http://dx.doi.org/10.62762/TETAI.2024.532253

RESEARCH ARTICLE

Short and Long-Term Renewable Electricity Demand Forecasting Based on CNN-Bi-GRU Model

Shuchen Zhao[1], **Zhongjin Xu**[2], **Zhefan Zhu**[3], **Xiaoxiang Liang**[4], **Zecheng Zhang**[5] and **Ruxue Jiang**[6,*]

[1] Department of Electrical & Computer Engineering, Duke University, Durham, NC 27708, United States
[2] University of Michigan-Dearborn, Dearborn, MI 48128, United States
[3] Carnegie Mellon University, Pittsburgh, PA 15213, United States
[4] Washington University in St. Louis, St. Louis, MO 63130, United States
[5] New York University, Brooklyn, NY 11201, United States
[6] Northeastern University, Seattle, WA 98109, United States

Abstract

With the increasing global focus on renewable energy and the growing proportion of renewable power in the energy mix, accurate forecasting of renewable power demand has become crucial. This study addresses this challenge by proposing a multimodal information fusion approach that integrates time series data and textual data to leverage complementary information from heterogeneous sources. We develop a hybrid predictive model combining CNN and Bi-GRU architectures. First, time series data (e.g., historical power generation) and textual data (e.g., policy documents) are preprocessed through normalization and tokenization. Next, CNNs extract spatial features from both data modalities, which are fused via concatenation. The fused features are then fed into a Bi-GRU network to capture temporal dependencies, ultimately forming a robust CNN-Bi-GRU model. Comparative experiments with ARIMA, standalone GRU, and EEMD-ARIMA (a hybrid model combining ensemble empirical mode decomposition with ARIMA) demonstrate the superiority of our approach in both short- and long-term forecasting tasks on the same dataset. This research offers a potential framework to enhance renewable power demand prediction, supporting the industry's sustainable growth and practical applications.

Keywords: multimodal information fusion, renewable electricity demand forecasting, CNN, Bi-GRU, predictive performance.

1 Introduction

In recent years, there has been a growing global emphasis on sustainable and renewable energy, driven by concerns over climate change and the finite nature of fossil fuels. Renewable electricity, a key component of the energy transition, continues to increase its

Submitted: 23 December 2024
Accepted: 10 February 2025
Published: 16 February 2025

Vol. 2, **No.** 1, 2025.
10.62762/TETAI.2024.532253

***Corresponding author:**
✉ Ruxue Jiang
snow0911@outlook.com

Citation
Zhao, S., Xu, Z., Zhu, Z., Liang, X., Zhang, Z., & Jiang, R. (2025). Short and Long-Term Renewable Electricity Demand Forecasting Based on CNN-Bi-GRU Model. *ICCK Transactions on Emerging Topics in Artificial Intelligence*, 2(1), 1–15.

© 2025 by the Authors. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

share in global power generation. As this transition unfolds, precise forecasting of renewable energy demand becomes essential. Accurate prediction of electricity demand is crucial in today's society, as it not only supports the achievement of sustainable energy development goals but also provides valuable insights for the transformation of electricity markets and the future of energy needs [1, 2]. By combining precise forecasts of electricity demand with renewable energy supply, new decision-making strategies can be proposed for global system planning. Therefore, developing an effective electricity demand forecasting framework is imperative.

In traditional electricity consumption forecasting methods, regression analysis and time series models are the most commonly used techniques. Bianco et al. [3] applied linear regression models to predict electricity consumption in Italy, while Arisoy et al. [4] used historical time series data to forecast residential electricity demand. Several other studies [5–7] employed ARIMA models and seasonal ARIMA models to predict electricity loads. Mitkov et al. [8] developed prediction models for future energy demand using ARIMA models based on the linear characteristics of historical electricity consumption data. However, econometric models tend to perform poorly in predicting nonlinear sequences due to their complexity [9]. Amjady [10] proposed an enhanced ARIMA model for short-term electric load forecasting, which incorporates both temperature and electric load data to predict hourly loads and daily peak electricity demand. Feinberg and Genethliou [11] integrated temporal factors, weather data, and customer segmentation in their short-term electric load forecasts. Despite these advances, linear regression models face significant challenges in capturing nonlinear relationships, especially when the demand for electricity and its influencing factors exhibit nonlinear dynamics. This makes it difficult for linear regression models to accurately capture nonlinear load demand patterns across different time periods [12].

With the advancement of artificial neural networks, the ability to address nonlinear problems has significantly improved [13, 14]. This improvement allows solutions to overcome the limitations of linear regression models in handling nonlinear relationships [15, 16]. As a result, some researchers have started combining regression models with artificial intelligence in demand forecasting [17]. For instance, Valenzuela et al. [18] integrated the nonlinear mapping capability of artificial neural networks with time series models, proposing an ARIMA-ANN model. In this approach, a fuzzy logic expert system was used to determine the differencing order of the ARIMA model, as well as the orders of the autoregressive (AR) and moving average (MA) components. Liu et al. [19] applied discrete wavelet transform to decompose renewable energy time series data and fed the decomposed components into an LSTM model, demonstrating that combining signal decomposition preprocessing with deep learning substantially improves short-term wind power prediction accuracy. Nie et al. [20] introduced a novel hybrid model combining data preprocessing, individual forecasting algorithms, and weight determination techniques, demonstrating higher accuracy and improved forecasting performance. However, renewable energy power systems are complex and dynamic, influenced by various factors such as weather, technological advancements, policy changes, and social trends [21]. These models may struggle when dealing with non-time-series electricity demand data, suggesting the need for further research into more advanced predictive models.

Multimodal information fusion refers to the process of collecting and integrating information from diverse data sources—such as historical load data, meteorological observations, and policy text—to improve forecasting accuracy and decision-making reliability [22, 23]. Recently, an increasing number of scholars have focused on applying multimodal fusion to electricity demand forecasting. For example, Xuan et al. [24] employed a feature selection algorithm based on random forests to identify key input features for load forecasting models. After selecting the input features, they proposed a hybrid neural network algorithm for short-term load forecasting, which leverages multimodal fusion. Ji et al. [25] developed an advanced multimodal fusion model with enhanced predictive performance by combining the outputs of three models—Gradient Boosting Decision Trees, Extreme Gradient Boosting, and Light Gradient Boosting Machine—through decision fusion. To capture more comprehensive temporal and spatial features from the original data, Kong et al. [26] proposed an integrated approach for short-term load forecasting, combining Empirical Mode Decomposition (EMD), the similar day method, and deep neural networks. Within this framework, a CNN-LSTM model based on EMD is used to extract multimodal spatiotemporal features. These multimodal fusion models not only increase the

complexity and dimensionality of the forecasting models but also improve prediction accuracy and reliability by capturing a broader range of influencing factors. However, there is currently no literature on how to effectively select the modality that best represents key factors influencing power demand, such as weather conditions, economic activities, political events, and social behavior patterns.

- Initially, we conduct an analysis of data relevant to short and long-term electricity demand. The data is categorized into short and long-term forecasting data. Subsequently, CNNs are employed to extract two types of latent features pertinent to the research objective from numerical time series data and textual data.
- Regarding the fusion methodology, we adopt a concatenation fusion approach to organically merge the temporal latent features and textual latent features. The fused features are then fed into a model constructed using Bi-GRU for prediction.
- Ultimately, the constructed short-term and long-term demand forecasting models are validated through experimentation, demonstrating their superior predictive performance compared to the ARIMA model, single GRU network, and combination model (EEMD-ARIMA).

