An Intelligent Assisted Assessment Method for Distribution Grid Engineering Quality Based on Large Language Modelling and Ensemble Learning

Wei Hu, Tingting Zheng, Puliang Du, and Zhiwei Chen

Abstract—**A safe and dependable supply of energy and power is directly correlated with the quality of distribution network engineering. The assessment and diagnosis of the design quality and economic viability of a distribution network engineering process are essential for guaranteeing the steady functioning of the corresponding power system. In this paper, an intelligent assisted assessment technique for distribution network engineering is proposed to address the issues of inefficiency, high manual dependence, and low utilization of vital information in the text during the evaluation of projects related to distribution network engineering. To improve the model's contextual learning ability, the robustly optimized bidirectional encoder representations from transformers pretraining approach and whole-word masking are adopted to extract useful features from the distribution network engineering project review text. Principal component analysis is then used to downscale the high-dimensional features, thereby greatly increasing the efficiency of downstream classification. The light gradient boosting machine performs classification on the downscaled text features, and the Bayesian optimization approach is utilized to identify the best hyperparameter combinations. This significantly lessens the impacts of random parameters on the model performance. Tenfold cross-validation results demonstrate that the model can quickly and accurately identify common problems in distribution network projects' technical and economic dimensions.**

Index Terms—**Large language model, ensemble learning, distribution network engineering, intelligent assisted assessment.**

Wei Hu, Tingting Zheng (corresponding author), and Puliang Du are with the School of Economics and Management, Shanghai University of Electric Power, Shanghai 201306, China (e-mail: hwshiep@shiep.edu.cn; zhtingtingsh@mail.shiep.edu.cn; dupuliang@ 163.com).

Zhiwei Chen is with the School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China (e-mail: ycczw2018@163.sufe.edu.cn).

I. INTRODUCTION

A. Background and Motivation

The "14th Five-Year Plan" for China highlights the need for a future distribution network to progressively transition into a new kind of active, intelligent, and collaborative network while increasing the capacity and quality of long-distance transmissions. The power supply and transmission line construction processes are comparatively efficient, and the overall electricity usage in society remains stable [1]. Currently, the planning procedure for the construction of a power grid primarily focuses on the distribution network rather than the backbone network. By 2025, 34% of terminal energy consumption is expected to come from electric energy, and the electrification level of energy consumption will continue to rise. Distribution networks are crucial for ensuring a steady and dependable supply of electric energy since they serve as the central links between the energy production, conversion, and consumption processes [2], [3]. The wide ranges of loads that are accessed, including energy storage facilities, electric vehicles, and renewable energy sources, have raised the construction quality standards of distribution network projects. Strictly reviewing a distribution network project, guaranteeing the construction quality of the network, and bolstering its grid security are essential for creating a robust and trustworthy intelligent distribution network [4], [5]. Distribution networks also contain complex data and a wide variety of devices. The creation of an effective data mining and analysis technique can standardize the way in which typical technical and financial issues are handled during the design phase, thus raising the intelligence and digitalization levels of the distribution network and ensuring the construction quality of the project.

High-quality distribution network design may facilitate energy access, strengthen the security and stability of the energy system, and encourage the sustainable growth of the energy industry [6], [7]. It is vital to invest in new building and repair projects for the distribution network since issues such as the ageing of poles, towers,

and circuit breakers increase as the service lives of distribution network facilities increase [8]. The relevant evaluation units must analyze the line direction, wire, selected equipment, network structure, economic feasibility, and other aspects of the project design to ensure that the design scheme satisfies the technical and financial standards of the distribution network project. When performing evaluation tasks, the evaluation units produce a large amount of textual data. Reducing the frequency of issues and improving the quality of the engineering design process can be achieved by effectively utilizing textual information acquired from distribution network engineering design quality problems, categorizing and prioritizing common design faults, and managing them through targeted, scientific methods. To increase the effectiveness of the review process, an intelligent auxiliary review method is urgently needed as most existing review activities are manual, challenging, and time-consuming. To achieve high-caliber construction and scientific planning for energy engineering projects, it is imperative to define common design errors and increase the evaluation depths of distribution network projects.

B. Literature Review

The offline expert review mode limits the quality and efficiency of distribution network project evaluations, while the wide variety of such projects further complicates the review process. The development of smart grids drives advancements in energy sector information technology, while with power-related data growing exponentially, it is critical to implement intelligently supported evaluations of distribution network projects [9]. The intelligent auxiliary evaluation process of distribution network engineering relies on easily analyzed, structured data. Based on the cost data accumulated in power grid projects, reference [10] designs a set of big data mining-based engineering construction cost management methods to predict the reasonable price of equipment and provide a foundation for determining whether various engineering construction costs are reasonable. To conduct accurate management and obtain lean measurements, reference [11] proposes the main characteristic indices of the total construction process of a distribution network project based on historical empirical data concerning the construction cost of the project. Academics tend to overlook high-value text data in favor of organized data from which it is simple to extract information, such as project costs and fees and historical data. Numerous important details, including project design difficulties and review comments, are gathered in review documents for distribution network engineering designs. Nevertheless, this review text information is not properly mined and utilized because of the intricacy of the text data themselves. The utilization of a large language model (LLM) for information mining and analysis is crucial for advancing the digital transformation and enhancement of distribution network engineering reviews. A standardized management method for common problems can reduce the rate of manual classification errors and provide improved review efficiency by thoroughly mining high-value information from expert review opinion texts and matching it with problem causes. Additionally, the frequency of common problems can be reduced by providing reference opinions for project design units to write feasibility and preliminary reports.

An LLM integrates deep learning and natural language processing (NLP) and carries out model pre-training and fine-tuning with the help of a large-scale text corpus [12]. Typical LLMs include bidirectional encoder representations from transformers (BERT), enhanced representation through knowledge integration (ERNIE), and ChatGPT. A research gap remains in the engineering and construction of power distribution networks, and LLMs are currently only used in the early stages of exploration in the field of electric power, e.g., power dispatching, power auditing, price prediction, power systems, and other related application scenarios [13]. A strategy for predicting crude oil prices that considers public opinion is creatively presented in [14], in which the characteristics of public attitudes about crude oil prices are retrieved using the BERT model, and are then decomposed twice to provide accurate crude oil price predictions. To address the power audit text classification problem, a fine-tuning model based on BERT is proposed, and the model outperforms other fully supervised models by large margins [15]. Robustly optimized bert pretraining approach (RoBERTa) addresses the static masking issue faced by the BERT model, and its context understanding ability is further enhanced by the use of more training data and dynamic masking operations. The substantial semantic feature parsing and extraction benefits of the RoBERTa model are confirmed by applying the model in feature extraction and terminology classification tasks involving power dispatch text [16]. Similarly, RoBERTa and attention are used to accomplish named entity recognition for scheduling text in the realm of power scheduling [17]. However, the above research do not involve the construction of a distribution network engineering process. In addition, LLMs have been mainly used to analyze text features and complete classification tasks without combining other classification models.

Bagging, boosting, and stacking are the three primary classes that comprise integration learning, which is a type of machine learning method that is trained to generate a powerful learner by joining numerous poorly supervised models [18]. The power industry has relatively developed integrated learning applications, which perform well in a variety of scenarios, such as load forecasting and wind power prediction. To evaluate the robustness of contemporary distribution networks, a probabilistic model is created based on random forest that represents the likelihood of overhead line outages due to crown snow [19]. Reference [20] proposes a probabilistic wind power prediction model considering

error correction, while the prediction results of the gradient boosting decision tree (GBDT) are also corrected to improve the accuracy of the model. This is done to satisfy the imposed energy transition requirements and ensure a balance between the supply and demand of urban energy. To anticipate the demand for regional energy systems, long short-term memory (LSTM) with extreme gradient boosting (XGBoost) is used, while the experimental findings demonstrates that XGBoost can greatly increase the training speed while maintaining a low error rate [21]. When working with large amounts of data, GBDT and XGBoost have significant restrictions because they must iteratively traverse the full training set multiple times. On the other hand, the light gradient boosting machine (LightGBM) has a quicker processing speed and a lower memory footprint, allowing it to handle concurrent training steps. LightGBM is used to forecast the flexibility of electric vehicle charging, and studies have demonstrated that this model offers clear benefits when handling large datasets [22]. In conclusion, the current electric power methods primarily rely on numerical data for prediction and classification using ensemble learning models, without utilizing textual data related to electric power.

C. Contributions

Currently, the majority of the research on reviewing and managing distribution network projects focuses on investment strategies and process optimization techniques, with little attention given to the analysis and handling of technical and financial design issues. The design of a distribution network project, which has an impact on energy field safety, is critical to the quality of the construction process. Artificial intelligence and big data must be included in engineering construction quality management to build a digital and intelligent new energy system. However, text data in the field of distribution network engineering have not yet been effectively mined and analyzed. Thus, this research provides an intelligent design problem diagnosis model that combines an LLM and ensemble learning based on historical project data derived from distribution network engineering scenarios.

