |

between words and phrases. To facilitate computer understanding of the text, the first step in Chinese information processing is typically word segmentation, which involves adding boundary markers between words in Chinese sentences. Common Chinese word segmentation methods include Jieba, HanLP, and FoolNLTK. Among these, Jieba is an excellent third-party Chinese word segmentation tool based on Python. The word segmentation process for Chinese data samples generally involves three main steps: first, using regular expressions to roughly segment Chinese paragraphs into different sentences. Then, constructing a directed acyclic graph for each sentence to preliminarily segment the sentences. Finally, applying the hidden markov mode to further divide continuous single characters.

After word segmentation, sentences need to be converted into vector form. This allows the model to quantitatively measure the relationships between words after training, facilitating the discovery of connections between words. In this paper, Word2Vec is used to vectorize all the words based on the segmentation results. A portion of the Word2Vec results is projected onto a two-dimensional plane to observe the positions of words in the vector space. The results are shown in Figure 1.

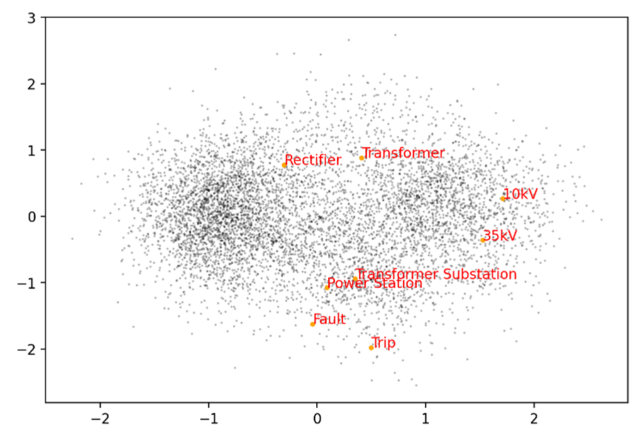


**FIGURE 1. Word vector projection.**

From the figure, it is evident that words with strong correlations are also relatively close to each other in the vector space after Word2vec vectorization. For example, "Power Station" and "Substation," as well as "10kV" and "35kV," are positioned near each other.

## III. DUAL-CHANNEL TEXT EVALUTION MODEL

### A. BERT

The BERT model, introduced by Google in 2018, achieved remarkable performance across various NLP tasks by using the Transformer architecture combined with large-scale unsupervised pre-training tasks. The core idea of BERT is to leverage the self-attention mechanism of the Transformer, enabling the model to effectively capture long-distance dependencies and internal sequence relationships for a more comprehensive understanding of semantic information.

BERT's model structure consists of multiple layers of bidirectional Transformer encoders, each composed of multi-head self-attention and feedforward neural networks. During the pre-training phase, BERT employs two tasks for unsupervised training. The first is the MLM task, where the model randomly masks parts of the input sequence and then predicts the masked tokens. The second is the NSP task, where the model determines whether two sentences are adjacent in meaning. These pre-training tasks enable BERT to learn general language representations.

Figure 2 shows an example of the sentence "未对新增负荷进行有效管控 (Failure to effectively control new loads)" as input to the Bert model. Before feeding the word vectors of text samples into the BERT module, the vectors undergo token embeddings, position embeddings, and segment embeddings. Token embeddings add two tokens, [CLS] and [SEP], to the word vectors, representing the beginning and end of the text, respectively. Position embeddings capture the positional information of words within a sentence since the meaning of a word can vary significantly depending on its position. Segment embeddings distinguish between two sentences in a sentence pair, used to learn whether two texts are semantically similar. However, since model proposed in this study training does not involve sentence pairs, the segment embeddings are all encoded as $E_A$.