2 Related Works

2.1 Forecasting-related Definitions

Prediction refers to estimating future trends using existing data or past experiences. Renewable energy demand forecasting involves multiple influencing factors, including energy consumption patterns, policy changes, technological advancements, climate conditions, and economic growth. Its primary objective is to assist governments, energy enterprises, and market decision-makers in making scientifically sound plans, improving energy supply chain management efficiency, optimizing grid dispatching, and promoting sustainable energy development. Based on the forecasting time horizon, renewable energy demand forecasting can be categorized into short-term, medium-term, and long-term predictions.

(1) Short-term renewable energy demand forecasting typically covers a few days, weeks, or months into the future. It primarily relies on recent weather data, grid load conditions, and energy consumption patterns. It is widely used in power dispatching, demand response management, and short-term energy market transactions. For example, given the high variability of wind and solar energy, short-term forecasting helps grid operators efficiently allocate energy storage resources and optimize the balance between electricity supply and demand.

(2) Medium-term renewable energy demand forecasting involves predicting energy demand trends over the coming months to several years. Compared to short-term forecasting, medium-term prediction requires considering a broader range of factors, such as economic development, social changes, policy adjustments, and industrial structure upgrades. For instance, governments can utilize medium-term forecasting results to design renewable energy subsidy policies, optimize energy infrastructure construction, and facilitate sustainable energy transitions.

(3) Long-term renewable energy demand forecasting spans several years to decades and involves more complex factors, including technological progress, global climate change, energy market transformations, and carbon neutrality goals. Although long-term forecasting entails greater uncertainty and risks, it is crucial for formulating national energy strategies, investing in new energy projects, and establishing carbon reduction plans. For example, energy companies can use long-term predictions to plan the development and investment of wind and solar energy projects, aligning with future energy structure adjustments.

2.2 Forecasting Methods

Traditional forecasting methods: In the research of electricity demand forecasting, traditional methods primarily rely on statistical analysis and time series modeling. Regression analysis is a common statistical approach; for instance, Bianco et al. [3] utilized a linear regression model to predict electricity consumption in Italy. Additionally, time series methods, such as ARIMA and its improved versions, have been widely applied in electricity load forecasting. For example, some employed ARIMA and seasonal ARIMA models [5–7] to forecast electricity load, respectively. Furthermore, Mitkov extended the use of the ARIMA model by incorporating the linear characteristics of historical electricity consumption data to establish an energy demand forecasting model [8]. However, traditional models often underperform when dealing with complex nonlinear features, especially when electricity demand is influenced by

multiple factors such as weather, policy changes, and economic fluctuations. Statistical models frequently fail to accurately capture the intricate dynamic patterns in the data, leading to diminished forecasting performance [9].

Deep learning prediction methods: ANN with their powerful nonlinear modeling capabilities, can effectively address the limitations of traditional methods in handling complex data patterns [13]. For instance, Valenzuela et al. [18] proposed an ARIMA-ANN hybrid model that integrates a fuzzy logic expert system to optimize the parameters of the ARIMA model, thereby improving prediction accuracy. Additionally, Liu et al. [19] employed discrete wavelet transform to preprocess renewable energy time series data and fed the decomposed signals into an LSTM network, constructing an integrated forecasting model that enhances short-term wind power prediction accuracy by combining signal decomposition with deep temporal modeling. Nie et al. [20] utilized a multi-objective grey wolf optimization algorithm to combine radial basis function networks, generalized regression neural networks, and extreme learning machines, further improving prediction performance. These studies demonstrate that integrating traditional statistical methods with deep learning techniques can uncover trends in electricity demand from time series data, significantly enhance the accuracy of electricity demand forecasting, and improve adaptability to complex load variations. However, although combining regression models with deep learning models can effectively leverage the strengths of each approach to improve prediction accuracy, renewable energy power systems remain inherently complex and dynamic. These systems are influenced by multiple factors such as weather patterns, technological advancements, policy changes, and social trends, which are often difficult to fully capture using single time series data alone.

Multimodal fusion methods: Multimodal information fusion involves integrating heterogeneous data sources to improve load forecasting accuracy and reliability [22], prediction [23], recognition, and classification tasks. In recent years, this approach has garnered increasing attention in electricity demand forecasting. For instance, Xuan et al. [24] employed a random forest-based feature selection algorithm to optimize input features for load forecasting models and proposed a hybrid neural network-based method for short-term load forecasting using multimodal fusion. Ji et al. [25] applied decision fusion by combining GBDT, XGBoost, and LightGBM to develop a multimodal fusion model with enhanced predictive performance. Furthermore, Kong et al. [26] introduced a short-term load forecasting method that integrates Empirical Mode Decomposition (EMD), the similar-day method, and deep neural networks. This study utilized an EMD-based CNN-LSTM approach to extract multimodal spatiotemporal features from the input data, improving prediction accuracy.

Despite significant progress in multimodal fusion research, several key challenges remain. For instance, effectively selecting key modalities to comprehensively represent core factors influencing electricity demand—such as weather conditions, economic activities, policy changes, and social behavior patterns—remains an unresolved issue. Additionally, traditional deep learning models still face limitations in handling long-sequence data, making it difficult to capture long-term dependencies fully. To address these challenges, this paper proposes a multimodal fusion-based forecasting model using CNN-Bi-GRU. This model leverages CNN's feature extraction capabilities and Bi-GRU's bidirectional learning ability in processing time-series data, enabling more efficient capture of spatiotemporal characteristics in electricity demand and enhancing prediction accuracy and stability.

3 Methodology

In order to combine the temporal information of time series data with the nature information of textual data, we propose a novel ensemble prediction modeling approach based on multimodal information fusion for quantitative analysis and forecasting of electricity demand and renewable energy supply. The fundamental principle of the prediction model lies in the concept of multimodal information fusion-driven modeling, where the complementary nature of time series data and textual data forms the foundational assumption for building effective models. The framework of this model is illustrated in Figure 1, consisting of four main steps:

(1) Data Preprocessing: The preprocessing of both modalities of data is conducted separately. Time series data is transformed into one-dimensional format using time window techniques, while textual data undergoes structured processing using the Word2Vec model, capable of extracting semantic features of text.

(2) Feature Extraction: Utilizing CNNs, we extract the hidden features of the processed time series and

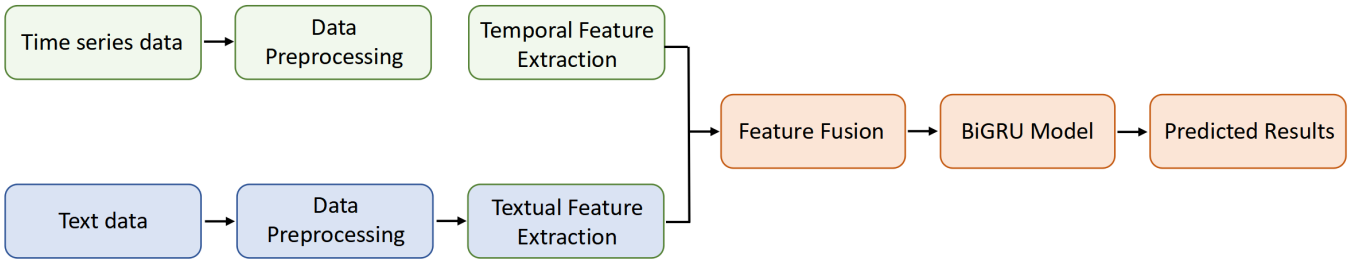


Figure 1. Prediction scheme based on multi-modal information fusion.

textual data, and then combine the features extracted from both data modalities as input for the prediction module.

(3) Feature Fusion: Considering the heterogeneity of the original data, this paper employs vector concatenation to merge and integrate the hidden features of the two modalities.