1) Advanced text feature extraction technology: The robustly optimized bidirectional encoder representations from transformers pretraining approach and whole-word masking (RoBERTa-wwm) model is creatively applied to the text of the power distribution network project evaluation to extract and analyze features. The RoBERTa-wwm model is more accurate at processing texts in the electric power domain than the Bert model since it is full-word masked at word granularity and pre-trained on a larger variety of datasets.

2) Efficient classification model and parameter tuning strategy: An integrated learning approach is employed to feed the dimensionality-reduced text features into the LightGBM model for effective classification. The model-specific gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB) algorithms

maximize the processing power for large-scale datasets, while the Bayesian technique is used simultaneously to search for the best super-parameter combination, thus greatly enhancing the model's performance.

3) Innovative intelligent evaluation model: To provide an intelligent assisted evaluation model for distribution network engineering, this research combines the semantic analysis of RoBERTa-wwm, the classification benefits of principal component analysis (PCA) downscaling, and Bayes-LightGBM. The comprehensive performance of the combined model in terms of numerous assessment indexes is thoroughly confirmed via ten-fold cross-validation.

4) Important practical application value: This research proposes an intelligent auxiliary assessment model of distribution network engineering quality, which is able to perform the intelligent classification of engineering problems and efficiently identify the main weaknesses in engineering design. In addition to ensuring energy security and fostering the intelligent and digital transformation of power systems, the model offers significant practical value in supporting the efficient landing of distribution network engineering projects.

II. GENERAL FRAMEWORK

The research framework of this paper is shown in Fig. 1. Text feature extraction and classification are the two parts of the intelligent approach developed for diagnosing the typical issues faced in distribution network engineering scenarios.

RoBERTa-wwm is used to extract semantic information from the input textual design for engineering a defective distribution network. First, the supplied text is subjected to text embedding and position embedding. Second, information is fed into the six-layer encoder structure of a transformer, which employs a multi-head attention technique to parse the semantic structures in different regions. The output is then fed into the next encoder, which passes through 12 transformers before producing the final text embedding.

Text features are extracted by RoBERTa-wwm and used in the downstream classification process. Since the 768-dimensional characteristics obtained from the upstream task may contain redundant information, PCA is used to reduce the dimensionality of the features. The resulting features are more representational and operate with a significant increase in efficiency following the dimensionality reduction process. The major features acquired after implementing dimensionality reduction are carried over to the LightGBM, whose unique histogram formation method and leaf node development strategy enable the model to significantly boost its computational efficiency without sacrificing accuracy. Finally, the LightGBM hyperparameters are tuned, and the Bayesian technique is used to determine the optimal hyperparameter combination.
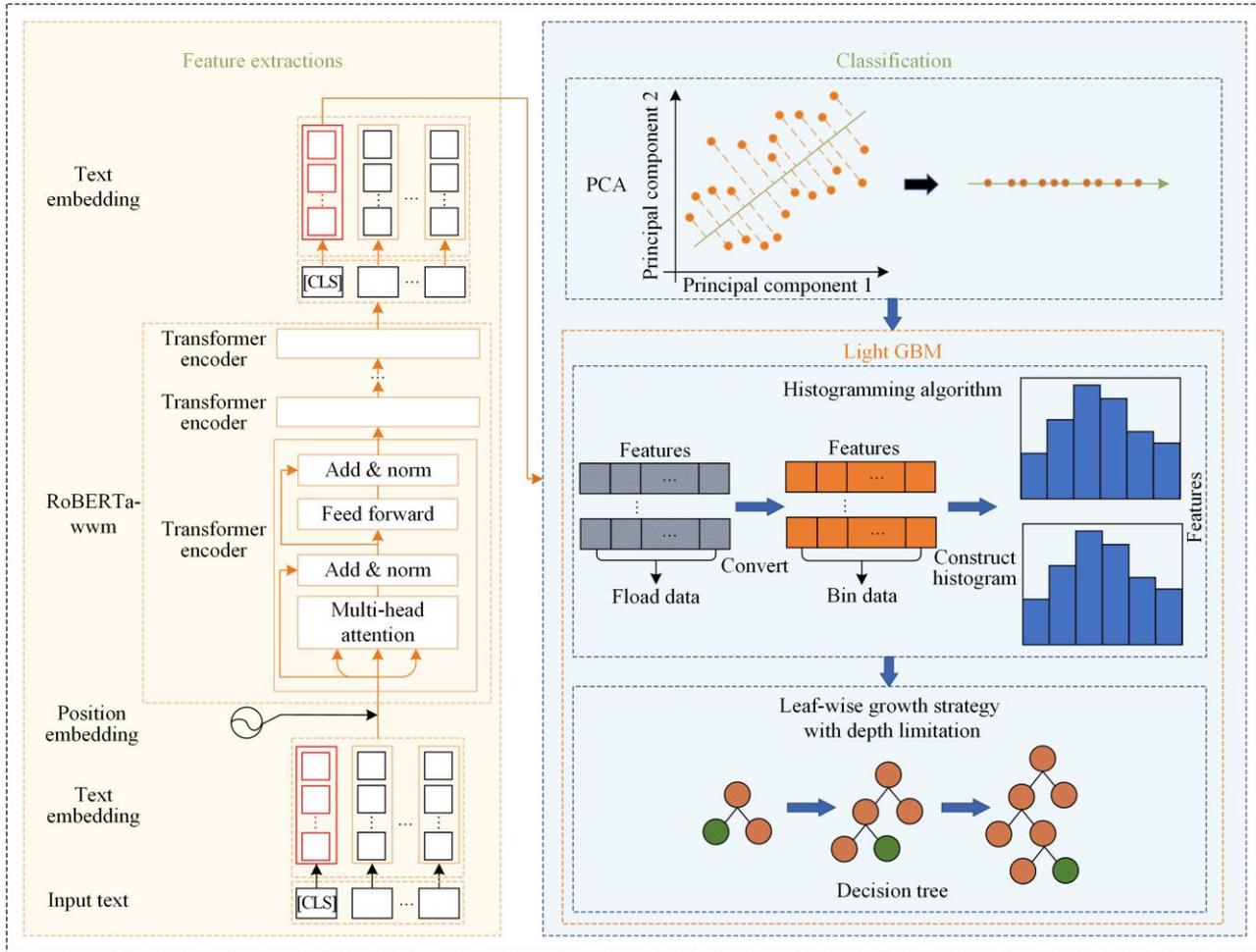
Fig. 1. Model framework.

III. BASIC METHODOLOGY

A. RoBERTa-wwm

To acquire the semantic properties of defective text for distribution network engineering design, RoB-ERTa-wwm is chosen in this paper. This approach is able to fully acquire the semantic information and rules of the Chinese language, which allows it to better comprehend contextual relationships. This is achieved by combining the advantages of the RoBERTa model with the whole-word masking strategy, which is a novel masking technique created specifically for Chinese expression features. Chinese word splitting is not accounted for when masking at the single word level, and thus, it is simple to overlook the dependencies between terms [23]. The three elements of RoBERTa-wwm are better than those of the BERT model.

1) RoBERTa-wwm derives a Chinese participle-based whole-word masking strategy from the single-character masking approach of BERT, as Fig. 2 illustrates the precise structure. In a case where the task is to "verify the number of removed cables," RoB-ERTa-wwm masks all of the letters in the target word together, and the whole-word masking policy treats "removed" as a group of words, masking two characters. By capturing the same phrases in various contexts, this masking method completely utilizes the semantic correlations between words and enhances the overall performance of model pretraining.

2) RoBERTa creates a fresh mask at random for every input, which contrasts with the static mask of BERT. Static masking, or using a fixed portion of the input token for mask processing, is what the classic BERT model does during pretraining [24]. The utilization of dynamic masking enables RoBERTa-wwm to perform autonomous mask operations on every training instance, hence enhancing the sample diversity and augmenting the processing capacity of the model for various scenarios. This dynamic masking technique enhances the efficacy and generalizability of the model in a variety of language tasks by helping RoB-ERTa-wwm learn and use the input context information more effectively.

3) During the pretraining phase, RoBERTa cancels the next-sentence-prediction (NSP) assignment. The NSP task is intended to ascertain whether two sentences are continuous. This process is used as a pretraining task in the BERT model. Studies have revealed that NSP tasks only slightly boost the performance of the resulting

model in numerous downstream tasks. And as a result, by removing the NSP process, RoBERTa-wwm concentrates

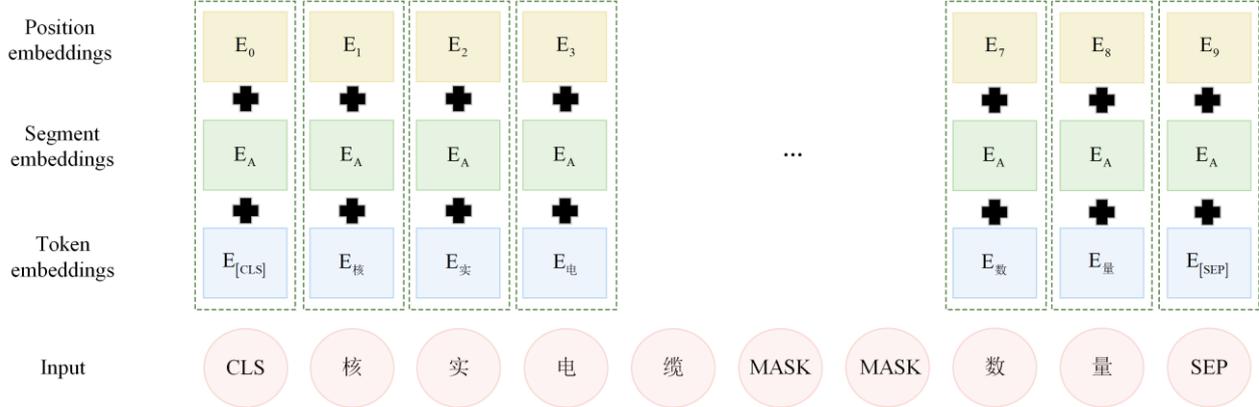more time and resources on language modelling and word-level prediction.