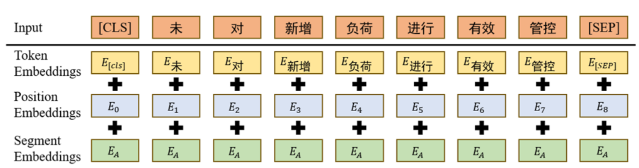


**FIGURE 2. Input of BERT.**

One of the features of the Transformer is that the number of inputs is equal to the number of outputs. Therefore, as shown in Figure 3, BERT's output consists of $C$, which corresponds to the [CLS] token. Since [CLS] is placed at the beginning of the input sample, its corresponding output $C$ is considered to represent the overall characteristics of the sentence. $T_i$ represents the outputs corresponding to the other tokens, which capture the semantic features of the words within the sentence.


**FIGURE 3.** Output of BERT.

For classification tasks, BERT employs $C$ as the sole input to the classification head. However, in the domain of power equipment defect level text categorization, the critical information for accurate classification is predominantly encapsulated within specific terms such as the names of the equipment and the types of defects. Given this context, it becomes essential to conduct further feature extraction from BERT's outputs to enhance the representation of these key terms. This additional step is crucial for improving the performance and robustness of the model in accurately categorizing power equipment defect levels.

In the subsequent dual-channel feature extraction module, $C$ is used as the input for the Bi-LSTM module, while $T_i$ serves as the input for the CNN module, further utilizing the information extracted by the BERT module.

### B. Bi-LSTM

The Bi-LSTM network [17] represents a significant advancement in the field of RNNs. Traditional RNNs face the problem of vanishing or exploding gradients when handling long sequences, making it difficult to capture long-term dependencies. LSTM addresses this issue by introducing gating units, such as forget gates and input gates, effectively enabling the network to capture and retain long-term sequence information. Bi-LSTM improves upon the LSTM structure by integrating bidirectional information flows, enhancing the model's ability to capture long-range dependencies.

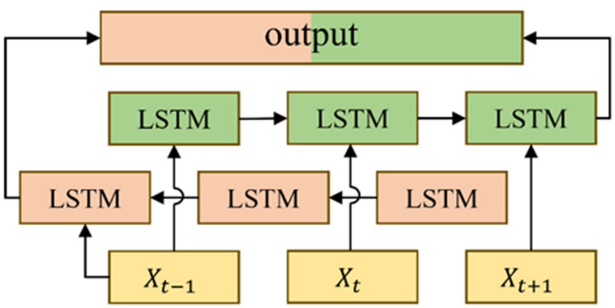


**FIGURE 4.** Bi-LSTM module.

The structure of the Bi-LSTM network is shown in Figure 4. It consists of two different LSTM layers: one processes information in the forward sequence, while the other processes information in the backward sequence. The outputs of these two layers are then concatenated. This allows the network to more comprehensively understand the context of the input data, enhancing its ability to model complex relationships within the sequence. In the proposed model, the Bi-LSTM module processes the output $C$ from the BERT module, and its output features are concatenated into the model's final FC layer module.

### C. CNN

CNNs were initially designed for image processing but can also be applied to sequence data. CNNs use convolutional kernels (filters) that slide over the sequence to perform convolution operations, aiming to detect local features. In text processing, convolutional kernels typically detect a series of adjacent words or phrases, similar to how kernels detect edges or textures in images. Figure 5 shows an example of CNN processing sequence data. After concatenating the input sequence, a $3 \times 3$ convolutional kernel performs convolution operations on the feature map to obtain the corresponding output feature map.

