

Defect text recognition and condition assessment method of power equipment based on BERT-BiLSTM

Jian Shao¹, Xuan Li¹, Tongbao Liu², Hongzhong Chen², Tonglei Wang¹ and Chao Wei²

¹ State Grid Jiangsu Electric Power Research Institute, Nanjing, China

² State Grid Jiangsu Electric Power Co., Ltd., Nanjing, China

ABSTRACT

The accumulated text of power equipment defects in operation are crucial for condition evaluation and maintenance. However, different structures of defect texts and their sensitivity to the experience of operation and maintenance personnel have led to difficulties in information mining, resulting in low efficiency of defect management. To address this issue, a defect text recognition method for power equipment is proposed based on Bidirectional Encoder Representations from Transformers-Bidirectional Long Short-Term Memory Neural Network (BERT-BiLSTM). First, the pre-training model of BERT is used to provide word vectors, which are fine-tuned by the BERT model based on the context of the defect text. Combined with the global semantic information, the vectors are inputted into the BiLSTM model. Then BiLSTM is used to encode and fuse features of the sequences to extract deep semantic features of defects. The obtained categories include defect severity, component, and type identification results. Based on algorithm comparison and case verification, the accuracy of the method is greater than 94% in identifying defect severity and component, proving the accurate recognition of defect text of power equipment. Finally, a defect condition assessment method is proposed to improve health management and condition maintenance.

Subjects Artificial Intelligence, Data Mining and Machine Learning, Natural Language and Speech, Optimization Theory and Computation, Text Mining

Keywords Transformer defects, Text categorization, BERT-BiLSTM, Defect text

INTRODUCTION

The intelligence of smart grid is developing increasingly. Digital empowerment integrates artificial intelligence and information technology with power systems, enhancing digital capabilities and transforming traditional operation and maintenance of equipment (*Pu et al., 2020; Jina et al., 2024*). Over long-term operation, transformers accumulate large amounts of structured and unstructured data, including text, audio, and images (*Liu et al., 2022; Meng et al., 2023; Lu et al., 2023*), which are crucial for status assessment and fault diagnosis (*Liao et al., 2023; Li et al., 2021*). Compared with audio and image data, manually recorded defect texts provide more precise defect information, but standardization is limited by personnel expertise, reducing text data utilization (*Ge et al., 2021; Liu et al., 2023*). Equipment health warning and fault diagnosis based on defect text mainly rely on manual processing, with decisions based on expert experience, leading to slow response

Submitted 20 May 2025
Accepted 1 October 2025
Published 19 December 2025

Corresponding author
Xuan Li, xuanya183174@163.com
Academic editor
Siddhartha Bhattacharyya
Additional Information and Declarations can be found on page 17

DOI 10.7717/peerj-cs.3318

© Copyright
2025 Shao et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

times (Liu et al., 2019). Leveraging massive defect text for transformer condition assessment and maintenance is thus critical for improving operational efficiency (Jia et al., 2023).

Several studies have explored transformer defect text mining in power systems (Wei et al., 2017; Liu et al., 2018; Jiang et al., 2019; Feng et al., 2020; Wang et al., 2024). Wei et al. (2017) proposed an Recurrent Neural Network-Long Short-Term Memory (RNN-LSTM) method for fault classification, improving the handling of unstructured data. Liu et al. (2018) applied Convolutional Neural Network (CNN) for transformer defect text classification. Jiang et al. (2019) used an RCNN-based approach to evaluate equipment health from defect levels. Feng et al. (2020) introduced a Bidirectional Long Short-Term Memory Neural Network (BiLSTM) with attention mechanism, enhancing classification accuracy and proposing a defect priority index. Wang et al. (2024) developed a knowledge-integrated manifold method to address fuzziness and diversity in defect descriptions. These studies primarily focus on defect text classification. Combining text recognition algorithms with multi-dimensional defect features could provide richer information for transformer condition assessment and maintenance, improving operational efficiency of transformer management systems.

To address limited defect text of power equipment mining and equipment maintenance capabilities, a Bidirectional Encoder Representations from Transformers (BERT)-BiLSTM-based transformer defect text recognition method is proposed. The method uses a BERT model to capture global semantic information of defect texts and BiLSTM to model contextual dependencies, enhancing feature extraction and classification. Through case studies, the effectiveness of the proposed BERT-BiLSTM text recognition model in extracting text information is confirmed by using the evaluation metrics such as classification accuracy, confusion matrix, precision, recall, and F1-score. Then it is compared with several commonly deep learning classification algorithms.

In addition, based on the classification model outputs, defect severity, component, and type, text recognition is integrated with transformer condition evaluation to derive a transformer health index. This indicator reflects the cumulative impact of historical defect records on equipment, and is used to determine the status level and guide the development of corresponding maintenance strategies. Unlike previous studies focusing solely on classification accuracy, this method provides a closed-loop framework converting unstructured defect descriptions into actionable insights for intelligent maintenance planning.

TEXT ANALYSIS OF POWER EQUIPMENT DEFECT

Overview of power equipment defect texts

Defect records of power equipment are mainly generated by operation and maintenance personnel, who enter defect information from daily inspections, maintenance, and equipment operation into the Power Management System (PMS) following a certain template (Wang, Cao & Lin, 2019). These records include objective and subjective details such as equipment type, defect location, date, description, severity, causes, and clearance time, each corresponding to a specific defect event of a transformer (Ye, Guo & Qi, 2023; Bhattarai et al., 2019).

Table 1 Example of a partially defect text record.

Number	Defect text	Defect severity
1	Discolouration of breathing apparatus silicone on No.1 main transformer over 2/3	General defect
2	The main transformer tank leakage, oil leakage rate per drop time is not faster than 5 s	General defect
3	The main transformer on-load tap-changer actuator switching off when gearing up	Serious defect
4	The on-load tap-changer silicone is completely discolored or the silicone is discolored from top to bottom	Serious defect
5	Poor grounding at the end screen causing discharge	Critical defect
6	The main transformer breathing apparatus is blocked.	Critical defect

Because defect texts are completed by personnel with varying experience, their formats are diverse, descriptions may be vague or oversimplified, and occasional errors exist in severity or component labels. Nevertheless, defect descriptions often contain critical information on defect occurrence, location, and phenomena, serving as an important supplement to other record fields—especially for equipment defects with unclear boundaries.

In practical grid operations, defect severity, typically classified as general, serious, or critical, is manually assigned, making the process labor-intensive, subjective, and inconsistent. However, severity classification is crucial for fault management, directly influencing required response times and maintenance prioritization. With advancements in natural language processing, deep learning techniques applied to historical defect text data enable automated extraction of key semantic features, achieving rapid and accurate classification. This approach reduces manual workload and provides a robust foundation for maintenance planning, fault diagnosis, and transformer health assessment ([Zhang et al., 2021](#); [Wang et al., 2019](#); [Liao et al., 2021](#); [Dong et al., 2019](#); [Tamma, Prasojo & Suwarno, 2021](#); [Zhang et al., 2023](#)).

Analysis of the characteristics of defect texts of power equipment

Partial defect text records are shown in [Table 1](#). Compared with general Chinese text, transformer defect records are highly variable due to manual entry and contain strong subjectivity. Their main characteristics are as follows:

- (1) Defect text belongs to a specialized domain of the power system and includes numerous technical terms. Influenced by personnel experience, professional level, and language habits, the same equipment may be described differently, such as “main transformer” and “main power transformer”, “oil storage tank” and “oil pillow”.
- (2) Significant variation in length, from a few characters to over 100, caused by differences in inspection methods, scope, and language style.
- (3) Frequent inclusion of letters, numbers, abbreviations, and formulas. Quantitative details, such as “discoloration exceeds 2/3 of the total amount”, often determine defect severity, adding to mining difficulty.
- (4) Semantic loss or redundancy from limited operator experience, such as missing severity levels, vague locations, or irrelevant details, which introduce noise and reduce mining accuracy.