Fig. 2. RoBERTa-wwm model.

Many professional-domain terms are used in engineering design quality assessments for distribution networks. To properly represent domain word properties and language rules, the RoBERTa-wwm model consists of 12 transformer encoder layers. To assist the model in extracting key features from many dimensions, each layer encoder is composed of a feedforward neural network and a multi-head self-attention mechanism. The structure of the coding layers is shown in the left module of Fig. 3, which is the structural diagram of the transformer. The encoder block of the transformer is made up of feedforward, add & norm, multi-head attention, and add & norm layers in order.

Fig. 3. Transformer structure.

This research uses the multi-head attention mechanism to extract text features. The inputs of the model or the outputs of the preceding coding layer are multiplied by initialized linear transformation matrices that are randomly generated to produce matrices $\boldsymbol{Q}_i$, $\boldsymbol{K}_i$, and $\boldsymbol{V}_i$. Eight self-attention layers form the multi-head attention mechanism, which allows the model to concentrate on parsing significant words by acquiring semantic information from many areas. To obtain the final output vector, the outputs of the eight self-attention layers are combined and processed through a linear transformation layer.

The calculation formulas are as follows:

$$\boldsymbol{Q}_i = \boldsymbol{X}\boldsymbol{W}_i^{\mathrm{Q}} \tag{1}$$

$$\boldsymbol{K}_i = \boldsymbol{X}\boldsymbol{W}_i^{\mathrm{K}} \tag{2}$$

$$\mathbf{V}_i = \boldsymbol{X}\boldsymbol{W}_i^{\mathrm{V}} \tag{3}$$

where $\boldsymbol{X}$ is the input matrix of the transformer; while $\boldsymbol{W}_i^{\mathrm{Q}}$, $\boldsymbol{W}_i^{\mathrm{K}}$ and $\boldsymbol{W}_i^{\mathrm{V}}$ are the weight matrices.

After the $\boldsymbol{Q}_i$, $\boldsymbol{K}_i$, and $\boldsymbol{V}_i$ matrices are obtained, the output of the self-attention process can be determined. The softmax function is used to calculate the attention coefficients between each word and other words after the inner product of the matrices is determined. Subsequently, the coefficients are multiplied by matrix $\boldsymbol{V}_i$ to obtain the output, as:

$$\boldsymbol{Z}_i = \mathrm{softmax}(\frac{\boldsymbol{Q}_i \times \boldsymbol{K}_i^{\mathrm{T}}}{\sqrt{D_{\mathrm{K}}}})\boldsymbol{V}_i \tag{4}$$

where D_{K} represents the dimensionality of $\boldsymbol{K}_i$. The inner product is divided by the square root of D_{K} to prevent it from growing too large.

The final output of the multi-head attention mechanism is obtained by splicing the outputs of different self-attention mechanisms to create a matrix. This matrix is then sent to the linear transformation layer, and its dimensions match those of the input matrix. A two-layer
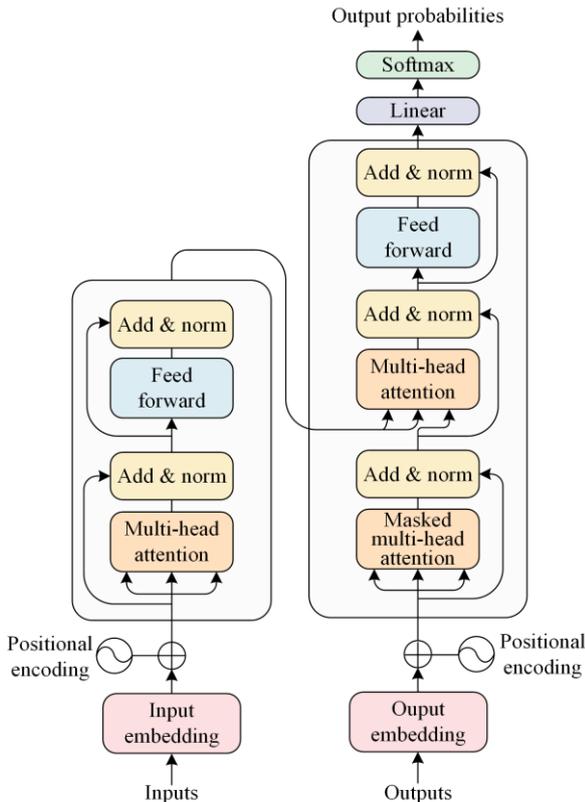
fully connected layer is obtained when the multi-head attention output is connected with the residuals, standardized processing is performed, and the final output matrix is obtained.

B. PCA

PCA is a linear dimensionality reduction technique that is used to extract significant features, eliminate redundant data, and map high-dimensional characteristics to a low-dimensional space. The following are the precise steps.

The sample features $s^{(i)}$ extracted by RoBERTa-wwm are first centred; m represents the total number of samples in the dataset used for PCA.

$$S^{(i)} = s^{(i)} - \frac{1}{m}\sum_{j=1}^{m} s^{(j)} \tag{5}$$

The covariance matrix of $s^{(i)}$ is then computed:

$$C = \frac{1}{m}SS^{\mathrm{T}} \tag{6}$$

Eigenvalue decomposition is performed on the covariance matrix C to find its eigenvalues and eigenvectors [25]. The eigenvector A corresponding to the largest eigenvalue d is found, and all eigenvectors are standardized to form an eigenvector matrix $(a_1, a_2, \cdots, a_d)$. Each sample $S^{(i)}$ in the sample set is then transformed into a new sample $G^{(i)}$:

$$G^{(i)} = A^{\mathrm{T}} S^{(i)} \tag{7}$$

C. LightGBM

This article uses LightGBM to intelligently classify engineering design quality problems involving distribution networks. Based on GBDT, LightGBM is an enhanced ensemble learning technique [26]. Its strategy is tailored to address issues such as the challenge of identifying the ideal segmentation point of a decision tree and the low effectiveness of the leaf growth mode. By ensuring the accuracy of the model, its running speed is greatly improved, and its memory consumption level is reduced. The design defect text features and problem category labels extracted by RoBERTa-wwm are transferred to the LightGBM model as inputs. The weight values corresponding to each leaf node are calculated according to the differences between gradients, and the scores of each category are obtained. The category with the highest scores is the final output classification result.

The objective function of LightGBM is shown in equation (8):

$$\mathrm{obj}(i) = L(Y, \hat{Y}^t) + \sum_{i=1}^{t} \Omega(f_i) \tag{8}$$

$$L(Y, Y^t) = \sum_{i=1}^{n} l(y_i, y_i^{(t)}) = \sum_{i=1}^{n} l(y_i, y_i^{(t-1)} + f_t(x_i)) \tag{9}$$

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \tag{10}$$

where Ω is the regularization term; n refers to the number of samples; y_i and $\hat{y}_i^{(t)}$ represent the true value and the predicted value of label i, respectively; $\hat{y}_i^{(t-1)}$ represents the predicted values for the previous $t-1$ steps; $f_t(x_i)$ denotes the prediction increment made by the model for the ith sample x_i at the tth iteration; $l(\cdot)$ serves as the loss function, which is employed to quantify the discrepancy between the predicted and true values; ω_j is the weight for controlling the j th leaf node; while γ and λ are the regularization parameters that prevent overfitting. A second-order Taylor expansion of the objective function with respect to $\hat{y}_i^{(t-1)}$ is produced, as:

$$\mathrm{obj}(i) = \sum_{i=1}^{n}\left[l(y_i, y_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i) \right] + \sum_{i=1}^{t} \Omega(f_i) \tag{11}$$

where g_i and h_i are expressed as the first and second partial derivatives of f_t, respectively. When the minimum value of the objective function is obtained, the corresponding $\hat{y}_i^{(t)}$ is the final prediction category.