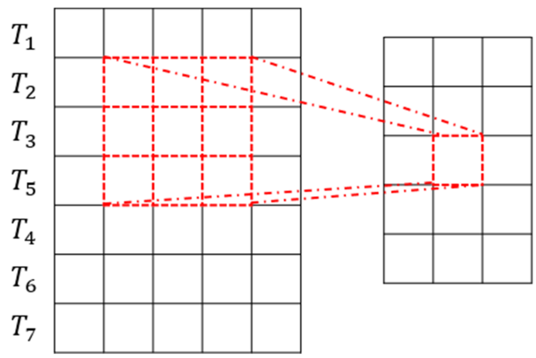


**FIGURE 5.** Convolutional operation of sequence data.

The BERT model has a substantial number of features in its output $T_1 \sim T_N$. Since convolution operations use the same convolution kernels to perform identical operations at different positions of the input, sharing parameters, it reduces the number of parameters required for training. Therefore, in the proposed model, the CNN module is used to process the output $T_1 \sim T_N$ from the BERT model. It extracts local features from the sequence through convolutional layers and utilizes pooling layers to reduce data dimensions while retaining critical information. Finally, the resulting output is flattened and concatenated into the model's final FC layer module.

### D. MODEL STRUCTURE

The dual-channel power equipment defect level text evaluation model based on BERT proposed in this paper is illustrated in Figure 6. The model structure consists of four modules: BERT, Bi-LSTM, CNN, and FC.

The BERT module is responsible for the initial feature extraction of word vectors. The first feature vector obtained is input into the Bi-LSTM module, while the remaining feature vectors are input into the CNN module. After further feature extraction by the Bi-LSTM and CNN modules, the output vectors are concatenated in the FC layer. The final output is obtained by dimension reduction through the FC layer.

Suppose the input sequence of power equipment defect levels is denoted as $X$. The operational process of the model

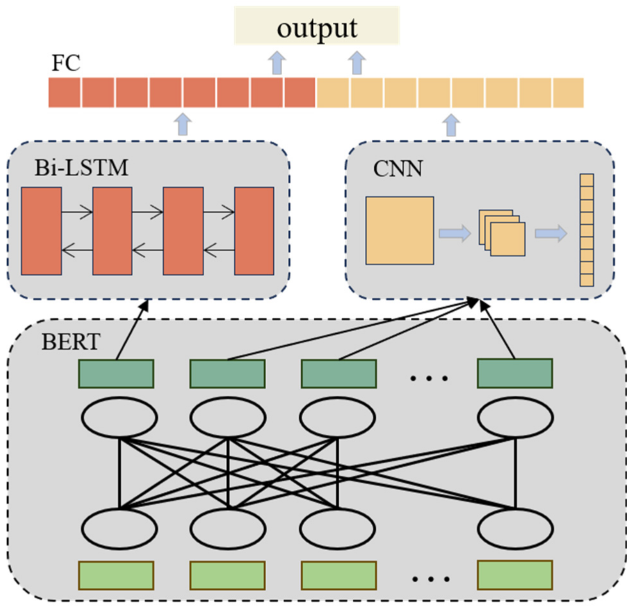


**FIGURE 6.** Model structure.

is as follows:

$$I = E(X) \tag{1}$$

$$[C, T_1, T_2, \cdots, T_n] = BERT(I) \tag{2}$$

$$[F_1, F_2, \cdots, F_m] = Bi\_LSTM(C) \tag{3}$$

$$[F_{m+1}, F_{m+2}, \cdots, F_{2m}] = CNN\ ([T_1, T_2, \cdots, T_n]) \tag{4}$$

$$O = FC\ ([F_1, F_2, \cdots, F_m, F_{m+1}, \cdots, F_{2m}]) \tag{5}$$

Equation (1) represents the preprocessing stage of the model, which involves converting the text into word vectors $X$ using Word2Vec and the embedding processing before inputting into BERT. The final input vector obtained is denoted as $I$. Equation (2) signifies the inputting of $I$ into the BERT module, resulting in the output feature vector $[C, T_1, T_2, \ldots, T_n]$. Subsequently, $C$ serves as the input for the Bi-LSTM module, while $[T_1, T_2, \ldots, T_n]$ serves as the input for the CNN module, as depicted in equations (3) and (4) respectively. The output feature vectors are concatenated and fed into the FC layer to obtain the final output $O$ as demonstrated in equation (5).