(4) Prediction Module: Model training is conducted using Bi-GRU, which exhibit favorable properties for processing sequential data.

It is worth noting that due to the differences between mid-term and long-term predictions in time series and textual data, this paper adopts the same model framework to train corresponding mid-to-long-term electricity demand prediction models using different time series and textual data.

3.1 Data Preprocessing

In most cases, the raw data we acquire tends to be messy and incomplete, making it challenging for machine learning models to extract meaningful information. To ensure that the data fed into the input feature extraction module meets the requirements, some preprocessing of the raw data is necessary.

Time series data preprocessing: Our process begins with an initial cleanup of the raw time series data, starting with the filling of missing values to ensure data completeness and continuity. This step is crucial for maintaining data quality and for subsequent analyses. Following this, we employ the sliding window technique for further processing. The sliding window technique moves a fixed-length window across the time series, extracting segments of data from each window. This transforms continuous data segments into a series of independent observations, making the processed data more suitable for input into CNN. This approach not only aids in capturing local features and trends within the time series data but also facilitates the model's learning of temporal dependencies, laying the groundwork for subsequent deep learning model training and prediction. Finally, to map the data within a fixed range, we utilize the method of min-max normalization to process the input data. The min-max normalization formula is expressed as follows:

$$\hat{x}_t = \frac{x_t - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

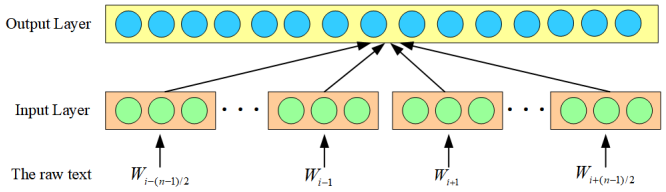


Figure 2. The Word2Vec model.

Textual data preprocessing: The raw textual data is considered unstructured data and requires data preprocessing before feature extraction. During the process of converting text data, it's essential to preserve as much useful information as possible. Basic preprocessing of textual data includes steps like tokenization and stop word removal. Subsequently, the Word2Vec model [27] is utilized for text representation, also known as text vectorization (Word Embedding). This technique enables quantitative measurement of the relationships between words, facilitating the exploration of connections between them. Text vectorization eliminates the cumbersome task of feature processing, maximally preserving the essential content within the text. The model's network structure comprises three layers: the input layer, the projection layer, and the output layer, as illustrated in Figure 2. The purpose of training this model is to maximize the value expressed as follows:

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\le j\le c}\log p(w_{t+j}|w_t) \tag{2}$$

where $c > 0$ represents the window size, and the

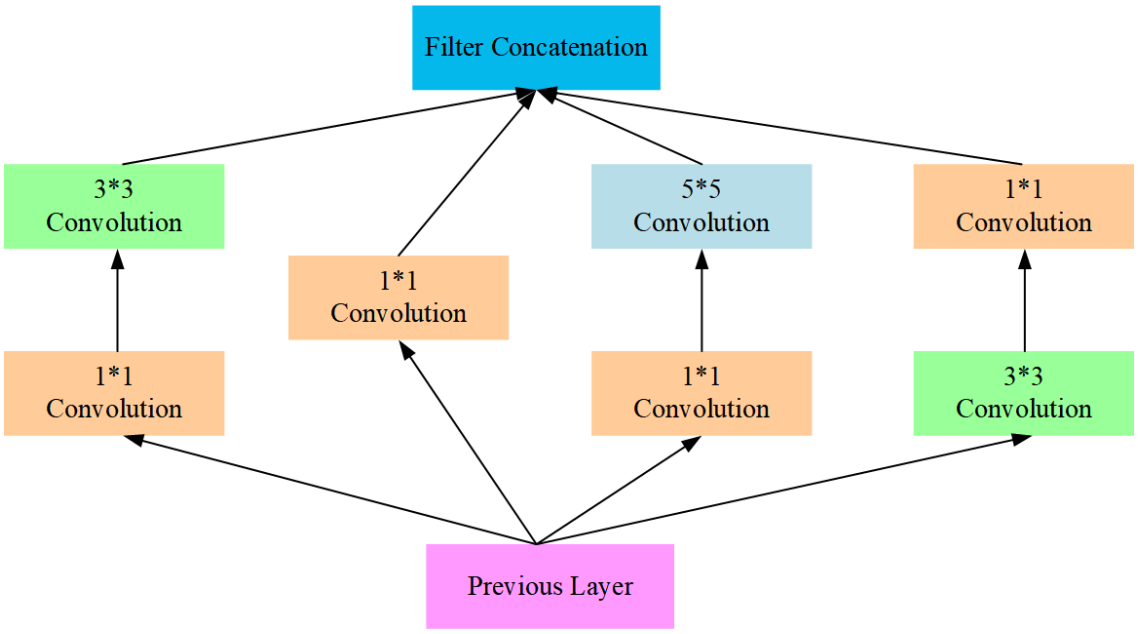


Figure 3. Inception structure diagram.

conditional probability $p(w_{t+j}|w_t)$ is designed as:

$$P(w_0|w_t) = \frac{\exp\left(e'(w)^T e(w_t)\right)}{\sum_{w=1}^{|v|} \exp\left(e'(w)^T e(w_t)\right)} \tag{3}$$

However, Word2Vec has limitations, such as its inability to fully capture contextual nuances and polysemy in textual data. To address this, alternative or complementary methods such as Bidirectional Encoder Representations from Transformers (BERT) or GloVe could be explored to enhance the representation of textual information and mitigate potential biases.

3.2 Feature Extraction

This section focuses on exploring how to perform feature extraction from preprocessed data, a key aspect of the method proposed in this paper. The input data consists of two distinct datasets, comprising time series data and related textual data. We aim to extract data features from these two types of data that are suitable for constructing a Bi-GRU prediction model.

Feature extraction of time series data: The Inception architecture is an innovative convolutional neural network design aimed at capturing multi-scale image features by paralleling convolutions of varying sizes, while employing convolutions for dimensionality reduction to enhance computational efficiency and minimize overfitting [28], as depicted in Figure 3. This architecture has undergone several iterations, incorporating optimizations such as batch normalization and residual connections, showcasing exceptional performance and efficient computation in tasks like image recognition and classification. The entire process of using CNNs to extract features from time series data is outlined as follows:

First, the convolutional layer extracts features from the input data.

$$c_i = f\left(W \otimes X_{i:i+h-1} + b\right) \tag{4}$$

where W, b, h denote the convolution kernel parameters.

Second, in order to accelerate the convergence speed, the ReLU function is added as an activation function after the convolutional layer. The formula is as follows:

$$f(x) = \max(x, 0) \tag{5}$$

Third, after the convolutional layer, a pooling layer is introduced to reduce the dimensionality of the feature maps and prevent the curse of dimensionality. The max-pooling strategy is applied to perform pooling operations on the feature maps obtained after the convolution operation, effectively reducing the number of parameters in the model while retaining useful information. This pooling operation is a common technique used to reduce the complexity of computations and improve the performance of the model by reducing the dimensionality of the input feature maps. The formula is as follows:

$$z_i = \max\{c_j\} \tag{6}$$

where c_i represents the feature map.

Finally, add a flattening layer to flatten the pooled feature maps into one-dimensional form, forming the final latent feature for the time series data.