The objective function definition of LightGBM is the same as that of XGBoost but with the addition of GOSS and EFB. The main goal of the GOSS algorithm is to retain all large gradient samples. Because of their minimal information gains, the small gradient samples are sampled at random. The large gradient samples are the first a samples of the absolute gradient values, and the small gradient samples are those that result from selecting $b \times (1-a)$ values at random from the remaining samples. Equation (12) displays the information gain determined by sampling based on the gradient size, where A is the large gradient sample; B is the retained small gradient sample; during feature splitting, both A and B are divided into left (A_1, B_1) and right (A_r, B_r) leaves based on feature j; g_i denotes the gradient value; j is the number of features; while m_1^j and m_r^j denote the numbers of left and right leaf nodes, respectively.

$$\tilde{V}_j(d) = \frac{1}{m}\left[\left(\sum_{x_i \in A_1} g_i + \frac{1-a}{b}\sum_{x_i \in B_1} g_i\right)^2 \Bigg/ m_1^j(d) + \left(\sum_{x_i \in A_r} g_i + \frac{1-a}{b}\sum_{x_i \in B_r} g_i\right)^2 \Bigg/ m_r^j(d) \right] \tag{12}$$

To accelerate the model training process, GOSS regulates the quantity of data in accordance with the gradient size. By combining mutually incompatible features, the EFB method further decreases the data quantity. By combining features with low conflict rates into a single feature and establishing a maximum

allowable conflict threshold K, the algorithm decreases sthe feature dimensionality [27]. The histogram algorithm of LightGBM facilitates the merging of mutually exclusive features. With the help of the histogram algorithm, only n bins need to be traversed instead of all the features to determine the optimal splitting point, thus striking a good balance between training efficiency and accuracy.

D. Bayesian Hyperparameter Optimization Algorithm

The fundamental concept of the Bayesian optimization process for the LightGBM hyperparameters is as follows. The objective function value is determined by continuously searching for new hyperparameter combinations based on the prior information formed by the historical data, the posterior probability distribution is updated, and the objective value is optimized in a step-by-step to direct the hyperparameter search procedure. The following are the basic steps.

A set of hyperparameter vectors is initialized in the given hyperparameter space:

$$X = [x_0, x_1, \cdots, x_n] \quad (13)$$

By substituting the hyperparameter combination into the model, we obtain the values of the objective function and the initial dataset D_0 [28] as (14), where $f(X)$ represents the objective function value of the hyperparameter combination X.

$$D_0 = \{X, f(X)\} \quad (14)$$

The next sampling point is determined by maximizing the collection function after its calculation:

$$x_t = \arg\max \alpha(x | D_{t-1}) \quad (15)$$

where α is a key value used in Bayesian optimization to balance exploration and utilization; and D_{t-1} represents the accumulated dataset of sampling data at the $(t-1)$th iteration, encompassing all evaluated combinations of hyperparameters along with their corresponding objective function values.

The objective function value $f(x_t)$ of the new sampling point x_t is calculated and added to the sampled dataset:

$$D_t = D_{t-1} \bigcup \{x_t, f(x_t)\} \quad (16)$$

When the maximum number of iterations is reached, the optimal sampling point $\{x^*, f(x^*)\}$ is returned.

E. Tenfold Cross-validation

This work uses the tenfold cross-validation method to evaluate the performance of the proposed model in terms of eliminating the uncertainty caused by the random division of data and to accurately assess the generalizability of the model. To guarantee the consistency of the data distribution, the feature dataset is stratified, sampled and split into ten mutually exclusive subsets following dimensionality reduction. To create 10 distinct pairs training and test sets, one subset is utilized as the test set, and the other nine subsets are utilized as the training set. Ten model training splits yield ten sets of test data. The overall performance of the model is determined by taking the mean of the findings produced for the ten sets.

Samples with the same quantity as those in the test set are randomly chosen as the validation subset, and the remaining samples serve as the training subset for Bayesian hyperparameter optimization. The parameters are adjusted in accordance with how well their hyperparameter combinations perform on the validation subset. The LightGBM hyperparameter combination serves as the input of the Bayesian framework, and the resulting accuracy is the output. Finally, the model performance is evaluated on the test set after the ideal hyperparameters are transferred to the training set for retraining. Figure 4 displays the unique experimental framework.
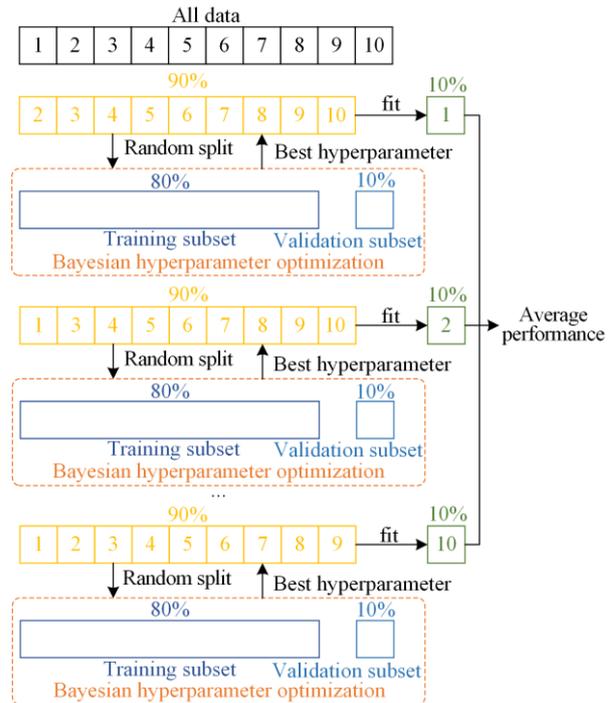

Fig. 4. Experimental framework.

F. Evaluation Indicators

The accuracy, precision, recall, and F1 score are popular metrics for assessing the performance of a model in binary classifications. In this study, we additionally consider the balanced error rate (BER) and computing speed of the model. Table I shows the constructed 2×2 contingency table. The number of samples judged by the classifier to be positive cases that are actually positive cases is E_{TP}, and the number of samples predicted by the classifier to be positive but are actually negative cases is E_{FP}. The number of samples judged by the classifier to be negative cases but are actually positive cases is E_{FN}, and the number of samples predicted by the model to be negative that are actually negative cases is E_{TN}.

TABLE I
2×2 CONTINGENCY TABLE

	Actual sample is a positive example	Actual sample is a negative example
Classifier judgement is a positive example	E_{TP}	E_{FP}
Classifier judgement is a negative example	E_{FN}	E_{TN}

The percentage of samples that the classifier correctly identifies as a whole is known as the $R_{Accuracy}$:

$$R_{Accuracy} = \frac{E_{TP} + E_{TN}}{E_{TP} + E_{TN} + E_{FP} + E_{FN}} \quad (17)$$

$R_{Precision}$ is the rate of correctness achieved when the classifier diagnoses samples as belonging to the corresponding category:

$$R_{Precision} = \frac{E_{TP}}{E_{TP} + E_{FN}} \quad (18)$$

The percentage of the total number of questions in a category that are accurately classified is known as the R_{Recall}:

$$R_{Recall} = \frac{E_{TP}}{E_{TP} + E_{FP}} \quad (19)$$

$R_{Precision}$ and R_{Recall} are averaged together to create the F1-score R_{F1}:

$$R_{F1} = \frac{2 \times R_{Precision} \times R_{Recall}}{R_{Precision} + R_{Recall}} \quad (20)$$

R_{BER} is the mean value of the proportion of positive cases that are misjudged and the proportion of negative cases that are misjudged:

$$R_{BER} = \frac{1}{2}(\frac{E_{FN}}{E_{TP} + E_{FN}} + \frac{E_{FP}}{E_{FP} + E_{TN}}) \quad (21)$$

The intelligent diagnosis of distribution network engineering design issues is a multiclassification problem. The average performance achieved across the n categories in terms of each indicator can then be computed. To do this, the ith category is considered a positive example, the remaining $n-1$ categories are considered negative examples, and so on.

IV. EXPERIMENTS AND RESULTS ANALYSIS

A. Dataset

The primary method for determining whether a distribution network construction project satisfies the imposed criteria is to assess the viability of the project design from both technical and economic standpoints. The business data of the Shanghai Institute of Economics and Technology (SITE) provided the dataset used in this article. The dataset includes 9290 audited data points related to substation projects, line projects, automation projects, and other projects related to the distribution network technology renovation. The distribution network engineering problems have a complex structure and a large amount of text data, making them challenging for standard model-driven approaches to

tackle, as revealed by an analysis of the dataset's features. There is vast amount of useful information in the distribution network engineering project review text, and the large language model can be used to extract the text features effectively. By mining the historical business data, a library of frequently asked questions on the project can be created, which can help the review experts make decisions and standardize engineering project management. However, integrated learning is necessary to complete the classification task after feature extraction. To address this complicated issue, this study develops a set of data-driven techniques that combine integrated learning and big language models. The application scenarios of the models that are created align with the research objectives of this work.