## IV. MODEL TRAINING AND TESTING

### A. MODEL TRAINING

The model is executed using PyCharm Community 2022.2.2 as the Integrated Development Environment (IDE), and the computer's operating system is Windows 11. All models are implemented using PyTorch 1.11.0 deep learning framework and Python 3.8. The hardware specifications for local execution are as follows: Intel i5 12400F CPU, 64GB RAM, NVIDIA 3090 GPU with 24GB VRAM.

The model in this paper is constructed based on the BERT architecture. The BERT model is initialized using the parameters provided by Hugging Face for the Chinese BERT model. The Bi-LSTM and CNN parts of the model are initialized with random initialization. The training is performed with a batch size of 32, a learning rate of 5 × 10.4, and the Adam optimizer.

The samples were randomly divided into training set, verification set and test set according to the ratio of 7:1.5:1.5 for the subsequent training and testing of the models. The established training dataset is utilized for model training. Additionally, during the training process, the validation dataset is employed for validation every 5 batches. The training process yielded the loss and accuracy of the model, as depicted in Figure 7. From the graph, it can be observed that the model achieves satisfactory accuracy on both the training and validation datasets after approximately 50 epochs. Moreover, the accuracy on the validation dataset does not decrease with increasing epochs, indicating good generalization of the model and the absence of overfitting.

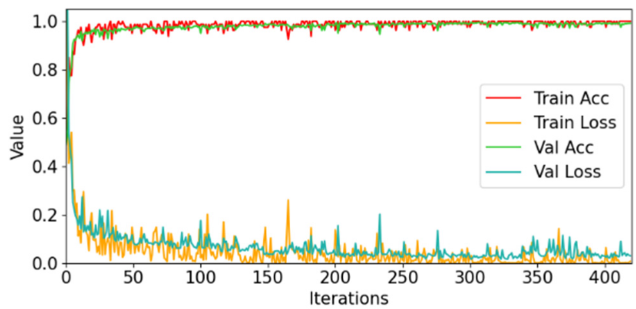


**FIGURE 7.** Training loss and accuracy.

### B. MODEL TESTING

The model's performance is evaluated using precision $P$, recall $R$, and $F_1$ score. The calculation formulas for each metric are shown in Equation (6).

$$\begin{cases} P_i = \dfrac{T}{T + F_P} \\ R_i = \dfrac{T}{T + F_N} \\ F_{1i} = \dfrac{2P_iPR_iR}{P_iP + R_iR} \end{cases} \tag{6}$$

In this context, for a specific type of defect level, $i$, $T$ represents the number of samples for which the prediction is correct, $F_P$ denotes the number of samples predicted as $i$ but with incorrect predictions, and $F_N$ indicates the number of samples for which the prediction is $i$ not and has incorrect predictions. The $F_1$ score, derived from $P$ and $R$, provides a comprehensive evaluation of precision and recall.

For the current model, since the predicted output includes three levels of defect severity, namely "general," "severe," and "critical," the metrics of the model are the average values of the metrics for these three different defect severity levels, as shown in Equation (7).

$$\begin{cases} P = \dfrac{\sum_{i=1}^{3} P_i}{3} \\ R = \dfrac{\sum_{i=1}^{3} R_i}{3} \\ F_1 = \dfrac{2PR}{P+R} \end{cases} \tag{7}$$

The testing dataset established in the previous section is input into the model for testing. To demonstrate the effectiveness of the model architecture proposed in this paper, comparisons and validations are conducted among LSTM, Bi-LSTM, BERT, BERT-LSTM, BERT-Bi-LSTM, and BERT-CNN using the same dataset. The obtained results are presented in Table 2.