Text data feature extraction: Traditional text feature extraction methods face challenges with high dimensionality and inefficiency. This study adopts a CNN-based approach proposed by Er et al. [29], which not only excels in feature extraction but also aligns structurally with previously extracted time series features. The text feature extraction process, illustrated in Figure 4, utilizes convolution and pooling as core mechanisms. These operations process fixed-length text sequences to distill primary features, which are then transformed into advanced features through further convolution and pooling. After being processed by a flattening layer, these features are directly output as the text features (TF) required for this study.

3.3 Feature Fusion

Having extracted latent features from both time series and textual modalities via the CNN-based Inception architecture, this section describes how these heterogeneous features are combined for downstream prediction.

This paper adopts the feature-level fusion strategy to build the prediction model, merging the latent features obtained from the two modalities through multimodal fusion methods. For feature fusion, we choose the direct concatenation of feature vectors to preserve the original data information as much as possible. In the linear fusion method based on manual rules, weight values need to be manually set to adjust the proportion of feature vectors, which may introduce subjectivity and limitations. The similarity matrix approach requires computing the similarity matrix between feature vectors, which may encounter complexity issues in matrix calculations. On the other hand, the direct concatenation of feature vectors is relatively simple and straightforward, as it only involves concatenating the feature vectors together. Especially in the case of heterogeneous data, the direct concatenation method does not require additional calculations and transformations, making it more convenient and efficient. In summary, when performing feature fusion, the most suitable fusion method should be selected based on the specific situation to maximize the extraction and preservation of data information.

Vector concatenation, or concatenation fusion as described by Yang et al. [30], involves joining latent features from time series and text data end-to-end for feature fusion, depicted in Figure 5. SF and TF denote latent features from time series and text, with elements like being vectors $[y_1, y_2, \ldots, y_n]$. This process leverages DenseNet's [31] concatenate operation to merge and aggregate flattened vectors post-extraction.

In this study, vector concatenation is adopted as the fusion method for textual and time series data, primarily due to its computational efficiency, interpretability, and suitability for electricity demand forecasting. Electricity demand forecasting involves time series data (such as historical load and meteorological factors) and textual data (such as policy announcements and news reports), which have distinct feature representations. Directly mapping them into the same space may result in information loss or distribution mismatches. Vector concatenation enables efficient information fusion while preserving the independent characteristics of each modality, avoiding the computational overhead and noise that more complex models might introduce. Moreover, electricity demand forecasting often requires high real-time performance, and vector concatenation offers a low computational cost, making it suitable for large-scale data processing and online prediction scenarios. Experimental comparisons demonstrate that this method achieves a good balance between prediction accuracy and computational efficiency in the fusion of time series and textual data, effectively meeting the practical application needs of electricity demand forecasting.

3.4 Prediction Model

The prediction module employs the Bi-directional Gated Recurrent Unit (Bi-GRU) neural network (as shown in Figure 6), using the fused features TSF as input. Bi-GRU is chosen for two key reasons: First, its bidirectional processing effectively models temporal dependencies and mitigates the vanishing gradient problem. Second, while CNNs excel at extracting local features, Bi-GRU's long-term memory capability complements this by capturing broader temporal relationships.

Unlike traditional RNNs, Bi-GRU processes sequences in both forward and backward directions, allowing it to incorporate past and future context. Its architecture consists of input layers, bidirectional GRU layers, and output layers, where each GRU layer contains two

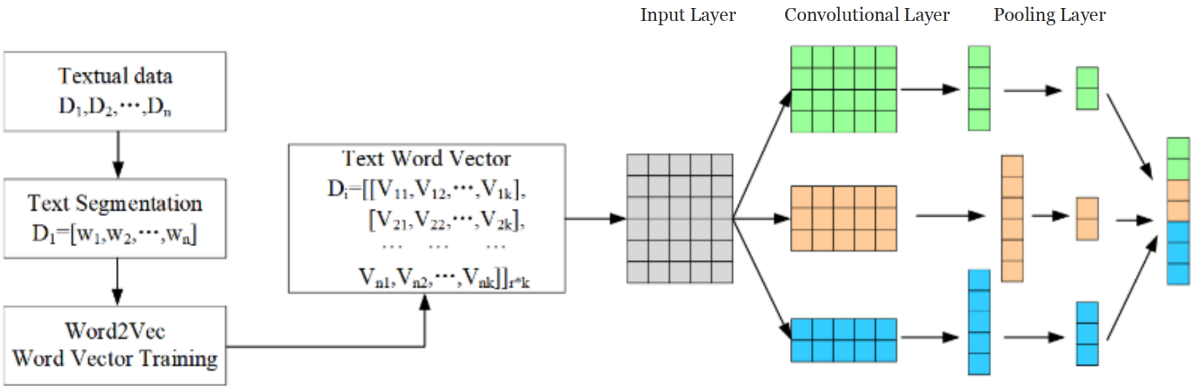


Figure 4. Text feature extraction process.

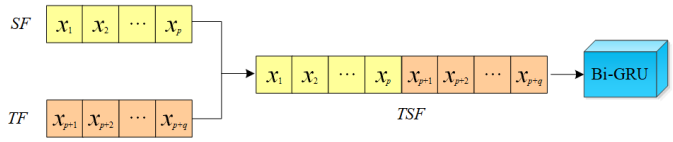


Figure 5. Feature fusion diagram.

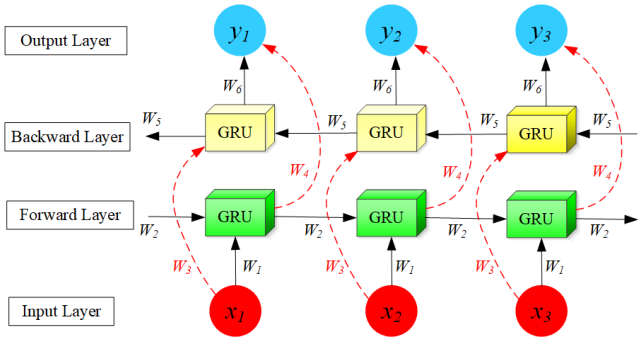


Figure 6. The architecture of Bi-GRU.

subunits processing data in opposite directions. This enables the model to better understand sequential patterns. Bi-GRU has proven effective in tasks such as natural language processing, speech recognition, and time series analysis due to its ability to capture long-range dependencies efficiently.

First, for every GRU, the update gate decides how much of the past information should be retained for the current time step. It takes as input the concatenation of the previous hidden state and the current input. This gate determines the proportion of the previous hidden state that will be combined with the candidate hidden state . It is calculated using a sigmoid activation function as follows:

$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t] + b_z\right) \tag{7}$$

The reset gate controls how much of the past information should influence the calculation of the candidate hidden state $\tilde{h}_t$. It also takes the concatenation of the previous hidden state h_{t-1} and the current input x_t as input and is computed using a sigmoid activation function as follows:

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t] + b_r\right) \tag{8}$$

The candidate hidden state is an intermediate representation computed based on the reset gate and the concatenation of the modified previous hidden state $r_t \odot h_{t-1}$ and the current input x_t. This candidate hidden state captures the potential new information for the current time step.

$$\tilde{h_t} = \tanh\left(W \cdot [r_t \odot h_{t-1}, x_t] + b\right) \tag{9}$$

Finally, the update gate z_t. decides how much of the candidate hidden state $\tilde{h_t}$ should replace the previous hidden state h_{t-1}.

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h_t} \tag{10}$$

where $\odot$ denotes element-wise multiplication.