To ensure the safety and dependability of power transmission, the design unit should, at the technical level, choose standard-compliant cables, switchgears, circuit breakers, transformers, and other supporting equipment in accordance with the voltage level, capacity, and performance requirements of various project types. However, safety risks are frequently encountered throughout the design phase of a construction program. The risks include flawed engineering designs, improper equipment selections and installation locations, and inadequate support materials. A few of the technical issues concerning distribution network engineering are listed in Table II, along with their related categories. For example, if mutual stranded cables are used to replace the distribution lines below a pole substation in the line rehabilitation project, but surge arresters are not installed, this poses a risk. Surge arresters are necessary to protect power equipment from atmospheric voltage transients and lightning strikes. Failing to install them during equipment replacement highlights a flaw in the engineering design.

TABLE II
DISTRIBUTION NETWORK ENGINEERING ISSUES AND THEIR CATEGORIES

Problem description	Question tag
Surge arresters should be added to the pole when changing lead wires to replace interstranded cables	Technical-engineering inadequacies
Circuit breaker installation should be changed to load switch installation	Technical-engineering approval of equipment for the installation site
An auxiliary grounding device needs to be installed at the end of the line	Technical-supporting documents, including drawings and photographs
The oil chromatograph device of the converter transformer should be installed in conjunction with the power outage of the converter transformer	Technical-supporting documents, excluding power work orders
Separate quotas should be for the suspension clamp and the tension clamp	Economic-quota recognition
The cost required for renovating the zero-flow pressure plate of the switch station should be verified	Economic-cost verification
The power cable work should be verified	Economic-quantity accounting

The budget estimates and budget quota documents of the National Energy Administration should be followed when creating estimates and budgets at the economic level. The costs of major materials and equipment should be based on the information prices published every quarter. Nonetheless, instances may occur in which the volume and cost of a project are not determined in accordance with the standards during the actual design phase of the construction program. A few of the issues related to the financial aspects of distribution network engineering are listed in Table II, along with their related categories. For example, in the line reconstruction project, various wire clip types are employed. To guarantee the stability and dependability of the line, overhanging wire clips and tension-resistant wire clips are used, depending on the distance and angle between the conductor and the tower, respectively. However, this design may encounter a quota confirmation issue because the two types of wire clips are not distinguished when the quota is applied.

B. Feature Extraction and Dimensionality Reduction

Upon analyzing the dataset, we find that the maximum number of tokens in the distribution network engineering review text is 69, with a total of 9290 texts, which constitutes a large volume of data. The RoB-ERTa-wwm model, which has a token limit of 512, demonstrates robust performance across various NLP tasks. It has a distinct advantage in terms of handling large-scale textual data, making it suitable for extracting features from the text of the distribution network engineering process. The design of the full-word mask can capture the semantics of each word in different contexts. Through the structure of the 12 transformer encoding layers in RoBERTa, the project engineering audit text is captured bidirectionally, enabling a better understanding of the contextual semantic information and the extraction of more complete, accurate, and rich semantic features [29].

The RoBERTa-wwm model maps each token to a 768-dimensional vector space based on a thorough consideration of the model capacity, the computational resources, and the model performance. This mapping step enhances the model's capacity to fully capture subtle features across languages and comprehend the relationships between semantics. RoBERTa-wwm converts the text of the distribution network engineering review into a 768-dimensional word vector [30]. PCA is then used to extract the important characteristics to minimize the use of processing resources and increase the training speed. By converting high-dimensional feature vectors to smaller dimensions, PCA can shorten the training period and minimize the amount of memory used by the model. To enhance the performance and generalizability of the classification model, superfluous information is removed from the input features while

retaining the most discriminative characteristics. The essential structure and content of the original data are retained in the features following the dimensionality reduction process implemented via PCA, making the model finding analysis more interpretable and easier to comprehend. The variance contribution ratio of each principal component is displayed in Fig. 5, and the cumulative variance contribution ratio of the first eight principal components surpasses 50%. This indicates that more redundant information is present in the extracted features and that the data retain a high level of information enrichment even after completing dimensionality reduction.
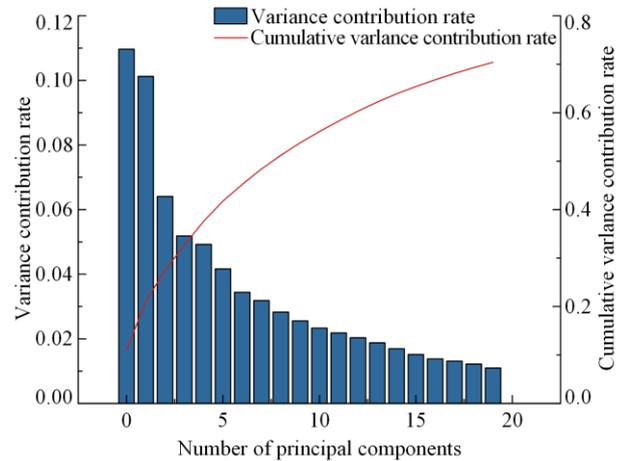

Fig. 5. Variance contribution of each principal component.

C. Classification and Hyperparameter Search

LightGBM has a strong parallel computing ability because of its proprietary GOSS and EFB algorithms, which are very helpful for resolving issues concerning the intelligent diagnosis of distribution network design quality. The benefits and drawbacks of the LightGBM hyperparameter combination, in addition to the model design, have direct impacts on the classification performance of the model. The hyperparameters can be broadly categorized into three groups based on their various effects: regulating the tree structure, enhancing the accuracy of the model, and avoiding overfitting. When choosing a greater depth, it is important to consider whether the model is experiencing overfitting. The max_depth is a common tree structure hyperparameter that controls the depth of the tree. Two significant hyperparameters that impact the model performance are learning_rate and n_estimators. The gradient descent step size can be adjusted with the learning_rate. If it is too large, the model may overshoot the optimal value, while a rate that is too small can slow down convergence. The number of decision trees is controlled by n_estimators. Using the tuning parameter, the model accuracy can be increased by determining the best combination of n_estimators and the learning_rate. Regularization parameters reg_lambda and reg_alpha

are utilized to improve the generalizability of the model while preventing overfitting. The subsample parameter specifies the proportion of data that is randomly sampled for each individual tree, whereas the colsample_bytree parameter regulates the proportion of columns that are randomly selected for each tree. The significance of the hyperparameters for one of the folds is shown in Fig. 6.
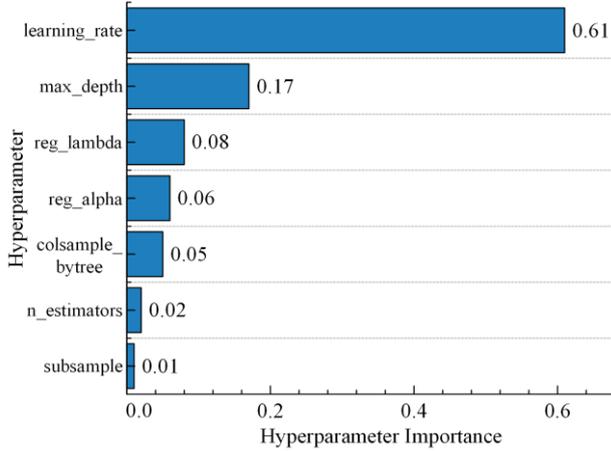

Fig. 6. Importance of the hyperparameters.

D. Results Comparison

An Intel Core i7-9700K processor, 64GB of RAM, and a 1TB SSD hard drive were utilized to ensure our suggested model could process the distribution network project evaluation material.

In this work, useful information is first extracted from distribution network project review text using RoBERTa-wwm. The retrieved structured features are then sent to PCA for feature dimensionality reduction. The output of RoBERTa-wwm is converted into 20-dimensional vectors by PCA from its original 768 dimensions. After performing dimensionality reduction, the primary features are sent to LightGBM for classification. The model performance is then determined by tenfold cross-validation, and the best hyperparameters are found with the aid of the Bayesian approach. A comparison study is carried out from four perspectives, i.e., feature extraction dimensionality, classification dimensionality, combined model dimensionality, and training time, to confirm the efficacy of the model described in this paper.