**TABLE 2. Dataset sample.**

| Model | P(%) | R(%) | F1(%) | Size(M) |
|---|---|---|---|---|
| Bi-LSTM | 77.8 | 78.6 | 78.2 | 172 |
| Bert | 88.9 | 86.7 | 87.8 | 390 |
| Bert-Bi-LSTM | 94.7 | 92.8 | 93.7 | 417 |
| Bert-CNN | 90.1 | 91.7 | 90.9 | 390 |
| Bert-Bi-LSTM-CNN (The proposed model) | **96.4** | **96.2** | **96.3** | 417 |

From the table, it can be observed that using the pre-trained BERT model yields better feature extraction performance compared to using non-pre-trained LSTM and Bi-LSTM models. Moreover, incorporating additional modules such as LSTM and CNN after the BERT model for further feature extraction leads to improved accuracy, and the model size does not show a significant increase. Ultimately, the proposed dual-channel model outperforms other models across all metrics, demonstrating the feasibility and effectiveness of the dual-channel approach for evaluating the severity of defects in power equipment text.

## V. CONCLUSION

This study addresses the inefficiency and high error rates associated with manual grading of power equipment defect texts. The model proposed in this study is a power equipment defect severity assessment model based on the pre-trained BERT model. Utilizing Bi-LSTM and CNN modules, semantic information is extracted from the BERT output, and the output features from the dual channels are concatenated to produce the model's output. The proposed model achieves precision, recall, and F1-score of 96.4%, 96.2%, and 96.3%, respectively, on the power equipment defect text dataset established in this study, showing improvements in various metrics compared to other algorithms. Future work will focus on expanding the dataset size, changing pre-trained models, and exploring multiple downstream tasks on power equipment defect text datasets to achieve intelligent processing of power text, providing references and support for the intelligent operation and maintenance of power grids.

## REFERENCES


[1] J. Zhan, J. Huang, L. Niu, X. Peng, D. Deng, and S. Cheng, "Key technologies of electric power big data and its application prospects in smart grid," *Proc. CSEE*, vol. 35, no. 3, pp. 503–511, 2015, doi: 10.13334/j.0258-8013.pcsee.2015.03.001.

[2] Z. Kong, C. Yue, Y. Shi, J. Yu, C. Xie, and L. Xie, "Entity extraction of electrical equipment malfunction text by a hybrid natural language processing algorithm," *IEEE Access*, vol. 9, pp. 40216–40226, 2021, doi: 10.1109/ACCESS.2021.3063354.

[3] R. Socher, C. Y. Lin, A. Ng, and C. D. Manning, "Parsing natural scenes and natural language with recursive neural networks," *Int. Conf. Mach. Learn.*, 2011, pp. 129–136.

[4] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1746–1751, doi: 10.3115/v1/d14-1181.

[5] W. Zhang, W. Guo, X. Liu, Y. Liu, J. Zhou, B. Li, Q. Lu, and S. Yang, "LSTM-based analysis of industrial IoT equipment," *IEEE Access*, vol. 6, pp. 23551–23560, 2018, doi: 10.1109/ACCESS.2018.2825538.

[6] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," in *Proc. NIPS Workshop Deep Learn.*, 2014, doi: 10.48550/arXiv.1412.3555.

[7] M. Henaff, J. Weston, A. Szlam, A. Bordes, and Y. LeCun, "Tracking the world state with recurrent entity networks," 2016, *arXiv:1612.03969*.

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017, *arXiv:1706.03762*.

[9] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.

[10] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A lite BERT for self-supervised learning of language representations," 2019, *arXiv:1909.11942*.

[11] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," 2019, *arXiv:1907.11692*.

[12] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized autoregressive pretraining for language understanding," 2019, *arXiv:1906.08237*.

[13] J. Hu, C. Jiang, G. Ma, J. Ding, Y. Wang, J. Xu, and Y. Wang, "Power entity information recognition method based on Bi-LSTM+CRF," in *Proc. Int. Conf. Adv. Electr. Equip. Reliable Operation (AEERO)*, Oct. 2021, pp. 1–5, doi: 10.1109/AEERO52475.2021.9708243.