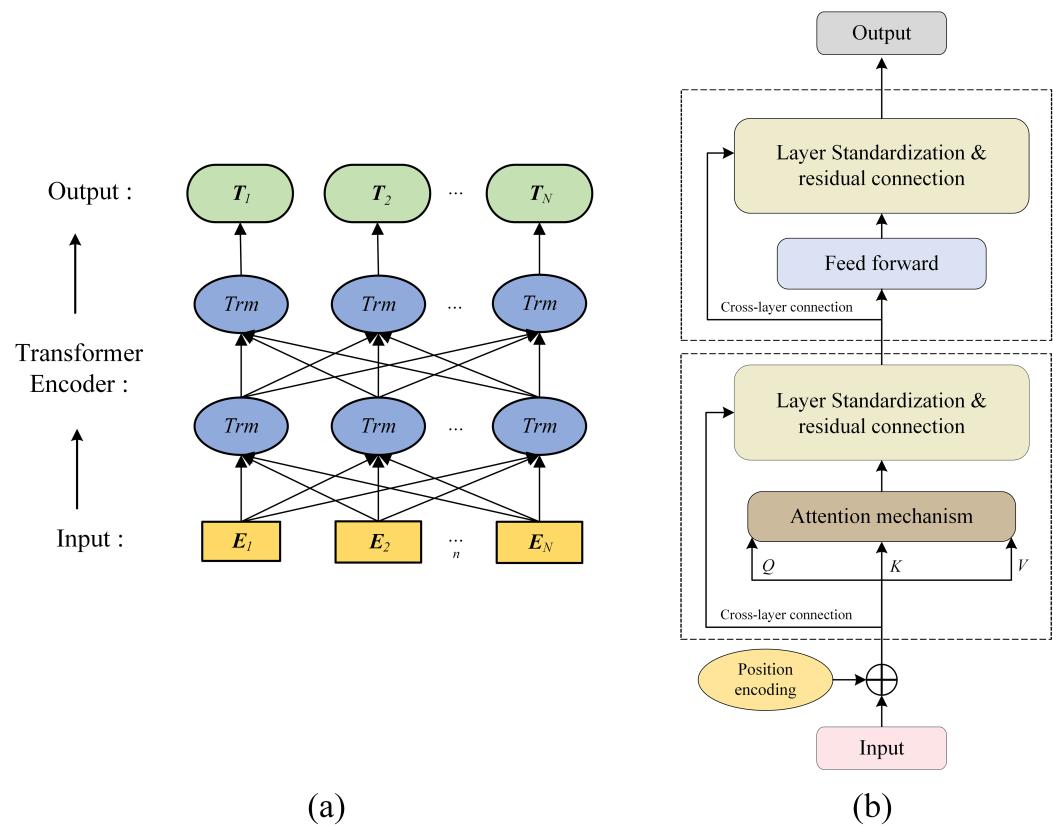


Figure 1 Structure of (A) BERT model and (B) Transformer encoding structure diagram.

Full-size DOI: [10.7717/peerj-cs.3318/fig-1](https://doi.org/10.7717/peerj-cs.3318/fig-1)

CONSTRUCTION OF BERT-BILSTM DEFECT TEXT RECOGNITION MODEL

Theory of BERT model

The BERT model, proposed by Google in 2018, employs a bidirectional Transformer encoder and unsupervised masked language modeling on large corpora to enhance natural language understanding and bidirectional representations (Devlin *et al.*, 2018; Vaswani *et al.*, 2017; Yang & Jiang, 2018). Unlike RNNs and CNNs, BERT's attention mechanism reduces token distance to one, effectively handling long-range dependencies. For transformer defect texts, BERT accurately extracts key information such as defect type, component, and severity. Its structure is shown in Fig. 1.

In Fig. 1A, the BERT feature extractor consists of four layers. The first layer (E_1-E_n) is the embedding layer, combining token and positional embeddings to preserve semantics and order. The second and third layers are stacked Transformer encoder blocks ("Trm") using self-attention to capture multi-level contextual dependencies for deep semantic modeling of defect texts, their structure is shown in Fig. 1B. The final layer (T_1-T_n) outputs contextualized token vectors for downstream BiLSTM sequence modeling.

In Fig. 1B, the Transformer core includes a self-attention mechanism and a Feedforward Neural Network (FNN), used to extract defect text vector features. BERT's self-attention

models relationships between any two tokens, capturing semantics and internal structure. The self-attention output is computed as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

where Q , K , and V are query, key, and value matrices, and d_k is their dimension.

Compared with a single self-attention mechanism, multi-head attention extends this by projecting into n subspaces with distinct W_i^Q , W_i^K and W_i^V , enabling the model to focus on multiple aspects simultaneously:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(h_1, \dots, h_i, \dots, h_n)W^o \quad (2)$$

$$h_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (3)$$

where *Concat* represents the concatenation of all h_i matrices obtained from the i -th self-attention mechanism, followed by matrix multiplication with the concatenation matrix; n is the heads in the multi-head self-attention mechanism, typically set to 8; W_i^Q , W_i^K and W_i^V are the weight matrices of matrices Q , K , V in the i -th subspace.

Residual connections and layer normalization are applied after self-attention to preserve gradients, stabilize training, and improve convergence (Chen *et al.*, 2020). The normalized output passes through an FNN with two fully connected layers, followed by another residual and normalization step, and this process is repeated. The output of the FNN model is calculated as follows:

$$x_{\text{FNN}} = \text{ReLU}(0, x \cdot w_1 + b_1) \cdot w_2 + b_2 \quad (4)$$

where x is the input to the feedforward neural network; w_1 , b_1 and w_2 , b_2 are the weights and biases of the two fully connected layers, respectively.

Theory of BiLSTM model

LSTM neural network

The LSTM neural network, proposed by Hochreiter and Schmidhuber (Jang *et al.*, 2020; Zhou *et al.*, 2019), enhances long-term dependency modeling by introducing three gates (input i , forget f , and output o) to control information flow within a memory cell. LSTM adapts to variable-length sequences and maintains contextual consistency, making it suitable for tasks such as Natural Language Processing (NLP), speech recognition, and time-series forecasting. Its internal structure, shown in Fig. 2A, comprises the three gates and a memory cell.

The forget gate determines how much of the previous cell state is retained or discarded. As shown in Eq. (5), it processes the previous hidden state \mathbf{h}_{t-1} and current input \mathbf{x}_t , outputting a value between 0 (complete discard) and 1 (full retention):

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \quad (5)$$

where f_t is the forget gate state; w_f and b_f are its weights and bias, and σ is the sigmoid activation function.