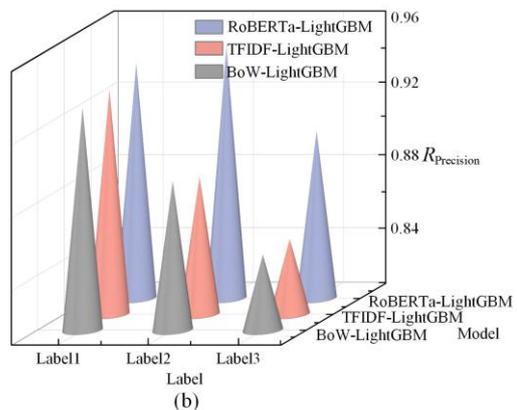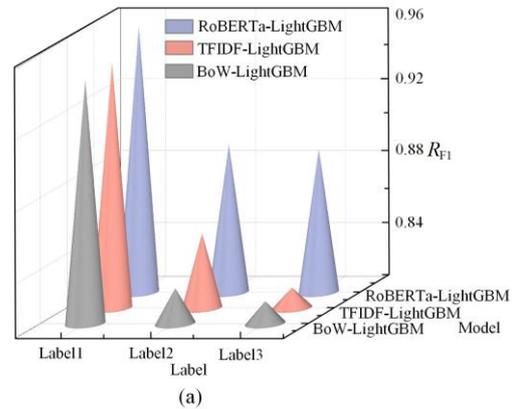### 1) Feature Extraction Dimensionality Comparison Among Different Models

In this research, two common static text embedding models, bag-of-words (BoW) and term frequency-inverse document frequency (TF-IDF), are chosen for comparison with the RoBERTa-wwm model via ablation experiments to validate the proposed model's strong semantic comprehension capacity. To verify the accuracy of the comparative experiment, the vectors output by the BoW and TF-IDF models are downscaled to 20 dimensions via PCA. Utilizing the same Bayesian

hyperparameter optimization method, the LightGBM classification approach is chosen for all three models so that their performances can be compared across five metrics. Table III displays the performance of each model. As seen, the model proposed in this work yields an accuracy decrease of 0.0029 but accuracy increases of 0.0024, 0.0206, and 0.013 over the second-best method in terms of the $R_{Accuracy}$, $R_{Precision}$, R_{Recall}, R_{F1}, and R_{BER} respectively. The advantage of the model developed in this research, which ensures a lower R_{BER} while having a high R_{F1}, is evident in terms of the overall performance.

TABLE III
COMPARISON AMONG THE FEATURE EXTRACTION
DIMENSIONALITY INDICATORS OF DIFFERENT MODELS

Model	$R_{Accuracy}$	$R_{Precision}$	R_{Recall}	R_{F1}	R_{BER}
BoW	0.9025	0.9063	0.8282	0.8558	0.1718
TF-IDF	0.8978	0.8894	0.7963	0.8265	0.2037
RoBERTa	0.8996	0.9269	0.8306	0.8688	0.1694

The performance achieved by each model under various classification labels is displayed in Fig. 7. Label 1 indicates a problem with economic cost adjustment, Label 2 indicates a problem with economic volume accounting, and Label 3 indicates a problem with the technical supporting materials, such as drawings and images. The three-dimensional diagrams shown in this research intuitively show that the models developed in this paper perform optimally on each label and have good generalizability.
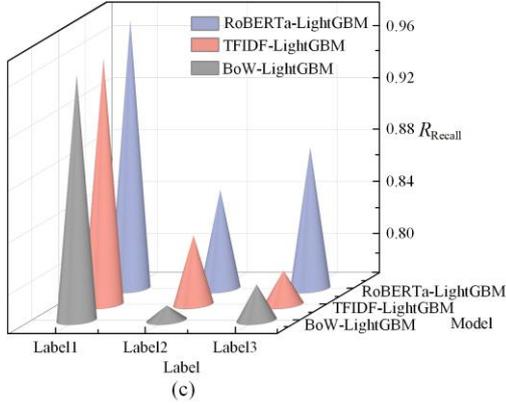

(a)

(b)

Fig. 7. Performance achieved by different feature extraction models for different labels. (a) R_{F1} . (b) $R_{Precision}$. (c) R_{Recall} .

2) Classification Dimensionality Comparison Among Different Models

Six classification models, including logistic regression (LR), decision tree (DT), random forest (RF), gradient boosting decision tree (GBDT), XGBoost, and multilayer perceptron (MLP), are compared in ablation experiments to confirm that the LightGBM model performs best in classification. Except for the classification model, the remaining experimental sets are the same. The feature extraction models are all selected from RoBERTa-wwm, and the same dimensionality reduction and hyperparametric search procedures are used. Figure 8 and Table IV show the comparative outcomes of the classification models. Except for precision, which is ranked second, all the metrics of the proposed method are ranked first. LightGBM performs 0.0038 in terms of precision relative to XGBoost, but achieves an F1 score gain of 0.0093 and lowers the equilibrium error rate by 0.0147. While integrated learning increases the model accuracy by merging several tree models, the tree structure provides stronger interpretability and robustness and is better at collecting the nonlinear properties of samples than classic classification models such as LR. Furthermore, in contrast to the other integrated learning models, such as DT, GBDT, and XGBoost, the special histogram method effectively of LightGBM increases the learning capacity and operating efficiency of the model.

TABLE IV
COMPARISON AMONG THE CLASSIFICATION DIMENSIONALITY
INDICATORS OF DIFFERENT MODELS

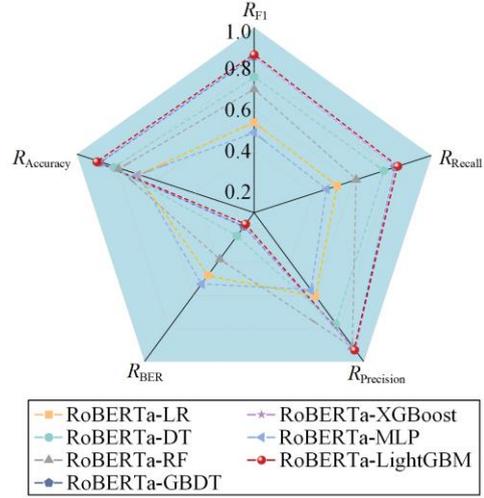| Model	$R_{Accuracy}$	$R_{Precision}$	R_{Recall}	R_{F1}	R_{BER}
LR	0.6913	0.6028	0.5222	0.5390	0.4787
DT	0.8126	0.7722	0.7601	0.7594	0.2399
RF	0.7885	0.9074	0.6189	0.6998	0.3811
GBDT	0.8874	0.9249	0.8174	0.8602	0.1826
XGBoost	0.8881	0.9307	0.8159	0.8595	0.1841
MLP	0.6909	0.5668	0.4698	0.4895	0.5302
LightGBM	0.8996	0.9269	0.8306	0.8688	0.1694

Fig. 8. Radar chart showing the performance of different classification models.

The performances achieved by several tree models on the three labels are displayed in Fig. 9. The three-dimensional conical graphs of the three measures show that LightGBM performs best across the board, whereas GBDT and XGBoost perform similarly but with slightly lower overall performance than that of LightGBM. Furthermore, the three enhanced tree models outperform the basic decision tree structure in terms of the metrics R_{F1}, $R_{Precision}$, and R_{Recall}.
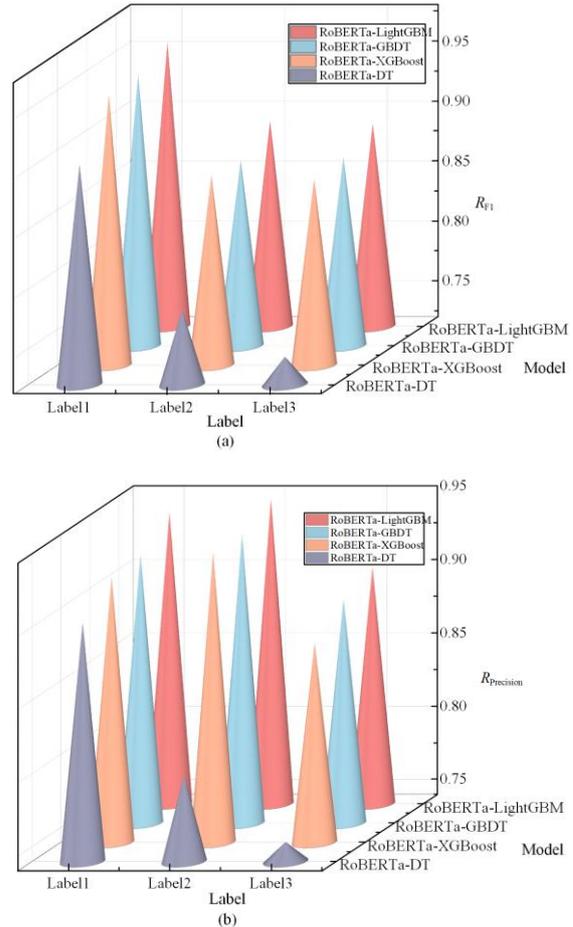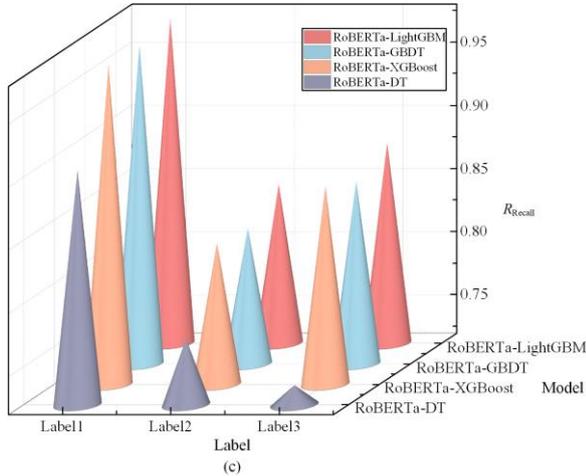
Fig. 9. Performance achieved by different classification models with different labels. (a) R_{F1}. (b) $R_{Precision}$. (c) R_{Recall}.

