J. Electrical Systems 20-2 (2024): 1924-1932

1,*Jinkai Sun
2Yulu Ren
3Junwei Zhang
4Xiaofang Chen

An anomaly Detection Method for Electricity Consumption Data Based on CNN-BiLSTM-Attention

Journal of Electrical Systems

Abstract: - As the complexity and uncertainty of smart distribution networks increase, data security issues in smart meters have become a pressing challenge, such as false data injection and electricity theft. To ensure fairness, safety, and overall economic efficiency of distribution networks, it is essential to accurately detect abnormal electricity consumption. However, traditional methods relying on on-site inspections by grid personnel suffer from low efficiency and high costs in detecting user anomalies. This paper proposes an electricity consumption data anomaly detection method based on CNN-BiLSTM-Attention. CNN is utilized to extract data features, while BiLSTM and attention mechanisms capture contextual information in sequence data. Furthermore, experiments conducted on data extracted from smart meters demonstrate that the proposed model outperforms other models in anomaly detection, with accuracy, recall, and F1-Score all exceeding 91%. These results validate the effectiveness and feasibility of the proposed method, providing an efficient solution for user anomaly detection in national power grids.

Keywords: Abnormal Electricity Consumption, CNN, BiLSTM, Attention.

I. INTRODUCTORY

With the growing popularity of the intelligentization of China's power grid equipment and the ongoing refinement of the electricity billing mechanism, the complexity of the environment and uncertainty in the actual operation of the smart distribution network have also increased. Furthermore, smart meters face data security issues, including false data injection and electricity theft, which have become more profound. Preventing the injection of false data into the electricity system, accurately detecting instances of electricity theft, and identifying abnormalities in the power information collection system are crucial to ensure the security and fairness of the distribution network, as well as the overall economy of the network. The standard method of detecting abnormal electricity consumption necessitates the installation of metering boxes and on-site inspection by grid personnel, leading to excessive expenditure in both time and resources. Electric power enterprises currently have access to a vast and intricate collection of data, encompassing sales, power consumption, user behavior, and other time series data. Our focus is on utilizing this data to detect abnormal user power consumption, conduct detailed analysis, and make efficient determinations of said anomalies.

II. RELATED WORK

Wherever Times is specified, Times Roman or Times New Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times. Avoid using bit-mapped fonts if possible. True-Type 1 or Open Type fonts are preferred. Please embed symbol fonts, as well, for math, etc.

The National Institute of Standards and Technology (NIST) released "Smart Grid Information Security Strategies and Requirements" in 2009. This publication was created to merge energy and information technology in order to establish a grid information security system that can efficiently manage security threats within the electric power system and upgrade the smart grid's security[1]. China's smart grid research began after advanced foreign countries, and the Provisions on Security Protection of Electric Power Monitoring System, enacted in 2014, elevated it to a national level for the first time. As a result, the security of the smart grid has become a growing concern, prompting more research by domestic and international scholars. Some recent studies by scholars like Zheng Shiying and Niu Qing[2]have shown that conducting analysis and research using electric power big data can contribute to the identification of anomalous power consumption patterns. The current research primarily concentrates on the instrument's prerequisites and relies mainly on manual extraction. However, the detection rate of abnormal electricity usage remains low. The detection of abnormal electricity consumption is typically executed through conventional means, such as manually inspecting the installation and configuration of the meter,

1State Grid Shanxi Marketing Service Center, Taiyuan, 030012, China
2 State Grid Shanxi Marketing Service Center, Taiyuan, 030012, China
3 State Grid Shanxi Marketing Service Center, Taiyuan, 030012, China
4 State Grid Shanxi Marketing Service Center, Taiyuan, 030012, China
*Corresponding author: Jinkai Sun

comparing meter readings from normal and abnormal meters, and examining the bypass transmission line. However, this approach can be arduous, expensive, and ineffective. Pan Jun[3] and other scholars argue that precise monitoring of abnormal electricity usage is crucial, and this can only be achieved through obtaining accurate information regarding electricity abnormalities from the user side of the smart grid. Zhang Guoqing[4] and other researchers, employed a multi-layer perceptron neural network to detect abnormal power in power systems. The method's resilience was evidenced through the neural network classifier's optimal design. You Qianqian[5] and fellow scholars constructed an abnormal power usage detection model using self-encoder and SVM. Feature extraction and classification were performed to reduce the false alarm rate while maintaining a high detection rate. Effective and accurate monitoring of abnormal power usage, utilizing abnormal power information obtained from the user side of the smart grid, is of paramount importance. Tao H.[6] and other researchers utilized energy consumption data from equipment to derive classification rules for users' normal and abnormal power usage models. They proposed an abnormal detection method for household power consumption based on a convolutional neural network.