For Bi-GRU, the input sequence is $[x_1, x_2, \ldots, x_T]$, where x_t is the t-th element in the sequence. The goal of Bi-GRU is to model the entire sequence and extract bidirectional information from it. Bi-GRU consists of two independent GRUs: one for processing the forward sequence and the other for processing the backward sequence. For the forward GRU z_t^f, the forward update gate determines how much the previous hidden state h_{t-1}^f and the current input x_t influence the current hidden state h_t^f. It is calculated using a sigmoid function. Forward Reset Gate r_t^f, the forward reset gate determines how much the previous hidden state h_{t-1}^f and the current input x_t influence the

calculation of the candidate hidden state $\tilde{h_t}^f$. Similarly, it is computed using a sigmoid function.

Finally, Bi-GRU concatenates the forward and backward hidden states to form the complete bidirectional context information. During training, Bi-GRU learns how to simultaneously consider both forward and backward information and generate the final sequential representation.

3.5 Evaluation Criteria

To assess predictive performance, the Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) are utilized to evaluate prediction accuracy. Additionally, the directional accuracy is evaluated using the Dstat metric. The calculation methods for these three metrics are as follows:

RMSE:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2} \tag{11}$$

where n is the total number of observations, Y_i represents the actual value, and $\hat{Y}_i$ denotes the predicted value.

MAPE:

$$\text{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \tag{12}$$

where n is the total number of observations, Y_i represents the actual value, and $\hat{Y}_i$ denotes the predicted value.

Dstat:

$$\text{Dstat} = \frac{1}{n}\sum_{i=1}^{n}\text{sign}(Y_i - Y_{i-1}) \cdot \text{sign}(\hat{Y}_i - \hat{Y}_{i-1}) \tag{13}$$

where n is the total number of observations, Y_i represents the actual value at time i, $\hat{Y}_i$ denotes the predicted value at time i, and $\text{sign}(\cdot)$ is the sign function. Indeed, smaller values of RMSE and MAPE indicate higher accuracy in horizontal predictions, indicating better performance of the prediction model. Conversely, a larger value of Dstat indicates better directional prediction accuracy. Here Dstat $\in [-1, 1]$, where a value closer to 1 indicates perfect directional accuracy and a value closer to -1 indicates consistently incorrect directional prediction. Time steps where $Y_i = Y_{i-1}$ or $\hat{Y}_i = \hat{Y}_{i-1}$ are excluded from the summation.

4 Results and Its Validation

4.1 Dataset

The predictive model requires heterogeneous data from multiple sources, including time series data and various types of textual data. In this study, datasets related to renewable energy supply time series data and relevant textual data were obtained. Each observation in the time series data is influenced by past observations, allowing for the prediction of future trends using its memory. For the renewable energy supply time series dataset, data were sourced from the National Bureau of Statistics of China (https://www.stats.gov.cn/). This dataset includes the electricity generation from major clean energy sources such as hydropower, wind power, and photovoltaic power, aggregated to monthly renewable energy supply. It covers 240 monthly observations from January 2003 to December 2022. As shown in Figure 7, with China actively promoting carbon peaking and carbon neutrality goals, and actively promoting the development of renewable energy, the country's renewable energy generation has been steadily increasing. Additionally, similar to electricity consumption, seasonal fluctuations are observed due to differences in power demand between seasons. Furthermore, due to China's energy resource endowment, the country's power supply is dominated by coal-fired power generation, and investment in renewable energy power generation is also constrained by existing large-scale coal-fired power installations, leading to a gradual slowdown in its growth rate.

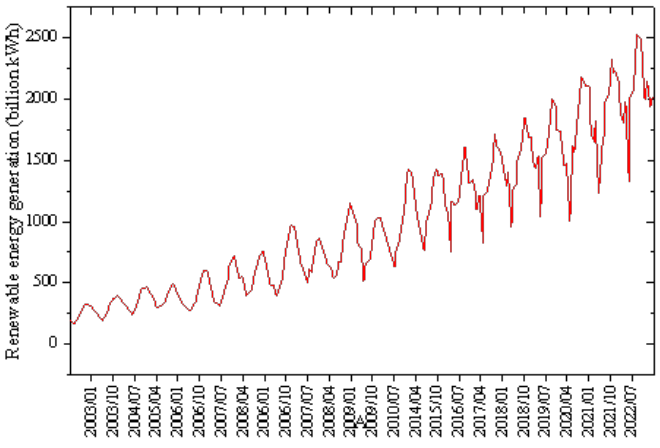


Figure 7. Monthly generation of renewable energy from 2003 to 2022.

The text data reveals that the demand for electricity is influenced by various factors, including political, economic, social, and technological aspects. Short-term electricity demand is primarily affected by seasonality, weather conditions, holidays, and special

events, while medium and long-term forecasts place more emphasis on macroeconomic trends, population growth, energy policies, technological advancements, and societal changes. Therefore, different approaches may be required for short-term and long-term predictions based on text data. To ensure data accuracy, we collected official data concurrent with the electricity time series data through various means, including government policy documents, official energy reports, and news articles related to renewable energy development in China. Ultimately, the text dataset covers 11,342 records from January 2000 to December 2022.

4.2 Model Construction

The experiments in this article were conducted on the Keras deep learning framework. After preprocessing the input data, which includes time series and textual information, it was fed into the proposed CNN-Bi-GRU model for training.

For time series data, the first step involves filling in missing values within the dataset. Next, the time series data needs to be reshaped into one-dimensional format. A parameter, n_steps, is set to determine the size of the time window. The dataset is then structured such that the input samples for the time series data have a length of the preceding n_steps periods, and the value for the next period $(n_steps + 1)$ is used as the label. Subsequently, the data undergoes normalization using the min-max scaling method to map it within a fixed range. Finally, the dataset is divided into training and testing sets at approximately a 9:1 ratio (210 training samples and 24 test samples out of 234 total entries).

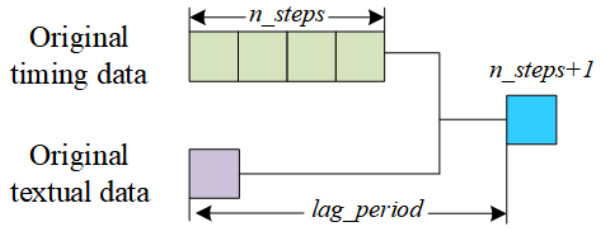


Figure 8. Multi-source data labeling.

Before formally constructing the text data processing module, it is necessary to build the Word2Vec model for Chinese electricity demand. Firstly, each category of text data is aggregated on a monthly basis. Subsequently, textual data from different categories will be combined in chronological order, forming a corpus comprising all text data. Subsequently, this corpus is utilized to train the Word2Vec model, which will serve as the corpus for text vectorization in subsequent steps. Through multiple experimental comparisons, the dimensionality parameter value for vectors is set to 200, and the sliding window parameter value is selected as 15. Additionally, a lag_period is set as the text's lagging period to solve the delayed effect of text data. The corresponding labels are illustrated in Figure 8.

Upon formally constructing the text data processing module, after the basic preprocessing steps such as tokenization and stop word removal, the raw unstructured text data is vectorized using the pre-trained Word2Vec model. Subsequently, the parameter lag_period is set, and in conjunction with the pre-allocated training and testing datasets for the time series data, the training and testing datasets for the text data are partitioned. Finally, a subset of text data with a length of lag_period is retained for predicting future trends. Once the complete dataset is constructed, text data of length lag_period is retained based on the varying lag periods. For experiments, the ratio of the dataset to the training set is 9:1. The model is trained using the training set and its predictive ability is evaluated using the test set.