3) Comparison Among Combined Models

The semantic information extraction and classification models mentioned above are combined to broaden the scope of the comparison. The same dimensionality reduction processing and Bayesian hyperparameter optimization techniques are then carried out to compare the performances of the 21 groups of combined models in terms of five indices. Among these 21 groups of models, the model developed in this work performs best in terms of the R_{Recall}, R_{F1}, and R_{BER}. The precision of the proposed model ranks second, 0.3% lower than that

of the first RoBERTa-wwm/XGBoost combination model, but it improves R_{Recall} by 1.47% and R_{F1} by 0.93%, and decreases R_{BER} by 1.47%. Significant accuracy, R_{Recall}, R_{F1}, and R_{BER} gains are provided by the model used in this work. Its comprehensive performance is high, and its overall performance surpasses that of the other models.

The performances of the 21 combined models are shown in Fig. 10 using R_{BER} and R_{F1} metrics. This work presents the optimal performance of the RoBERTa-wwm and LightGBM combination models, which achieve the lowest equilibrium error rate while preserving accuracy. The BoW-MLP model has the lowest performance in terms of both accuracy and the error. This suggests the presence of a significant discrepancy between the classification capacity of the integrated learning model and that of the standard BoW model and that it is challenging to extract useful features from the former approach. With their strong feature extraction capabilities and gradient boosting classification mechanisms, the model combinations consisting of RoBERTa-wwm and the three major boosting algorithms are found to be the top three performers based on R_{F1} metrics. These models can accurately identify the design problems of distribution network engineering projects, and they all maintain low equilibrium error rates.
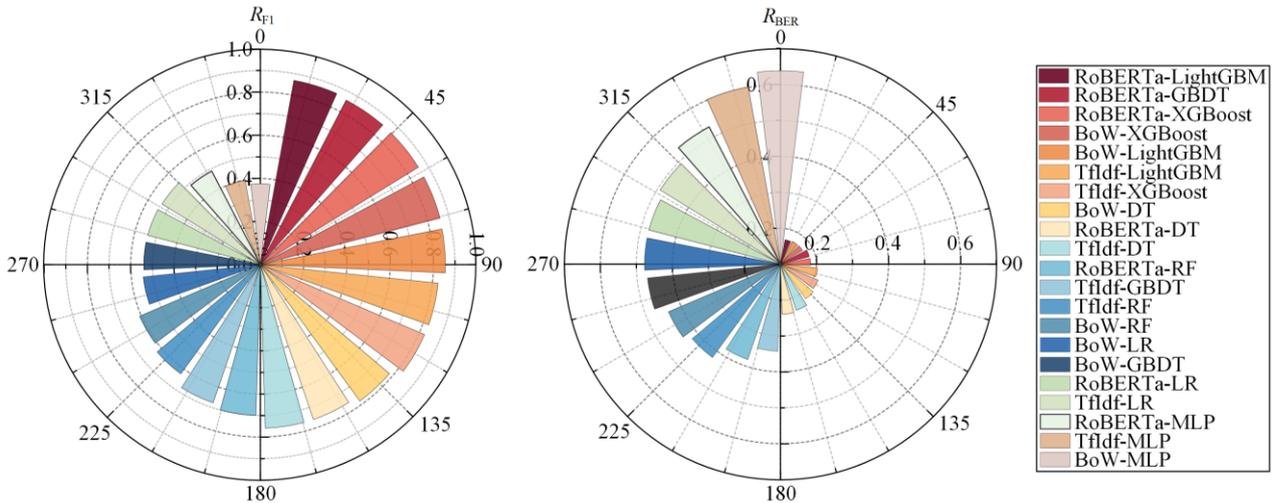

Fig. 10. R_{F1} and R_{BER} performances of the combined models.

4) Speed Comparison

There is no discernible variation in the inference speed of the 21 combined models in terms of their temporal complexity, and all of them are able to meet the real-time criterion. However, compared with the classical methods, our proposed method exhibits a notable advantage in terms of training time. By comparing the R_{F1} and training speed, each combinatorial model's overall performance is assessed. The performance of

each model is displayed in Fig. 11 in terms of their training speeds and R_{F1}. It is evident from Fig. 11 that the model used in this paper has the highest R_{F1}, indicating that the transformer structure and full-word masking operation of the RoBERTa-wwm model are able to learn the text features more thoroughly and that the optimized hyperparameters of LightGBM perform better than those of the other models in terms of classification. Additionally, the gradient boosting-based tree

model performs significantly better with respect to the R_{F1} among the classification models.

LR has the worst training time performance. However, the calculation times of DT and MLP are faster because of their smaller model structures, and their classification performances are worse. The model presented in this paper increases the R_{F1} by 0.1657, while its training time is closer to that of RF. The model developed in this paper has a training time of 8.023 seconds, which is 62.623 seconds shorter than that of XGBoost and much shorter than that of other gradient boosting models, with guaranteed accuracy. This indicates that the designs of the GOSS and EFB algorithms of LightGBM attains improved training efficiency. In conclusion, the model designed in this paper is suitable for multiclassification scenarios with high training and accuracy requirements and has significant advantages in classifying textual problems involving power distribution network engineering reviews.
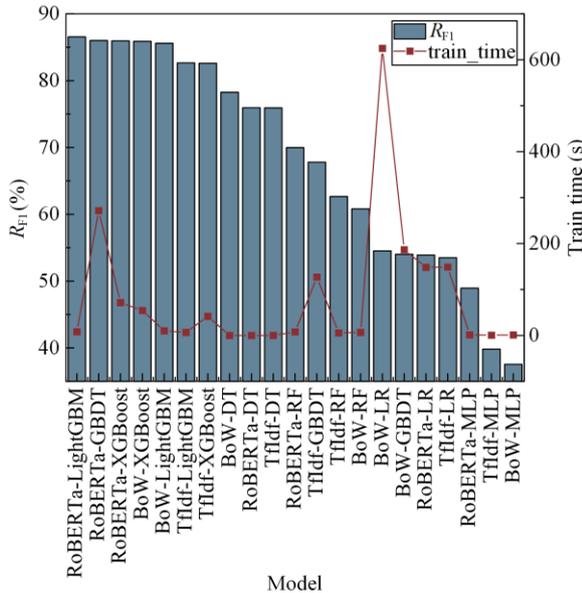

Fig. 11. Training speed and accuracy of each model.

V. Conclusion

This research suggests an intelligent diagnosis approach based on integrated learning and an LLM for addressing typical distribution network engineering issues. The methodology developed in this research ensures the quality of a distribution network engineering design by intelligently diagnosing issues through the extraction of semantic information from expert review text. First, features are extracted from the expert review text of distribution network engineering projects using RoBERTa-wwm. Second, to preserve the principal components with significant variance contributions, the RoBERTa-wwm output is subjected to feature dimensionality reduction via PCA. Ultimately, the LightGBM classification model receives this information. Ten folds

are created from the input dataset, the validation set undergoes a Bayesian hyperparametric search, and the test set yields the average performance achieved across the ten folds. Additionally, the tested approaches are compared and analyzed from four perspectives: training speed, model combination, feature extraction, and classification. The following deductions are made.

1) To extract the textual aspects of professional vocabulary in the field of power distribution network engineering, in this research, we adopt RoBERTa-wwm and fully utilize the powerful comprehension capacity of an LLM for processing large-scale corpora. With respect to the metrics R_{BER}, $R_{Precision}$, R_{Recall}, and R_{F1}, the RoBERTa-wwm model performs best according to a comparative analysis conducted based on feature extraction dimensions. Effective text features are provided for the downstream classification task by the RoBERTa-wwm model.

2) This research analyzes the solutions produced by seven classification models for a common problem in distribution network engineering to validate the efficacy of LightGBM in the proposed intelligent diagnosis approach. Experiments are carried out under the premise of ensuring the consistency of basic conditions such as the feature extraction method, feature dimensionality, number of iterations, and Bayesian hyperparameter search. According to the findings, LightGBM performs best in terms of the metrics R_{BER} across all four performance criteria. This study presents a model that ensures correctness when parsing and categorizing large-volume distribution network problems, provides a more effective parallel training procedure, processes vast amounts of data quickly, and exhibits significantly increased operating efficiency.

3) Twenty-one combined models, which are created by combining the feature extraction and classification models presented with respect to the first two comparison dimensions, are used for comparison experiments. The performance of the combined model presented in this study ranks first in terms of every metric, and a noticeable improvement in the performance of the model is observed with respect to R_{BER}, R_{Recall}, and R_{F1}. When the textual feature capture ability of RoBERTa-wwm, the downscaling strength of PCA, and the classification capability of Bayes-LightGBM are applied together, the overall performance of the model can be enhanced to conduct more precise intelligent diagnoses of distribution network engineering issues.

4) The model in this research has an ideal R_{F1}, which considers both $R_{Precision}$ and R_{Recall}, based on a comparison among the training durations and R_{F1} of the 21 combined models. In terms of classification performance, all of the gradient boosting-based tree models outperform the other methods, although LightGBM

does so considerably better. The model developed in this study ensures precision while significantly reducing the computational time required by previous gradient boosting methods, particularly in terms of the training speed.