The input gate regulates new information entering the cell state. It evaluates \mathbf{h}_{t-1} and \mathbf{x}_t , to decide which information to store, with a tanh layer generating candidate memory

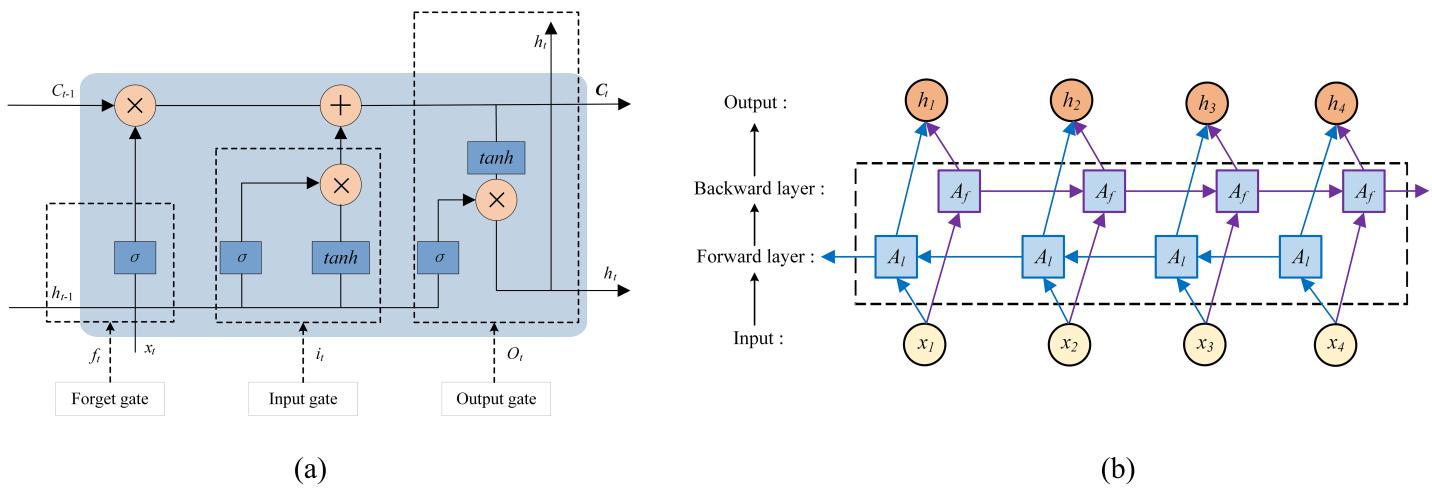


Figure 2 Structure of the (A) LSTM and (B) BiLSTM model.

Full-size DOI: 10.7717/peerj-cs.3318/fig-2

values. The updated cell state C_t combines retained and new information, as expressed in Eq. (6).

$$i_t = \sigma(w_i h_{t-1} + w_i x_t + b_i) \quad (6)$$

$$c_t = \tanh(w_c h_{t-1} + w_c x_t + b_c) \quad (7)$$

$$C_t = C_{t-1} \otimes f_t + i_t \otimes c_t \quad (8)$$

where i_t is the input gate state at time step t , controlling the amount of information from x_t passed to C_t ; w_i and b_i are the weights and biases of the input gate respectively; w_c and b_c are the weight matrix and bias term of the cell state; \tanh is the hyperbolic tangent activation function, and \otimes denotes the Hadamard product.

The output gate determines which part of the cell state is exposed as the hidden state. A sigmoid layer selects relevant cell state elements, which are then scaled by \tanh to produce the output h_t , also serving as the next hidden state input. The output gate is calculated as follows:

$$o_t = \sigma(w_o h_{t-1} + w_o x_t + b_o) \quad (9)$$

$$h_t = o_t \otimes \tanh(C_t) \quad (10)$$

where o_t is the output gate state at time step t ; h_{t-1} is the previous time step's hidden state; w_o and b_o are the weight matrix and bias of the output gate, respectively.

BiLSTM neural network

Although LSTM captures long-term dependencies in sequences, it operates in a single temporal direction. BiLSTM enhances this by integrating forward and backward LSTM units, leveraging both past and future context to model long-range dependencies and richer semantic relationships (Zhou et al., 2019). As shown in Fig. 2B, forward units A_f and backward units A_b process the input x_i , BERT-generated word vectors in this study, and their concatenated outputs h_i form the extracted semantic features.

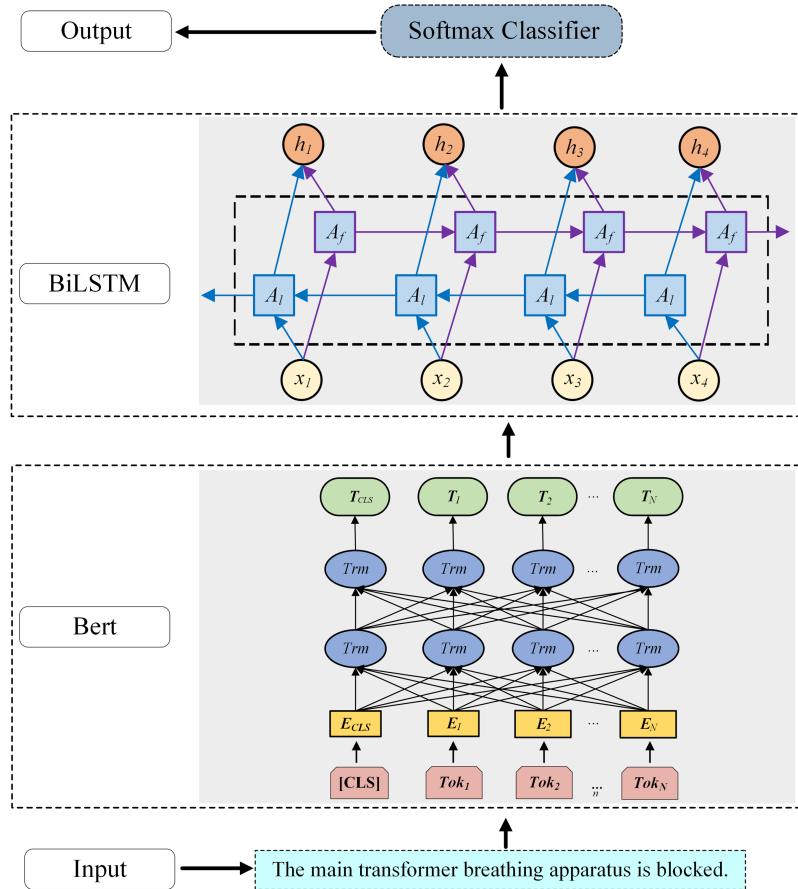


Figure 3 Structure diagram of the defect text recognition model based on BERT-BiLSTM.

Full-size DOI: [10.7717/peerj-cs.3318/fig-3](https://doi.org/10.7717/peerj-cs.3318/fig-3)

BERT generates rich contextual embeddings *via* transformer-based attention, yet transformer defect texts are often short, irregular, and telegraphic, which may hinder its ability to capture sequential dependencies or resolve syntactic ambiguities. To mitigate this, token embeddings $\{x_1, x_2, \dots, x_n\}$ from BERT's last hidden layer are fed into a BiLSTM, producing forward \vec{h}_i and backward \bar{h}_i hidden states that are concatenated into h_i for each token. This hybrid structure combines BERT's global semantic modeling with BiLSTM's local sequential encoding, enhancing robustness in handling ambiguous or irregular defect texts.

Defect text recognition model

To effectively extract important features from defect texts and accurately identify defect characteristics such as performance, component, and severity in the defect texts, a defect text classification model based on BERT-BiLSTM is established, as shown in Fig. 3.

First, the word embedding layer converts the input sequence into word vectors using the pre-trained BERT model. In Fig. 3, Tok_1 - Tok_N correspond to the characters or words in the defect text. The $[CLS]$ token added at the beginning of the text represents the vector T_{CLS} obtained after training the BERT model, which can be used for subsequent text

Table 2 Description of defect classification categories.

Category label	Label types	Label categories	Meanings
Defect performance	4	Mechanical, Insulation, Thermal, Others	Types of stresses that cause defects to occur.
Defect component	9	Iron core, Winding, Oil insulation, Tank, Conservator, On-load tap-changer, Bushing, Cooling system, Non electricity protection.	The specific component or locations where defects occur.
Defect severity	3	General, Serious, Critical	The extent to which defects affect the operation of the power grid.

classification. Contextual word vectors are then generated by BERT as input to the BiLSTM. Here, the Chinese BERT-Base model is used, comprising 12 transformer layers with hidden size 768, pre-trained on large-scale Chinese corpora and fine-tuned on the cleaned dataset to capture defect-domain linguistic patterns. The BiLSTM encodes forward and backward context, capturing long-range dependencies in the BERT output sequence. Finally, the decoding layer maps BiLSTM outputs to classification space *via* fully connected layers, and the Softmax classifier predicts defect labels.