Numerous studies have indicated that security issues within smart grids have become a prominent concern. One such issue is the detection of abnormal power usage in these grids, which has received increasing attention from researchers. Currently, researchers are primarily focused on the performance aspects of detecting abnormal power consumption data. With the rapid growth of power user data and equipment, the dimension and data volume of power user data are increasing rapidly, leading to the challenge of low performance in existing power data anomaly detection algorithms. To resolve this issue, our study analyzes the temporal and high-dimensional properties of power user data and presents an anomaly detection algorithm for power user data based on CNN-BiLSTM-Attention. Through experimentation, it has been demonstrated that the algorithm presented in this paper outperforms the traditional algorithm and enhances algorithmic performance.

III. CNN-LSTM-ATTENTION MODEL STRUCTURE

A. *Convolutional Neural Network*

The Convolutional Neural Network (CNN) can effectively capture local spatial information and associations between features, making it a frequently employed feature extraction network in the field of deep learning[7]. It is capable of automatically learning useful features from the input data and transforming them into more meaningful representations for subsequent tasks. As depicted in Figure 1[8], the convolutional layer of a CNN carries out local perception and feature extraction by utilizing a group of filters that are modifiable. These filters identify various patterns, shapes, and textures in the input data. The paper uses the tanh hyperbolic tangent activation function to enable the extraction of more complex features after the convolutional layer. Stacking multiple convolutional layers gradually captures features at various levels, ranging from simple to complex and from low to high level. Subsequent pooling layers extract crucial features and maintain important spatial information by decreasing the dimensionality and size of the feature map, ultimately reducing the sensitivity of the model to noise and extraneous details in the input data, thereby improving the model's robustness. By utilizing convolutional and pooling layers, a CNN can extract abstract feature representations from raw data that capture important attributes and patterns. This allows for more efficient processing and analysis of complex data.

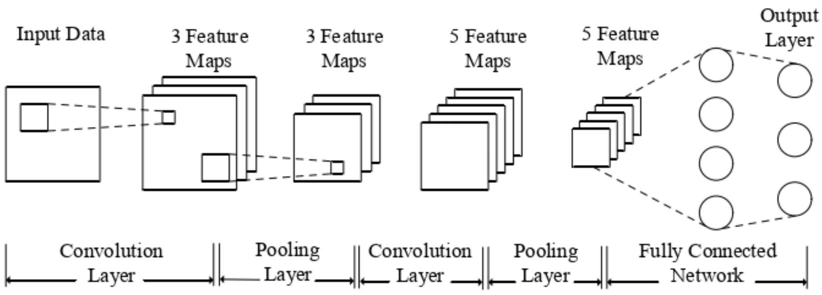


Figure 1: Structure of CNN Network

B. *Long Short-Term Memory*

Long Short-Term Memory (LSTM) is a type of Recurrent Neural Network (RNN) that deals with the problem of long-term dependency in traditional RNNs by capturing and processing such dependencies effectively in time series data[9]. The components that constitute the LSTM network are illustrated in Figure 2, where in C, he memory cell, controls the sequence transmission through three gating cells[10].

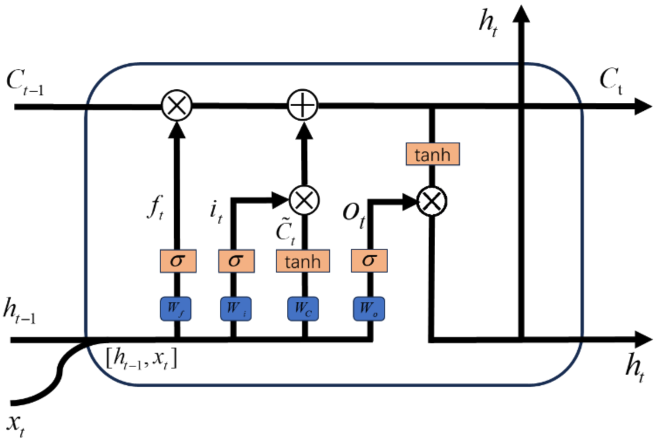


Figure 2: Structure of LSTM Network

The Forget Gate gets the output h_{t-1} from the previous moment and the input x_t from the current moment. It uses a sigmoid activation function to output a value between 0 and 1, indicating the amount of information retained in the memory cell from the previous moment. A value close to 1 indicates that the majority of the information is retained, whereas a value close to 0 indicates that most of the information is forgotten.

Input Gate: The input gate comprises a sigmoid activation function that decides which information requires updating, alongside a tanh activation function that generates a novel candidate value. The output of the sigmoid is multiplied by the output of the tanh and added to the output of the forgetting gate to derive an updated memory cell. Output Gate: The output gate is composed of a sigmoid activation function that determines the relevant parts of the output and a tanh activation function that scales the output. The knowledge state of the model at time t-1, denoted by C_{t-1}, is updated with newly acquired information, denoted by $\tilde{C}_t$, through a process of multiplication with corresponding weighting parameter and subsequent summation, resulting in an updated knowledge state at time t. The weighting parameter determines the amount of knowledge retained at each moment. The above process corresponds to the following equation:

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C_t} \tag{1}$$

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{2}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3}$$

$$\widetilde{C_t} = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{4}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t * \tanh(C_t) \tag{6}$$

Where $[h_{t-1},x]$ is the new vector created by combining the input from the previous moment and the current moment. W_f represents the weight of the forgetting gate, and bf represents the bias value of the forgetting gate. W_c represents the weight of the memory cell, and b_c represents the bias value of the memory cell. W_i represents the weight of the input gate, and b_i represents the bias value of the input gate. Finally, W_o represents the weight of the output gate, and b_o represents the bias value of the output gate.