[14] P. Chen, B. Tai, Y. Shi, Y. Jin, L. Kong, and J. F. Wang, "Text entity extraction of power equipment defects based on BERT-BI-LSTM-CRF algorithm," *Power Syst. Technol.*, vol. 47, no. 10, pp. 4367–4376, 2023.

[15] Q. Meng, Y. Song, J. Mu, Y. Lv, J. Yang, L. Xu, J. Zhao, J. Ma, W. Yao, R. Wang, M. Xiao, and Q. Meng, "Electric power audit text classification with multi-grained pre-trained language model," *IEEE Access*, vol. 11, pp. 13510–13518, 2023, doi: 10.1109/ACCESS.2023.3240162.

[16] B. Stephen, X. Jiang, and S. D. J. McArthur, "Extracting distribution network fault semantic labels from free text incident tickets," *IEEE Trans. Power Del.*, vol. 35, no. 3, pp. 1610–1613, Jun. 2020, doi: 10.1109/TPWRD.2019.2947784.

[17] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," 2015, *arXiv:1508.01991*.


**ZHENAN ZHOU** received the B.S. degree in electrical engineering and automation from China University of Geosciences (Beijing), China, in 2020, where he is currently pursuing the Ph.D. degree. His major research interest includes image processing of power equipment.

 

**CHUYAN ZHANG** (Member, IEEE) was born in Qinghai, China, in June 1986. He received the B.S. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China, in 2008 and 2013, respectively. From 2013 to 2014, he was an Engineer with State Grid Corporation of China. From 2014 to 2016, he was a Research Assistant with the Graduate School at Shenzhen, Tsinghua University, Shenzhen, China. Since July 2016, he has been an Associate Professor with the School of Information Engineering, China University of Geosciences (Beijing), China. He is currently the author of more than 40 articles. His research interests include high-voltage external insulation, UHV insulation technology, power equipment condition sensing, and insulating material.

**XINYI LIANG** is currently pursuing the master's degree with China University of Geosciences (Beijing). Major in control engineering. Her main research interest includes the combination of deep learning and power field.

**HUIFANG LIU** was born in Henan, China, in November 1975. She received the Ph.D. degree in geological engineering from China University of Geosciences (Beijing), China, in 2011. She is currently an Associate Professor with China University of Geosciences (Beijing). Her research interests include renewable energy grid connection and HVDC power transition technology.

**MINGGUANG DIAO** was born in Inner Mongolia, China, in January 1970. He received the Ph.D. degree in information technology from China University of Geosciences (Beijing), China, in 2010. He is currently an Associate Professor with China University of Geosciences (Beijing). He is mainly engaged in the teaching of software engineering theory and methods, and related courses in software engineering technology. In the interdisciplinary field of new-generation information technology and geoinformation engineering, with a focus on the direction of large-scale software engineering methods, conducts research in the development of large complex software in the intersection of software engineering and geology, and in the area of formal methods in software engineering. He has presided more than 25 scientific research projects. He presided more than 103 research outcomes, including academic articles, invention patents, and software copyrights.

**YU DENG** (Member, IEEE) was born in Jiangxi, China, in April 1987. He received the Ph.D. degree in electrical engineering from Tsinghua University, Beijing, China, in 2014. He is currently the Director of the High Voltage Insulation Materials Research Department, High Voltage Research Institute, China Electric Power Research Institute. He is a member of IEC TC36 MT19 and Tibet Yangbajing High Altitude Electrical Safety and Electromagnetic Environment National Observation and Research Station. He has participated in three national key research and development projects and presided more than six scientific and technological projects of State Grid Company. His research interests include high voltage and insulation technology, electrical materials, and dielectrics. He has been awarded one prize of Electric Power Science and Technology Innovation Award and two prizes of China Electric Power Science and Technology Progress Award.