CLASSIFICATION RESULT ANALYSIS OF TEXT RECOGNITION MODEL FOR TRANSFORMER DEFECT

Transformer defects text dataset setup

Based on 10,550 Alternating Current (AC) transformer defect records above 35 kV from a provincial power grid over 5 years, a defect text dataset was constructed through manual cleaning and screening. Original records, written by personnel with varying expertise, included inconsistent terminology (e.g., “main transformer” *vs.* “power transformer”), vague expressions (e.g., “abnormal noise” without location), and inaccurate component labels. Cleaning involved standardizing terms (e.g., unifying “cooling fan” variants), resolving ambiguities (e.g., cross-referencing components), and verifying labels (e.g., correcting severity ratings from context). After removing duplicates, errors, and invalid entries, and manually correcting missing information, 2,464 text records remained, split into training and testing sets at a 3:1 ratio. Classification labels, defect performance, component, and severity, were assigned based on the dataset and standards. The label categories and their meanings are listed in [Table 2](#).

The labels of each defect text in the defect text dataset are determined. The distribution of each label category in the defect dataset is shown in [Fig. 4](#).

In the three label categories, sample distribution varies notably for defect severity. This due to the substantial differences in the original defect records. During routine inspections, most defects are classified as General or Serious, with few Critical cases due to low risk of major accidents. For defect components, the top five classes, tanks, cooling systems, on-load tap-changers, bushings, and non-electrical protection, have similar record counts, while defects in iron cores and windings are rare, as they are harder to detect. Regarding defect types, mechanical, insulation, and thermal defects have comparable sample sizes, whereas other types are scarce ([Li & Li, 2017](#)), consistent with expert assessments.

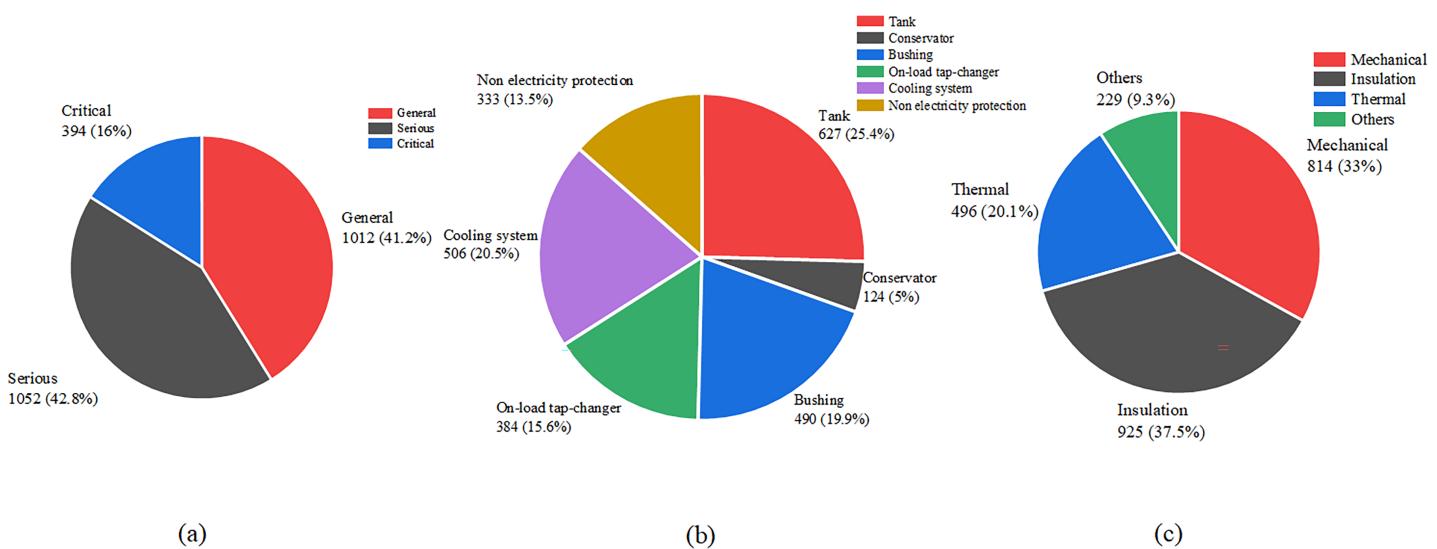


Figure 4 Label distribution of defect text datasets. (A) Defect severity. (B) Defect component. (C) Defect type.

Full-size DOI: 10.7717/peerj-cs.3318/fig-4

Table 3 Parameter values of the classification model.

Parameter	Value
Hidden_size	768
Batch_size	128
Dropout	0.5
Activation function	ReLU
Forward hidden_size	384
Backward hidden_size	384
Learning_rate	5×10^{-5}
Optimizer	Adam
Num_epochs	30
Word vector dimension	128

Classification result analysis of the BERT-BiLSTM model

This model is based on the Windows operating system and uses Python 3.7 programming language. The testing and analysis are conducted based on the TensorFlow deep learning framework and the Keras neural network framework. The key parameters of the model are selected, as shown in Table 3.

Hyperparameters were determined *via* empirical tuning and transformer text classification studies. A learning rate of 5×10^{-5} and batch size of 128 ensured stable convergence. To mitigate overfitting on the moderate, noisy dataset, a dropout layer (0.5) followed the BiLSTM, with early stopping applied based on validation loss. Hidden layer size balanced expressiveness and computation, and 30 epochs sufficed for convergence. Rectified Linear Unit (ReLU) activation and Adam optimizer were used, with Adam adaptively minimizing loss.

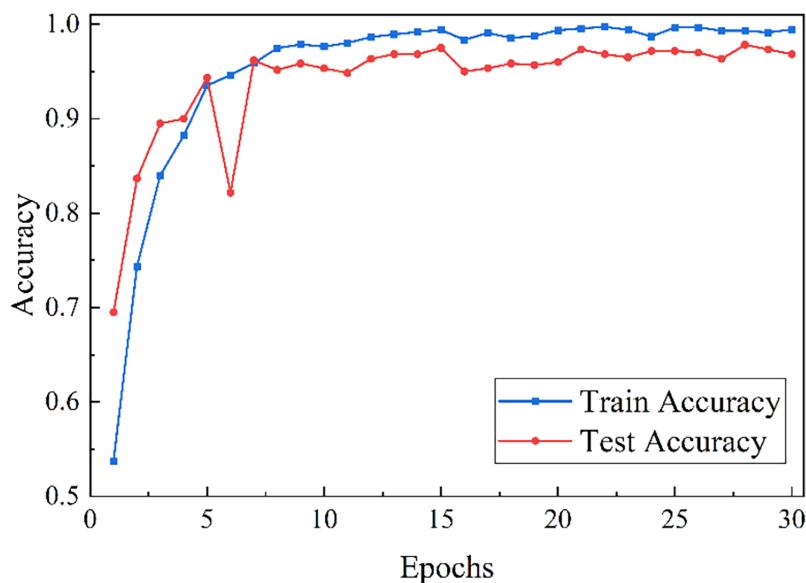


Figure 5 Trend of model classification accuracy and loss. [Full-size](#) DOI: 10.7717/peerj-cs.3318/fig-5

To address class imbalance, class weights—set inversely proportional to class frequency—were incorporated into the loss function, increasing the gradient impact of minority classes and reducing bias toward majority classes without resampling or architectural changes.