5) To provide a common problem library, this paper analyzes and organizes the issues encountered in distribution network engineering projects. This approach offers a fresh perspective on how to handle design flaws in distribution network engineering projects. With the intelligent assisted review tool developed in this study, reviewers can rapidly find and address design flaws, decreasing the required review time and increasing productivity. Reducing the chance of accidents, increasing the dependability of the power system, and ensuring a steady supply of electricity to the grid are all possible by effectively controlling quality faults. Additionally, this method can reduce financial losses caused by poor distribution network engineering designs, enhance the quality of engineering work, and further advance social and economic advancement by increasing evaluation efficiency and enabling accurate, timely issue diagnosis.

Both model optimization and application expansion will be the focus of future research. More sophisticated language models can be created in the future to increase the computational efficiency and lower the memory occupation rate. At the model level, the model created in this study still consumes significant computing resources, while at the application level, the main focus of this paper is text mining for the evaluation opinions that are generated during the distribution network project evaluation process. Future work can broaden the scope of text information mining, extract information from text materials such as the preliminary design and feasibility study reports, and conduct a more thorough assessment of the quality of distribution network projects.

DECLARATIONS

Competing interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

AUTHORS' INFORMATION

Wei Hu received a Ph.D. degree in management science and engineering from Shanghai Jiaotong University, Shanghai, China. He is now a professor and the deputy dean of the School of Economics and Management at Shanghai University of Electric Power. His main research interests include artificial intelligence, smart energy, power engineering management, energy and power optimisation and decision making.

Tingting Zheng is currently pursuing her master's degree at Shanghai University of Electric Power. Her main research interests include large language modelling, power engineering management, and smart energy.

Puliang Du received a Ph.D. degree in electrical engineering from Southeast University, Nanjing, China. He is currently teaching at the School of Economics and Management, Shanghai University of Electric Power. His main research interests are the application of decision theory and methods in power systems, the optimal operation of distribution networks, and artificial intelligence.

Zhiwei Chen is currently pursuing a Ph.D. at Shanghai University of Finance and Economics. His primary research interests are in the areas of natural language processing, information technology, and smart energy.

REFERENCES

[1] R. Li, P. Wong, and K. Wang *et al.*, "Power quality enhancement and engineering application with high permeability distributed photovoltaic access to low-voltage distribution networks in Australia," *Protection and Control of Modern Power Systems*, vol. 5, no. 3, pp. 1-7, Jul. 2020.

[2] L. Cheng, T. Wang, and Y. Wang, "A novel fault location method for distribution networks with distributed generations based on the time matrix of traveling-waves," *Protection and Control of Modern Power Systems*, vol. 7, no. 1, pp. 46, Jan. 2022.

[3] U. U. Uma, D. Nmadu, and N. Ugwuanyi *et al.*, "Adaptive overcurrent protection scheme coordination in presence of distributed generation using radial basis neural network," *Protection and Control of Modern Power Systems*, vol. 8, no. 4, pp. 1-19, Oct. 2023.

[4] Y. Wang, J. Gu, and L. Yuan, "Distribution network state estimation based on attention-enhanced recurrent neural network pseudo-measurement modeling," *Protection and Control of Modern Power Systems*, vol. 8, no. 2, pp. 1-16, Apr. 2023.

[5] Y. Wang, T. Wang, and L. Liu, "A fault segment location method for distribution networks based on spiking neural P systems and Bayesian estimation," *Protection and Control of Modern Power Systems*, vol. 8, no. 3, pp. 1-12, Jul. 2023.

[6] X. Sun, J. Chen, and H. Zhao *et al.*, "Sequential disaster recovery strategy for resilient distribution network based on cyber-physical collaborative optimization," *IEEE Transactions on Smart Grid*, vol. 14, no. 2, pp. 1173-1187, Mar. 2023.

[7] T. Jain, D. Ghosh, and D. K. Mohanta, "Augmentation of situational awareness by fault passage indicators in distribution network incorporating network reconfiguration," *Protection and Control of Modern Power Systems*, vol. 4, no. 4, pp. 1-14, Oct. 2019.

[8] Y. Xiao, J. Ouyang, and X. Xiong *et al.*, "Fault protection method of single-phase break for distribution network considering the influence of neutral grounding modes," *Protection and Control of Modern Power Systems*, vol. 5, no. 2, pp. 1-13, Apr. 2020.

[9] Y. Gao, W. Wang, and J. Shi *et al.*, "Batch-constrained reinforcement learning for dynamic distribution network reconfiguration," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5357-5369, Nov. 2020.

[10] W. Chen, S. Zhuang, and J. Lin *et al.*, "Application of data mining in the settlement management of distribution network projects," *Security and Communication Networks*, vol. 2022, no.1, Feb. 2022.

[11] Y. Sha, W. Li, and J. Yan *et al.*, "Research on investment scale calculation and accurate management of power grid projects based on three-level strategy," *IEEE Access*, vol. 9, pp. 67176-67185, May 2021.

[12] Y. Zhong, Y. Chen, and Y. Zhou *et al.*, "The artificial intelligence large language models and neuropsychiatry practice and research ethic," *Asian journal of psychiatry*, vol. 84, Jun. 2023.

[13] R. S. Bonadia, F. C. L. Trindade, and W. Freitas *et al.*, "On the potential of ChatGPT to generate distribution systems for load flow studies using OpenDSS," *IEEE Transactions on Power Systems*, vol. 38, no. 6, pp. 5965-5968, Nov. 2023.

[14] J. Li, S. Qian, and L. Li *et al.*, "A novel secondary decomposition method for forecasting crude oil price with twitter sentiment," *Energy*, vol. 290, Mar. 2024.

[15] Q. Meng, Y. Song, and J. Mu *et al.*, "Electric power audit text classification with multi-grained pre-trained language model," *IEEE Access*, vol. 11, pp. 13510-13518, Jan. 2023.

[16] Z. Tan, Y. Chen, and Z. Liang *et al.*, "Named entity recognition for few-shot power dispatch based on multi-task," *Electronics*, vol. 12, no. 16, Aug. 2023.

[17] Y. Chen, D. Lin, and Q. Meng *et al.*, "Named entity identification in the power dispatch domain based on RoBERTa-Attention-FL model," *Energies*, vol. 16, no. 12, Jun. 2023.

[18] C. Wan, M. Niu, and Y. Song *et al.*, "Pareto optimal prediction intervals of electricity price," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 817-819, Jan. 2017.

[19] R. Otto, S. Susanne, and H. Jouni *et al.*, "Crown snow load outage risk model for overhead lines," *Applied Energy*, vol. 343, Aug. 2023.

[20] L. Ye, B. Dai, and Z. Li *et al.*, "An ensemble method for short-term wind power prediction considering error correction strategy," *Applied Energy*, vol. 322, Sep. 2022.

[21] J. Runge and E. Saloux, "A comparison of prediction and forecasting artificial intelligence models to estimate the future energy demand in a district heating system," *Energy*, vol. 269, Apr. 2023.

[22] E. Genov, C. D. Cauwer, and G. V. Kriekinge *et al.*, "Forecasting flexibility of charging of electric vehicles: Tree and cluster-based methods," *Applied Energy*, vol. 353, Jan. 2024.

[23] N. A. Semary, W. Ahmed, and K. Amin *et al.*, "Improving sentiment classification using a RoBERTa-based hybrid model," *Frontiers in human neuroscience*, vol. 17, Dec. 2023.

[24] J. Bed and D. Toshniwal, "CitEnergy: a BERT based model to analyse citizens' energy-tweets," *Sustainable Cities and Society*, vol. 80, May 2022.

[25] J. Nyangon and R. Akintunde, "Principal component analysis of day-ahead electricity price forecasting in CAISO and its implications for highly integrated renewable energy markets," *Wiley Interdisciplinary Reviews: Energy and Environment*, vol. 13, no. 1, Nov. 2023.

[26] J. Park and E. Hwang, "A two-stage multistep-ahead electricity load forecasting scheme based on LightGBM and attention-BiLSTM," *Sensors*, vol. 21, no. 22, pp. 7697, Nov. 2021.

[27] J. Liang, "Estimation of stellar atmospheric parameters with light gradient boosting machine algorithm and principal component analysis," *The Astronomical Journal*, vol. 163, no. 4, pp. 153, Apr. 2022.

[28] H. Fang, J. Xiao, and Y. Wang, "A machine learning-based detection framework against intermittent electricity theft attack," *International Journal of Electrical Power & Energy Systems*, vol. 150, Aug. 2023.

[29] L. You, F. Han, and J. Peng *et al.*, "ASK-RoBERTa: a pretraining model for aspect-based sentiment classification via sentiment knowledge mining," *Knowledge-Based Systems*, vol. 253, Oct. 2022.

[30] J. Liang, D. Li, and Y. Lin *et al.*, "Named entity recognition of chinese crop diseases and pests based on RoBERTa-wwm with adversarial training," *Agronomy-Basel*, vol. 13, no. 3, Mar. 2023.