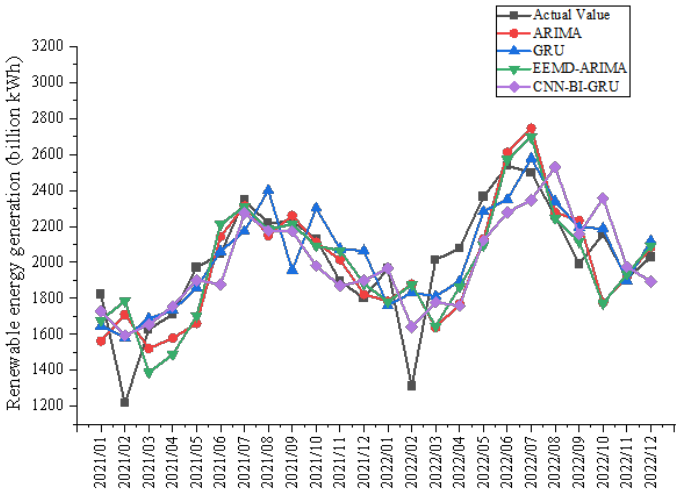


Figure 9. Short-period forecast curves for four models.

Subsequently, when constructing the temporal CNN neural network, the network is set up according to the Inception structure with adjustable width. For each channel of the network, an input layer, convolutional layer, pooling layer, and flattening layer are added, with the number of convolutional kernels in the convolutional layer set as an adjustable parameter. The outputs of the flattening layers from each channel are then aggregated to form the output of the temporal CNN neural network. As for the text CNN neural network, the operations are similar to the temporal CNN, except that an Embedding layer is added at the

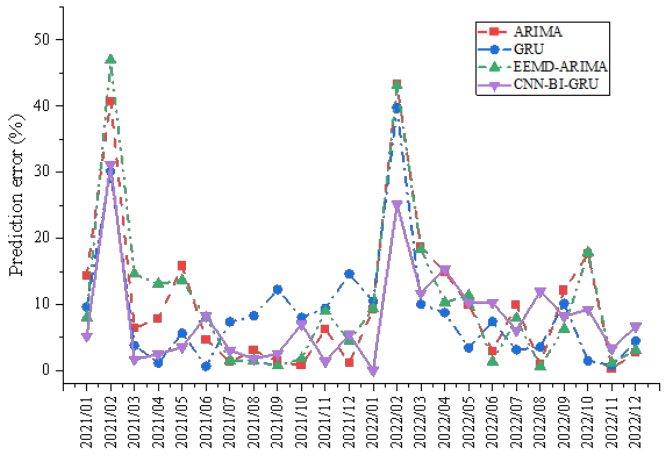


Figure 10. Short-period forecast error curves for four models.

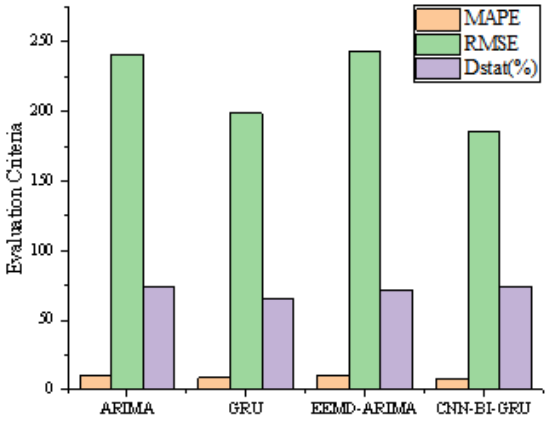


Figure 11. Evaluation criteria for short-period forecasting of four models.

input layer to pre-train word vectors.

Finally, the Bi-GRU prediction module is constructed. The two types of latent features obtained from the feature extraction module are aggregated as inputs to the Bi-GRU neural network. The number of hidden neurons in the network is set as an adjustable parameter. MSE is utilized as the loss function, and Adam is employed as the optimizer to train the model.

4.3 Model Evaluation

The dataset is labeled using the renewable energy generation data from the previous 6 months and aggregated textual data from 36 months ago to predict the data for the next month. The final dataset comprises 234 entries, with 210 entries in the training set and 24 entries in the test set. This section will compare the effectiveness of the proposed renewable energy supply prediction model based on multimodal information fusion with ARIMA model, single GRU network, and composite model EEMD-ARIMA (Ensemble Empirical Mode Decomposition combined with ARIMA) to validate its efficacy. This experiment is primarily divided into two parts: the first part focuses on short-term forecasting by month, while the second part deals with long-term forecasting on a semi-annual basis.

Figure 9 demonstrates the prediction curves of the real value of electricity with the four models ARIMA, GRU, EEMD-ARIMA, and CNN-BI-GRU for renewable electricity. Figure 10 presents the prediction error curves for the four models. Figure 11 displays the three evaluation metrics MAPE, RMSE, Dstat for the four models. Based on these short-period forecast results, we can draw the following conclusions.

First, Figure 9 presents a comparison between the actual electricity values and the predictions made by four different models: ARIMA, GRU, EEMD-ARIMA, and CNN-BI-GRU. It is noticeable that while all models follow the trend of the actual values to some extent, they each have varying degrees of accuracy and precision. The short-term nature of the forecast likely introduces significant uncertainty due to the influence of many unpredictable factors that can affect electricity consumption and production on a short scale. Additionally, Figure 10 illustrates the forecasting errors of the same models. It appears that some models have larger spikes in error, potentially indicating moments when the actual values had sudden changes that the models failed to predict accurately. The variability and sudden changes in electricity demand or supply can be difficult for some models to capture, especially in a short-term context. In addition, Figure 11 shows the evaluation metrics MAPE, RMSE, and Dstat. These metrics provide a quantitative measure of the models' performance. The results show that the proposed CNN-Bi-GRU model achieves the lowest MAPE and RMSE values among the four models, indicating superior average prediction accuracy. Besides, among these four models, the Dstat of the model proposed in this paper is relatively high, which also indicates that the prediction direction of the model proposed in this paper is more accurate.

Figure 12 presents the prediction curves of the actual power values by four different models: ARIMA, GRU, EEMD-ARIMA, and CNN-Bi-GRU for renewable electricity. Figure 13 depicts the error curves of these models' predictions. Figure 14 displays the performance of these four models across three evaluation metrics: MAPE, RMSE, and Dstat. Based on the results of these long-term forecasts, we can draw

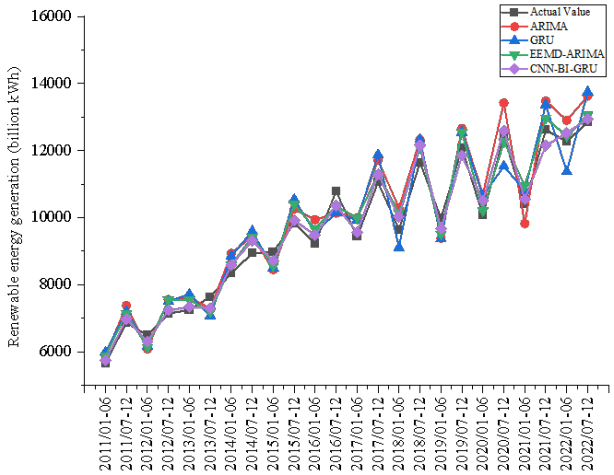


Figure 12. Long-period forecast curves for four models.