Effective text classification is an important module for assessing the performance of text recognition models. In this article, model performance is evaluated using precision, recall, F1-score, accuracy, and the confusion matrix, defined as follows:

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (11)$$

$$Precision = \frac{T_P}{T_P + F_P} \quad (12)$$

$$Recall = \frac{T_P}{T_P + F_N} \quad (13)$$

$$F1\text{-score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (14)$$

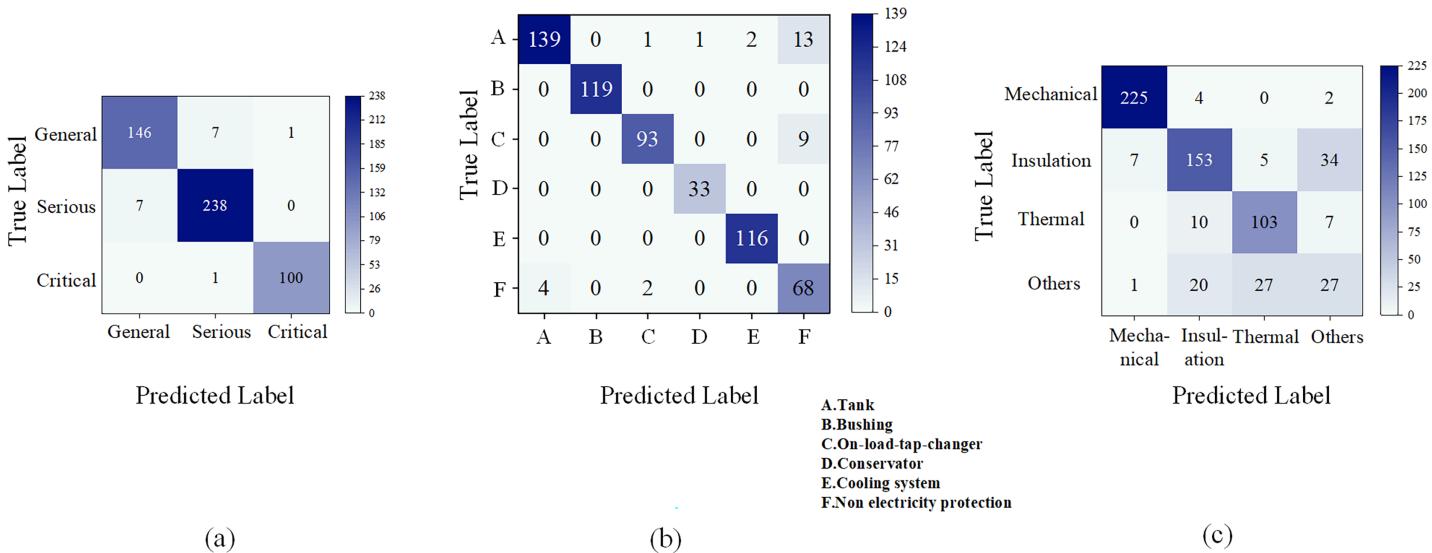
where T_P , T_N , F_P , and F_N correspond to the numbers of true positives case, true negatives case, false positives case, and false negatives case, respectively.

Taking defect severity classification as an example, Fig. 5 presents the model's performance on the training and testing sets. Although fluctuations occur during training, accuracy and loss gradually converge with increasing epochs, stabilizing around epoch 18. At 30 epochs, the model achieves an accuracy of 96.83%.

The evaluation results of the BERT-BiLSTM model for the three defect text classification tasks are shown in Table 4. The model achieves over 94% accuracy for defect severity and defect component classification, while defect type classification attains a comparatively lower accuracy of over 84%.

Table 4 Test effect of BERT-BiLSTM model.

Label type	Precision	Recall	F1-score	Accuracy
Defect severity	0.968	0.970	0.969	0.970
Defect component	0.948	0.947	0.945	0.947
Defect performance	0.845	0.848	0.846	0.848

**Figure 6** Confusion matrix of the classification models. (A) Defect severity. (B) Defect component. (C) Defect type.Full-size DOI: [10.7717/peerj-cs.3318/fig-6](https://doi.org/10.7717/peerj-cs.3318/fig-6)

The confusion matrix results for the three types of text classification problems are shown in Fig. 6. The BERT-BiLSTM model has good performance in defect severity and component classification, but low in defect type recognition.

Misclassification of defect severity mainly arises between general and severe levels due to vague descriptions. Component errors frequently occur among tank, conservator, and non-electrical protection devices when multiple components are mentioned ambiguously. Defect type classification is hindered by sample imbalance and fuzzy semantic boundaries, particularly between mechanical and thermal faults. Many “other” defects, such as auxiliary system issues, share textual features with overheating or insulation faults, leading to misclassification. Moreover, independent category prediction in the BERT-BiLSTM model neglects correlations between defect type and component, reducing classification accuracy.

Comparison with general classification models

Based on the above evaluation metrics, the effectiveness of the BERT-BiLSTM model and commonly used text classification models such as CNN, BERT, and LSTM are compared, as shown in Fig. 7.

The results show that the evaluation metric distribution across the three classification tasks is consistent with other models, but BERT-BiLSTM achieves the highest F1-score

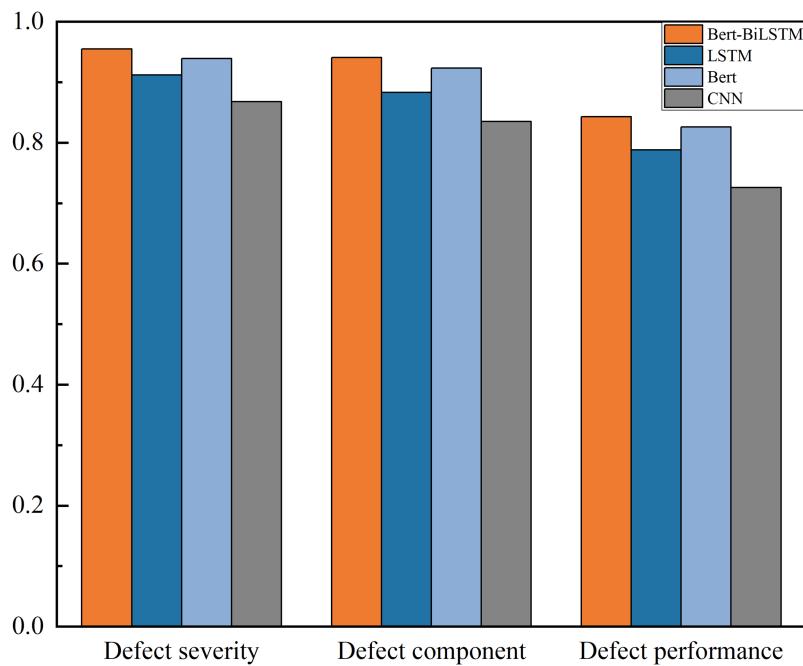


Figure 7 Comparison of model effects under different text classification problems.

Full-size DOI: [10.7717/peerj-cs.3318/fig-7](https://doi.org/10.7717/peerj-cs.3318/fig-7)

(>84%), 2–5% higher than alternatives, indicating superior semantic extraction for transformer defect texts. The F1-score ranking is BERT-BiLSTM > BERT > LSTM > CNN. The reasons may be: (1) LSTM alone outperforms BERT, suggesting that while LSTM efficiently mines defect text features, its global feature extraction contains noise that BERT’s bidirectional encoding can mitigate. (2) CNN performs worst, particularly with long texts, due to its shallow sequential feature extraction and inability to capture deep semantic context.

In addition to classification performance, the computational efficiency and resource demands of the four models are shown in Table 5. The BERT-BiLSTM model has the longest training time (1 h 13 min for 30 epochs), the largest parameter count (~113.5 M), and peak GPU usage (~6.3 GB) due to its dual-layer architecture. Once trained, inference is efficient, taking only a few hundred milliseconds per sample. By contrast, CNN and LSTM consume fewer resources but capture complex semantic dependencies less effectively, leading to lower classification accuracy, highlighting the advantage of BERT-BiLSTM for tasks requiring precise diagnostics.

CONDITION ASSESSMENT OF POWER TRANSFORMER

Transformer condition assessment

Transformer condition assessment, fundamental for equipment maintenance decision-making, evaluates equipment health based on operation, maintenance, and testing. Condition maintenance uses these assessments to determine operational status and guide maintenance plans. Results are expressed as condition scores (quantitative)

Table 5 Performance comparison of different algorithms.

Models	Training times (30 epochs)	Parameters (Million)	Peak GPU memory usage
CNN	45 min	~2.1 M	~1.2 GB
LSTM	58 min	~5.4 M	~1.8 GB
BERT	1 h 05 min	~110 M	~4.5 GB
BERT-BiLSTM	1 h 13 min	~113.5 M	~5.3 GB

and grades (qualitative), which correspond to each other, though grade assignment may vary across standards or procedures.

In transformer health management systems, condition assessment often depends on online monitoring data, offline diagnostic test results, and analysis of feature quantities such as dissolved gases to evaluate the health status of the transformer. However, the influence of historical defect events (recorded as unstructured textual descriptions) on equipment condition should not be neglected. Despite their availability, these records are rarely used as direct inputs of health assessment models.

To address this gap, an assessment method that integrates defect text recognition with transformer condition evaluation is proposed, as shown in [Fig. 8](#).

This method first uses a BERT BiLSTM classification model to extract defect-related information (such as defect type, component, and severity). Then, according to the national transformer health assessment standards, these features are mapped to structured evaluation criteria, enabling automatic generation of condition scores and grades based on defect history. This framework not only broadens the data sources used in condition evaluation, but also supports intelligent fault triage and maintenance decision-making in scenarios where sensor data may be unavailable or insufficient.

Association rules for status evaluation

In actual work, based on the understanding of equipment principles and operating conditions, a reasonable equipment defect classification method can be formulated. Considering the internal structure of the equipment, external environment, and operating conditions, combined with the actual status of the equipment, the defect rating score for power transformer and the weighting of defect status score are shown in [Table 6](#).

In this framework, the BERT-BiLSTM model's classification outputs—defect severity, component, and performance—serve as inputs to the scoring system. Predicted severity levels are mapped to score deductions ([Table 5](#)), and component types are assigned weights ([Table 6](#)). Label definitions and scoring thresholds follow national standards (e.g., DL/T 572, Q/GDW 11247) and are validated by maintenance experts.

Component weights reflect criticality and operational risk, determined *via* expert judgment and historical fault analysis. An adaptive correction mechanism dynamically adjusts health scores based on abnormal operation indicators, improving contextual accuracy. Expert validation indicates this approach better reflects operational risks than fixed weights, although formal ablation studies were not conducted.

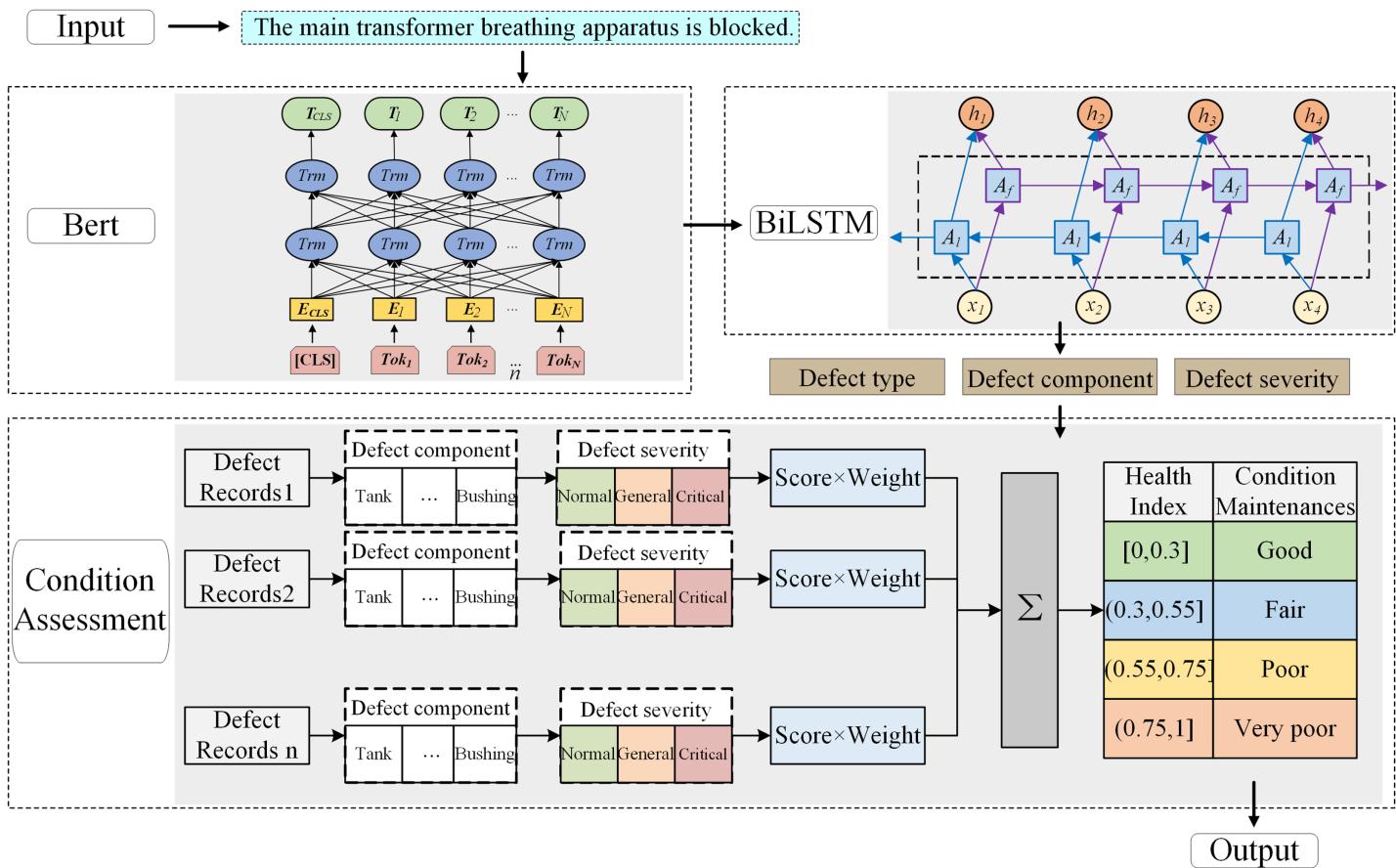


Figure 8 Framework of the assessment method.

Full-size DOI: 10.7717/peerj-cs.3318/fig-8

According to the above condition evaluation rules, the deduction points of its transformer equipment are obtained, corresponding to the health status. Maintenance strategies can be developed based on the results of health index. The health index is calculated as follows:

$$n_i = a_0 k_0 + a_1 k_1 + a_2 k_2 + a_3 k_3 \quad (15)$$

$$HI = \frac{\sum_{i=1}^m w_i \times n_i}{\sum_{i=1}^m n \max w_i} \quad (16)$$

where a and k are respectively the number and score of different defect classes; n_i represents the overall score of the component; w_i represents the weight of the component, as shown in Table 6, and m represents the total number of classifications of defect component.