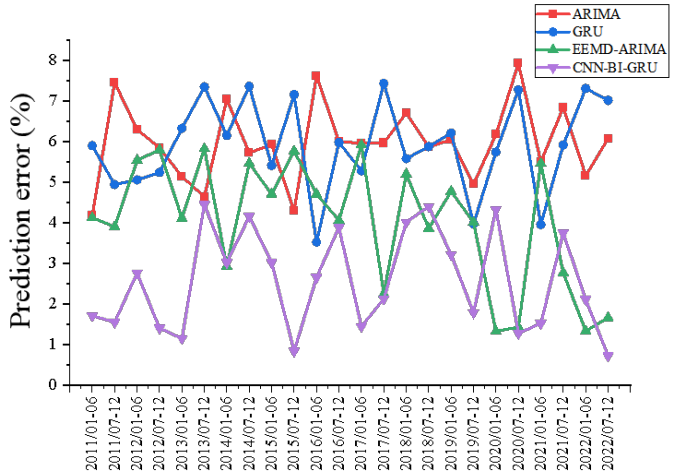


Figure 13. Long-period forecast curves for four models.

the following conclusions.

Firstly, Figure 12 compares the actual power values with the forecasted values by the four distinct models: ARIMA, GRU, EEMD-ARIMA, and CNN-Bi-GRU. These models exhibit trends similar to the actual data in long-term forecasting, demonstrating higher predictive accuracy compared to short-term forecasts. This suggests that the impact of many unpredictable factors is significantly reduced in long-term forecasting. Figure 13 presents the predictive error of these models, indicating that long-term forecasting errors are generally lower than those of short-term forecasts, which underscores the reliability and credibility of long-term predictions. Additionally, the evaluation metrics of MAPE, RMSE, and Dstat shown in Figure 14 provide a quantitative assessment of the models' performance. The proposed model in this study exhibits smaller values in MAPE and RMSE compared

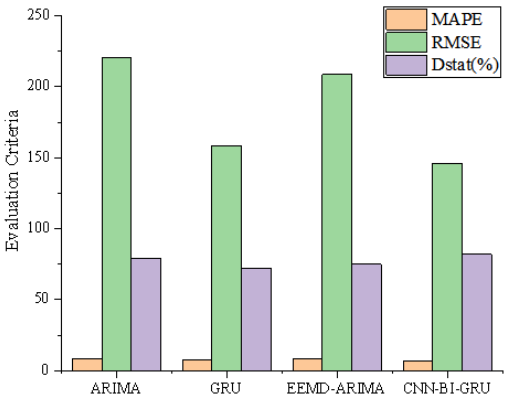


Figure 14. Evaluation criteria for long-period forecasting of four models.

to the ARIMA, GRU, and EEMD-ARIMA models, indicating superior average performance. Moreover, the higher Dstat value for the proposed model suggests more accurate predictive directionality.

In addition, we conducted paired sample t-tests to evaluate the performance differences between CNN-Bi-GRU and the baseline models (ARIMA, GRU, EEMD-ARIMA). We conducted paired sample t-tests (two-tailed, $\alpha = 0.05$, $df = 23$) on the 24 test-set observations. The results indicate that CNN-Bi-GRU demonstrates statistically significant improvements over ARIMA ($p = 0.01$), GRU ($p = 0.03$), and EEMD-ARIMA ($p = 0.02$), confirming that the observed performance enhancements are not due to random variations but result from the structural optimizations of the proposed model.

For the comparison between CNN-Bi-GRU and ARIMA, the p = 0.01 result suggests a highly significant difference. This indicates that deep learning models can more effectively capture the complex nonlinear characteristics of electricity demand compared to traditional time series methods such as ARIMA. Since ARIMA relies on linear assumptions, it performs well in stable time series forecasting but struggles to generalize when faced with high volatility and multi-factor influences in electricity demand. In contrast, CNN-Bi-GRU overcomes this limitation by utilizing deep feature extraction and bidirectional temporal modeling.

In the comparison between CNN-Bi-GRU and GRU, the p = 0.03 result indicates a statistically significant improvement. This can be attributed to the bidirectional structure of Bi-GRU, which enables the model to consider both past and future information in a time series, whereas unidirectional GRU only

leverages historical data for prediction. As a result, GRU has certain limitations in modeling long-term dependencies. Additionally, the integration of CNN enhances feature extraction, allowing CNN-Bi-GRU to outperform GRU in both RMSE and MAPE metrics.

For the comparison between CNN-Bi-GRU and EEMD-ARIMA, the $p = 0.02$ result confirms a statistically significant performance gain. Although EEMD-ARIMA improves traditional statistical methods by using EEMD to denoise and smooth time series data, it still relies on ARIMA's linear modeling capabilities. This makes it less effective in capturing the complex fluctuations in electricity demand. In contrast, CNN-Bi-GRU leverages CNN to extract local temporal features and utilizes Bi-GRU to model temporal dependencies, providing superior adaptability in handling nonlinear, cyclical, and abrupt changes.

In summary, whether for short-term or long-term demand forecasting of renewable energy, the proposed multimodal information fusion CNN-Bi-GRU model outperforms the ARIMA, GRU, and EEMD-ARIMA models in terms of predictive performance. These experimental results validate the effectiveness of the predictive design scheme proposed in this paper. It is worth noting that the proposed model still has certain limitations when dealing with specific types of data or particular scenarios. For instance, extreme unexpected events, such as power grid failures or abrupt policy changes, may cause significant shifts in historical data patterns. Since the CNN-Bi-GRU model relies on time series features and textual information, it may struggle to quickly adapt to such changes, thereby affecting prediction accuracy. Additionally, data missingness or poor data quality poses a significant challenge. When the input data contains a large number of missing values or anomalies, the model may fail to effectively learn trend features, leading to prediction biases.

5 Conclusion

This study presents a novel integrated forecasting modeling approach, CNN-Bi-GRU, for short-term and long-term renewable power demand prediction, based on the concept of multimodal information fusion. In this approach, we delve into the analysis of data essential for power demand forecasting and employ the robust feature extraction capabilities of CNNs to extract relevant latent features from both time series and textual data. Subsequently, a concatenation fusion method is applied to organically merge these features, harnessing the complementary nature of information from both modalities. Finally, the fused features are inputted into a Bi-GRU model for prediction. Through comparative experiments with traditional forecasting models and ensemble models, the proposed model exhibits superior directional and level accuracy over widely adopted time series forecasting models, thus effectively validating the superiority of our model in terms of predictive performance. The proposed model is primarily applied to smart grid optimization, electricity market scheduling, and renewable energy generation forecasting. However, challenges remain in practical deployment, including computational costs, data acquisition and quality issues, and extreme unexpected events. Additionally, this study has certain limitations. For instance, the feature extraction module employs a convolutional neural network to extract textual features, making it difficult to interpret the actual significance of the extracted information. Furthermore, the chosen multimodal information fusion method is based on feature-level fusion, which limits the depth of text data mining in this study. Future work will focus on further optimizing the model structure, improving multimodal fusion efficiency, and enhancing the model's anomaly detection capabilities.

Data Availability Statement

The renewable energy time series data used in this study are publicly available from the National Bureau of Statistics of China (https://www.stats.gov.cn/). The text dataset is available from the corresponding author upon request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Singh, A. K., Ibraheem, S. K., Muazzam, M., & Chaturvedi, D. K. (2013). An overview of electricity demand forecasting techniques. *Network and complex systems*, 3(3), 38-48.