In traditional state evaluation, indicator weights remain fixed once determined. However, when an indicator exceeds normal values, its abnormal information may be underestimated due to a small original weight, which is impractical. To address this, a dynamic weight calculation with adaptive correction is employed. When an indicator shows a high state value, its importance in accurately assessing the current equipment

Table 6 Scoring method of transformer condition assessment.

Defect rating		Defect status		Transformer condition maintenance		
Equipment defects	Grade interval (k)	Condition	Weight (w_i)	Equipment condition	Health index	Condition maintenances
Normal	0	Tank	0.3	Good	[0, 0.3]	Normal maintenance or extension for 1 year.
General	4	Conservator	0.1	Fair	(0.3, 0.55]	Appropriate reduction of maintenance cycles
Serious	8	Bushing	0.22	Poor	(0.55, 0.75]	Schedule maintenance in due course.
Critical	12	Cooling system	0.12	Very poor	(0.75, 1]	Immediately arrange maintenance.
		On-load tap-changer	0.18			
		Non-electricity protection	0.08			

health increases, making it a key factor in the comprehensive evaluation. Its weight is appropriately adjusted to reflect its critical impact on transformer condition. When n indicators exceed abnormal thresholds, they are recorded as set A, while m normal indicators form set B. The weights of the abnormal indicators are then adjusted as follows:

$$w'_i = w_i + \left(1 - \sum_{k \in A} w_k\right) w_i \quad i \in A. \quad (17)$$

The remaining normal indicators are adjusted as follows:

$$w'_j = w_j - \left(1 - \sum_{k \in B} w_k\right) w_j \quad j \in B \quad (18)$$

$$\sum_{i \in A} w'_i + \sum_{j \in B} w'_j = 1. \quad (19)$$

Using the above scoring rules, the impact of historical defects on a transformer can be assessed to determine its health status and guide appropriate maintenance strategies. According to oil-immersed transformer condition evaluation guidelines, transformer status is categorized into four levels: good, fair, poor, and very poor. The corresponding maintenance programs for each status are listed in [Table 6](#).

Case study

To verify the applicability of the proposed method for transformer condition assessment, we selected a real case involving a 35 kV AC transformer with multiple recorded defect events in the past year. The specific defect instances and their corresponding classification results are shown in [Table 7](#).

Each defect text in [Table 7](#) was processed by the BERT-BiLSTM model to extract three core elements: defect severity, component, and type. Component weights (w_i) were assigned based on a predefined expert knowledge base, and severity levels (k_i) were quantified *via* mapping rules. The overall Health Index (HI) was calculated using [Eqs. \(16\)](#) and [\(17\)](#), with a weighted aggregation across components. An HI of 0.44 corresponds to the “Fair” condition zone in [Table 6](#), indicating moderate degradation. Although a critical defect (e.g., oil tank smoke) exists, its impact is offset by less severe defects and

Table 7 Transformer defect statistics.

Transformer	Defect text	Defect severity	Defect condition	Health index	Equipment condition
A 35kV AC transformer	[Transformer, cooling system, air-cooled] Failure of a single fan.	General $k = 4$	cooling system $w_i = 0.12$	$HI = 0.44$	Fair
	[Transformer, oil-immersed transformer, tank] Oil leakage was detected from the transformer tank at a rate of approximately one drop per minute.	General $k = 4$	tank $w_i = 0.3$		
	The temperature of the B-phase bushing of the transformer reached 55.2 degrees.	Serious $k = 8$	Bushing $w_i = 0.22$		Appropriate reduction of maintenance cycles
	Smoking occurred in the transformer tank.	Critical $k = 12$	Tank $w_i = 0.3$		

component weights. The recommended maintenance strategy is to shorten inspection intervals and perform targeted fault tracing, particularly for the tank and cooling system.

This case demonstrates the feasibility and effectiveness of the proposed framework and highlights the potential of NLP-based methods in digital power equipment management.

CONCLUSION

In intelligent power systems, natural language processing and deep learning are employed to classify defect texts. For transformer equipment, this approach can significantly accelerate fault diagnosis and localization by maintenance personnel, reduce maintenance costs and workload, support scientific maintenance decision-making, and provide essential evidence for assessing current equipment health.

A transformer defect text recognition model based on BERT-BiLSTM was proposed. It can identify features of the severity level, component, and type of equipment defects based on defect texts. The results indicate that the model can achieve an accuracy of more than 94% in identifying the defect severity and defect component. Comparative experiments further validate that the F1-score of the text classification model is more than 2% higher than that of models such as BERT, CNN and LSTM.

After adopting text recognition methods, health assessment rules are determined based on the relevant state assessment guidelines and standards for transformers by evaluating the status of defective equipment. After conducting health indicator evaluation calculations, the healthy operating status of the equipment can be obtained, and reasonable arrangements can be made based on this. The maintenance plan can effectively optimize the allocation of transportation and inspection resources.

Although the proposed approach demonstrates robust performance in defect classification and condition assessment, its generalization may be affected by variability in real-world input. Future work will optimize the model architecture, incorporate correlations between defect locations and types, and integrate real-time operational data,

such as sensor measurements, to enhance diagnostic accuracy and temporal responsiveness. The methodology will be deployed in intelligent maintenance decision support systems and validated through on-site applications and expert feedback.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Science and Technology Project of State Grid Corporation of China under Grant 5700-202455338A-2-1-ZX. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

Science and Technology Project of State Grid Corporation of China: 5700-202455338A-2-1-ZX.

Competing Interests

Tongbao Liu, Hongzhong Chen and Chao Wei are employed by State Grid Jiangsu Electric Power Co., Ltd. The authors declare that they have no competing interests.

Author Contributions

- Jian Shao conceived and designed the experiments, performed the computation work, prepared figures and/or tables, and approved the final draft.
- Xuan Li conceived and designed the experiments, performed the computation work, prepared figures and/or tables, and approved the final draft.
- Tongbao Liu performed the experiments, prepared figures and/or tables, and approved the final draft.
- Hongzhong Chen performed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Tonglei Wang analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Chao Wei analyzed the data, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw measurements and code are available in the [Supplemental Files](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj-cs.3318#supplemental-information>.

REFERENCES

Bhattarai BP, Paudyal S, Luo Y, Mohanpurkar M, Cheung K, Tonkoski R, Zhang X. 2019. Big data analytics in smart grids: state-of-the-art, challenges, opportunities, and future directions. *IET Smart Grid* **2**(2):141–154 DOI [10.1049/iet-stg.2018.0261](https://doi.org/10.1049/iet-stg.2018.0261).

Chen K, Mahfoud RJ, Sun Y, Nan D, Wang K, Haes Alhelou H, Siano P. 2020. Defect texts mining of secondary device in smart substation with GloVe and attention based bidirectional LSTM. *Energies* **13**(17):4522 DOI [10.3390/en13174522](https://doi.org/10.3390/en13174522).

Devlin J, Chang MW, Lee K, Toutanova K. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. ArXiv DOI [10.48550/arXiv.1810.04805](https://doi.org/10.48550/arXiv.1810.04805).

Dong M, Zheng H, Zhang Y, Shi K, Yao S, Kou X, Guo L. 2019. A novel maintenance decision making model of power transformers based on reliability and economy assessment. *IEEE Access* **7**:28778–28790 DOI [10.1109/ACCESS.2019.2897606](https://doi.org/10.1109/ACCESS.2019.2897606).

Feng B, Zhang YW, Tang X, Guo CX, Wang JJ, Yang Q, Wang HF. 2020. Power equipment defect record text mining based on BiLSTM-attention neural network. *Proceedings of the CSEE* **40**(S1):1–10 DOI [10.13334/j.0258-8013.pcsee.200530](https://doi.org/10.13334/j.0258-8013.pcsee.200530).

Ge Q, Wang M, Jiang H, Lu Z, Yao G, Sun C. 2021. Health management of dry-type transformer based on broad learning system. *IEEE Transactions on Industrial Electronics* **69**(3):3027–3036 DOI [10.1109/TIE.2021.3065616](https://doi.org/10.1109/TIE.2021.3065616).

Jang B, Kim M, Harerimana G, Kang SU, Kim JW. 2020. Bi-LSTM model to increase accuracy in text classification: combining Word2vec CNN and attention mechanism. *Applied Sciences* **10**(17):5841 DOI [10.3390/app10175841](https://doi.org/10.3390/app10175841).

Jia J, Yang Q, Fu H, Yang JG, He YD. 2023. Research on pre-training language model construction and text semantic analysis based on power equipment big data. *Proceedings of the CSEE* **43**(3):1027–1037 DOI [10.13334/j.0258-8013.pcsee.212629](https://doi.org/10.13334/j.0258-8013.pcsee.212629).

Jiang YW, Li L, Li ZW, Su C, Wang GJ, Peng XS. 2019. An information mining method of power transformer operation and maintenance texts based on deep semantic learning. *Proceedings of the CSEE* **39**(14):4162–4172 DOI [10.13334/j.0258-8013.pcsee.181457](https://doi.org/10.13334/j.0258-8013.pcsee.181457).

Jina E, Zhang Y, He W, Zhang W, Cao Y, Lv G. 2024. A new method for oil-immersed transformers fault diagnosis based on evidential reasoning rule with optimized probabilistic distributed. *IEEE Access* **12**(5):34289–34305 DOI [10.1109/ACCESS.2024.3370866](https://doi.org/10.1109/ACCESS.2024.3370866).

Li S, Li J. 2017. Condition monitoring and diagnosis of power equipment: review and prospective. *High Voltage* **2**(2):82–91 DOI [10.1049/hve.2017.0026](https://doi.org/10.1049/hve.2017.0026).

Li J, Li G, Hai C, Guo M. 2021. Transformer fault diagnosis based on multi-class adaboost algorithm. *IEEE Access* **10**(10):1522–1532 DOI [10.1109/ACCESS.2021.3135467](https://doi.org/10.1109/ACCESS.2021.3135467).

Liao CB, Yang JX, Hu X, Li K, Li THF, Liu XT. 2023. Hierarchical diagnosis method for transformer faults driven by mixed data and experience. *High Voltage Engineering* **49**(5):1841–1850 DOI [10.13336/j.1003-6520.hve.20221839](https://doi.org/10.13336/j.1003-6520.hve.20221839).

Liao W, Yang D, Wang Y, Ren X. 2021. Fault diagnosis of power transformers using graph convolutional network. *CSEE Journal of Power and Energy Systems* **7**(2):241–249 DOI [10.17775/CSEEJPES.2020.04120](https://doi.org/10.17775/CSEEJPES.2020.04120).

Liu YP, Liu YJ, Liu G, Zhao T, Xu ZQ, Yang N, Jia PF. 2023. Digital twin conception of intelligent operation and maintenance of power transformer. *Proceedings of the CSEE* **43**(22):8636–8652 DOI [10.13334/j.0258-8013.pcsee.221544](https://doi.org/10.13334/j.0258-8013.pcsee.221544).

Liu YP, Liu YJ, Lu FC, Xie Q, Gao SG, Gao F. 2022. Application prospect and key technology of digital twin in power transmission and transformation equipment. *High Voltage Engineering* **48**(5):1621–1633 DOI [10.13336/j.1003-6520.hve.20220450](https://doi.org/10.13336/j.1003-6520.hve.20220450).

Liu ZQ, Wang HF, Cao J, Qiu J. 2018. A classification model of power equipment defect texts based on convolutional neural network. *Power System Technology* 42(2):644–651 DOI 10.13335/j.1000-3673.pst.2017.1377.

Liu YP, Xu ZQ, Li G, Xia YW, Gao SG. 2019. Review on applications of artificial intelligence driven data analysis technology in condition based maintenance of power transformers. *High Voltage Engineering* 45(2):337–348 DOI 10.13336/j.1003-6520.hve.20190130001.

Lu YC, Liao CB, Li Q, Wang TL, Shao J, Zhang Y. 2023. Transformer fault diagnosis method based on voiceprint feature and ensemble learning. *Electric Power Engineering Technology* 42(5):46–55.

Meng Q, Song Y, Mu J, Lv Y, Yang J, Xu L, Meng Q. 2023. Electric power audit text classification with multi-grained pre-trained language model. *IEEE Access* 11:13510–13518 DOI 10.1109/ACCESS.2023.3240162.

Pu TJ, Qiao J, Han X, Zhang GB, Wang XY. 2020. Research and application of artificial intelligence in operation and maintenance for power equipment. *High Voltage Engineering* 46(2):369–383 DOI 10.13336/j.1003-6520.hve.20200131001.

Tamma WR, Prasojo RA, Suwarno. 2021. High voltage power transformer condition assessment considering the health index value and its decreasing rate. *High Voltage* 6(2):314–327 DOI 10.1049/HVE2.12074.

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin L. 2017. Attention is all you need. ArXiv DOI 10.48550/arXiv.1706.03762.

Wang H, Cao J, Lin D. 2019. Deep analysis of power equipment defects based on semantic framework text mining technology. *CSEE Journal of Power and Energy Systems* 8(4):1157–1164 DOI 10.17775/cseejp.2019.00210.

Wang XL, Gu YL, Zhang HR, Liu LH, Liu HS, Li QQ. 2024. Data augmentation and application of defect texts for power equipment based on knowledge integration manifold. *Power System Technology* 48(4):1690–1702 DOI 10.13335/j.1000-3673.pst.2023.0713.

Wang CY, Jiang QY, Tang YJ, Zhu BS, Xiang ZM, Tang J. 2019. Fault diagnosis of power dispatching based on alarm signal text mining. *Electric Power Automation Equipment* 39(4):126–132 DOI 10.16081/j.issn.1006-6047.2019.04.019.

Wei D, Wang B, Lin G, Liu D, Dong Z, Liu H, Liu Y. 2017. Research on unstructured text data mining and fault classification based on RNN-LSTM with malfunction inspection report. *Energies* 10(3):406 DOI 10.3390/en10030406.

Yang CC, Jiang L. 2018. Enriching user experience in online health communities through thread recommendations and heterogeneous information network mining. *IEEE Transactions on Computational Social Systems* 5(4):1049–1060 DOI 10.1109/tcss.2018.2879044.

Ye Z, Guo B, Qi D. 2023. BERT-TriF: an inductive short text classification model for power equipment defect records. *Engineering Reports* 5(10):e12671 DOI 10.1002/ENG2.12671.

Zhang Y, Chen HC, Du Y, Chen M, Liang J, Li J, Yao X. 2021. Power transformer fault diagnosis considering data imbalance and data set fusion. *High Voltage* 6(3):543–554 DOI 10.1049/HVE2.12059.

Zhang D, Chu Z, Gui Q, Wu F, Yang H, Ma Y, Tao W. 2023. Transformer maintenance decision based on condition monitoring and fuzzy probability hybrid reliability assessment. *IET Generation, Transmission & Distribution* 17(4):976–992 DOI 10.1049/GTD2.12718.

Zhou J, Lu Y, Dai HN, Wang H, Xiao H. 2019. Sentiment analysis of Chinese microblog based on stacked bidirectional LSTM. *IEEE Access* 7:38856–38866 DOI 10.1109/ACCESS.2019.2905048.