[2] Mir, A. A., Alghassab, M., Ullah, K., Khan, Z. A., Lu, Y., & Imran, M. (2020). A review of electricity demand forecasting in low and middle income countries: The

demand determinants and horizons. *Sustainability*, 12(15), 5931. [CrossRef]

[3] Bianco, V., Manca, O., & Nardini, S. (2013). Linear regression models to forecast electricity consumption in Italy. *Energy Sources, Part B: Economics, Planning, and Policy*, 8(1), 86-93. [CrossRef]

[4] Arisoy, I., & Ozturk, I. (2014). Estimating industrial and residential electricity demand in Turkey: A time varying parameter approach. *Energy*, 66, 959-964. [CrossRef]

[5] Pappas, S. S., Ekonomou, L., Karamousantas, D. C., Chatzarakis, G. E., Katsikas, S. K., & Liatsis, P. (2008). Electricity demand loads modeling using AutoRegressive Moving Average (ARMA) models. *Energy*, 33(9), 1353-1360. [CrossRef]

[6] Erdogdu, E. (2007). Electricity demand analysis using cointegration and ARIMA modelling: A case study of Turkey. *Energy policy*, 35(2), 1129-1146. [CrossRef]

[7] Wang, Y., Wang, J., Zhao, G., & Dong, Y. (2012). Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: A case study of China. *Energy Policy*, 48, 284-294. [CrossRef]

[8] Mitkov, A., Noorzad, N., Gabrovska-Evstatieva, K., & Mihailov, N. (2019, June). Forecasting the energy consumption in Afghanistan with the ARIMA model. In *2019 16th conference on electrical machines, drives and power systems (ELMA)* (pp. 1-4). IEEE. [CrossRef]

[9] Hippert, H. S., Pedreira, C. E., & Souza, R. C. (2001). Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on power systems, 16*(1), 44-55. [CrossRef]

[10] Amjady, N. (2001). Short-term hourly load forecasting using time-series modeling with peak load estimation capability. *IEEE Transactions on power systems*, 16(3), 498-505. [CrossRef]

[11] Feinberg, E. A., & Genethliou, D. (2005). Load forecasting. *Applied mathematics for restructured electric power systems: optimization, control, and computational intelligence*, 269-285. [CrossRef]

[12] Kyriakides, E., & Polycarpou, M. (2007). Short term electric load forecasting: A tutorial. *Trends in neural computation*, 391-418. [CrossRef]

[13] Kuo, P. H., & Huang, C. J. (2018). A high precision artificial neural networks model for short-term energy load forecasting. *Energies, 11*(1), 213. [CrossRef]

[14] Perrusquía, A., & Yu, W. (2021). Identification and optimal control of nonlinear systems using recurrent neural networks and reinforcement learning: An overview. *Neurocomputing*, 438, 145-154. [CrossRef]

[15] AlShafeey, M., & Csáki, C. (2021). Evaluating neural network and linear regression photovoltaic power forecasting models based on different input methods. *Energy Reports*, 7, 7601-7614. [CrossRef]

[16] Wu, C., Wang, J., Chen, X., Du, P., & Yang, W. (2020). A novel hybrid system based on multi-objective optimization for wind speed forecasting. *Renewable energy*, 146, 149-165. [CrossRef]

[17] Bedi, J., & Toshniwal, D. (2019). Deep learning framework to forecast electricity demand. *Applied energy, 238*, 1312-1326. [CrossRef]

[18] Valenzuela, O., Rojas, I., Rojas, F., Pomares, H., Herrera, L. J., Guillén, A., ... & Pasadas, M. (2008). Hybridization of intelligent techniques and ARIMA models for time series prediction. *Fuzzy sets and systems*, 159(7), 821-845. [CrossRef]

[19] Liu, Y., Guan, L., Hou, C., Han, H., Liu, Z., Sun, Y., & Zheng, M. (2019). Wind power short-term prediction based on LSTM and discrete wavelet transform. *Applied Sciences, 9*(6), 1108. [CrossRef]

[20] Nie, Y., Jiang, P., & Zhang, H. (2020). A novel hybrid model based on combined preprocessing method and advanced optimization algorithm for power load forecasting. *Applied Soft Computing*, 97, 106809. [CrossRef]

[21] Saglam, M., Spataru, C., & Karaman, O. A. (2022). Electricity demand forecasting with use of artificial intelligence: the case of Gokceada Island. *Energies*, 15(16), 5950. [CrossRef]

[22] Zhang, W., Quan, H., & Srinivasan, D. (2018). An improved quantile regression neural network for probabilistic load forecasting. *IEEE Transactions on Smart Grid, 10*(4), 4425-4434. [CrossRef]

[23] Lee, S. J. (2023). *Multimodal Data Fusion for Estimating Electricity Access and Demand* (Doctoral dissertation, Massachusetts Institute of Technology).

[24] Xuan, Y., Si, W., Zhu, J., Sun, Z., Zhao, J., Xu, M., & Xu, S. (2021). Multi-model fusion short-term load forecasting based on random forest feature selection and hybrid neural network. *IEEE Access*, 9, 69002-69009. [CrossRef]

[25] Ji, Q., Zhang, S., Duan, Q., Gong, Y., Li, Y., Xie, X., ... & Zhao, X. (2022). Short-and medium-term power demand forecasting with multiple factors based on multi-model fusion. *Mathematics*, 10(12), 2148. [CrossRef]

[26] Kong, Z., Zhang, C., Lv, H., Xiong, F., & Fu, Z. (2020). Multimodal feature extraction and fusion deep neural networks for short-term load forecasting. *IEEE Access*, 8, 185373-185383. [CrossRef]

[27] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*. [CrossRef]

[28] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).

[29] Er, M. J., Zhang, Y., Wang, N., & Pratama, M. (2016). Attention pooling-based convolutional neural

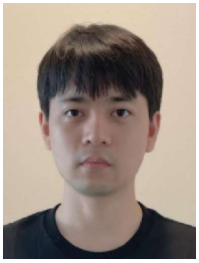
network for sentence modelling. *Information Sciences*, 373, 388-403. [CrossRef]

[30] Yang, J., Yang, J. Y., Zhang, D., & Lu, J. F. (2003). Feature fusion: parallel strategy vs. serial strategy. *Pattern recognition*, 36(6), 1369-1381. [CrossRef]

[31] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).



Shuchen Zhao He received the M.Sc. degree in Electrical and Computer Engineering from Duke University, Durham, NC, in 2020. He is currently an engineer at Amazon, Inc., focusing on data infrastructure and platform development for logistics and delivery services. (Email: shuchen.zhao.1996@gmail.com)



Zhongjin Xu He received the M.Sc. degree in Electrical and Computer Engineering from Duke University, Durham, NC, in 2020. He is currently an engineer at Amazon, Inc., specializing in data pipeline design and system optimization for delivery operations. (Email: zhongjin@umich.edu)

Zhefan Zhu received the B.Sc. degree in biological technology from Nankai University, Tianjin, China in 2015 and the M.Sc. degree in computational biology from Carnegie Mellon University, Pittsburgh, United States in 2018. He is currently working as a software engineer, where he applies his expertise in machine learning and software engineering to industrial challenges. (Email: zhefanzhu@gmail.com)

Xiaoxiang Liang received his Master's degree in Business Analytics from Washington University in St. Louis, Missouri, in 2022. He currently contributes to Capital One, Inc., where he focuses on innovating, developing, and maintaining robust data solutions to address complex business challenges. Leveraging cutting-edge machine learning technologies, Xiaoxiang works to uncover valuable opportunities that empower everyday individuals to save money, save time, and reduce the complexities in their financial lives. (Email: l.xiaoxian@outlook.com)



Zecheng Zhang received a B.Sc. degree in Mathematics from the Chinese University of Hong Kong in 2021 and an M.Sc. degree in Computer Science from NYU in 2023. He is currently working on building machine learning infrastructure at Bytedance Inc. (Email: roderickzzc@gmail.com)



Ruxue Jiang received the B.Sc. degree in Data Science from Fudan University, Shanghai, China in 2021 and the M.Sc. degree in Computer Science from Northeastern University, Seattle, in 2024. She is currently working on building the #1 Integration platform in the world with Salesforce Inc. (Email: snow0911@outlook.com)