LSTM networks have the capability to selectively and adaptively forget, update, and output information when learning, which makes them ideal for capturing long-term dependencies. This is particularly significant for processing time-series data, especially for tasks requiring the consideration of long-range dependencies.

C. Attention Mechanism

Attention Mechanism (AM) is a technique that simulates how much attention individuals allocate to different parts of the information they process, enabling the model to concentrate on the vital information and enhance the model's input processing. The research of Bahdanau et al. (2014) and Luong et al. (2015) is regarded as the foundation of Attention Mechanisms in Deep Learning work[11]. The focus of this paper is to weigh the features that were extracted by the preceding CNN and LSTM. The aim is to capture the spatial and temporal fusion features of the original sequence, and subsequently input them into the model for prediction. Figure 3 illustrates the general structure of the model.

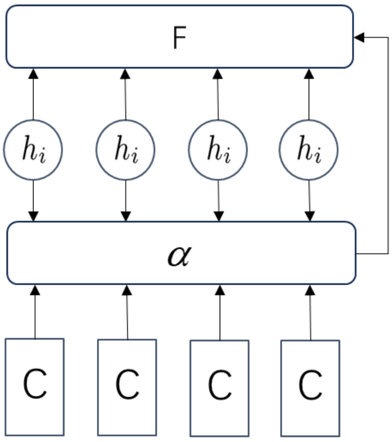


Figure 3: Structure of Attention

$$\varphi(h_i, C) = \tanh(h_i \cdot W_\alpha \cdot C_t + b_\alpha) \tag{7}$$

$$\alpha_i = \frac{\exp(\varphi(h_i,C))}{\sum_{j=1}^{n} \exp(\varphi(h_j,C))} \tag{8}$$

$$F = \sum_{i=1}^{n} \alpha_i h_i \tag{9}$$

In the above equations[12], the feature vector C is extracted after the CNN network, the hi feature vector is extracted by the LSTM network at moment i, and the weight is denoted by W_α , with the bias term being denoted by b_α . Eq. (7) weights the features extracted by CNN and LSTM and obtains the fused weights through the activation function. Eq. (8) passes the weights through the softmax function and finally, multiplies the output value of LSTM with the weights to obtain the final features through Eq. (9). The data is fed through a network's fully-connected layer, and the output is used to classify it as either 0 (normal) or 1 (abnormal) using the softmax function.

D. Model Structure

The structure of the CNN-BiLSTM-Attention model proposed in this paper is shown in Figure 4.

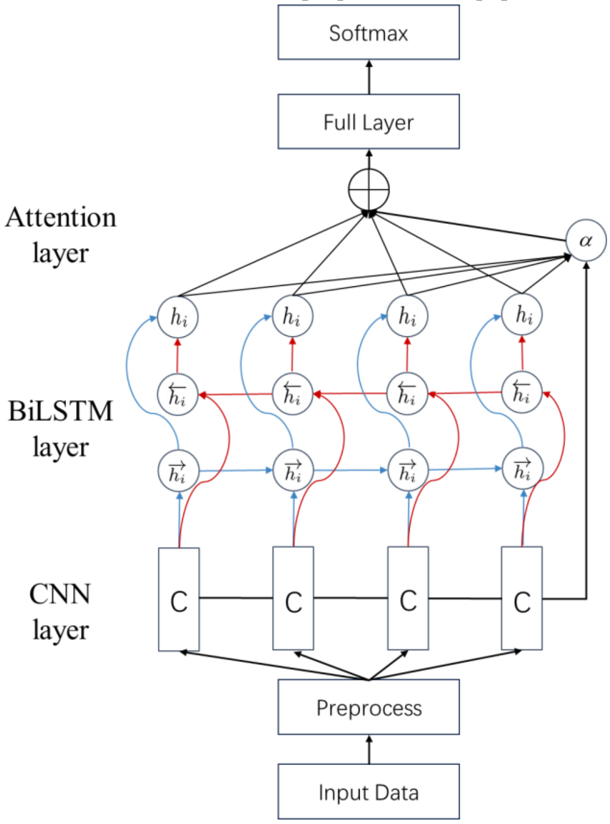


Figure 4: Structure of CNN-BiLSTM-Attention Model

The algorithm flow for detecting abnormal electricity usage in residential areas, which is designed in this paper based on CNN-BiLSTM-Attention architecture, is outlined as follows:

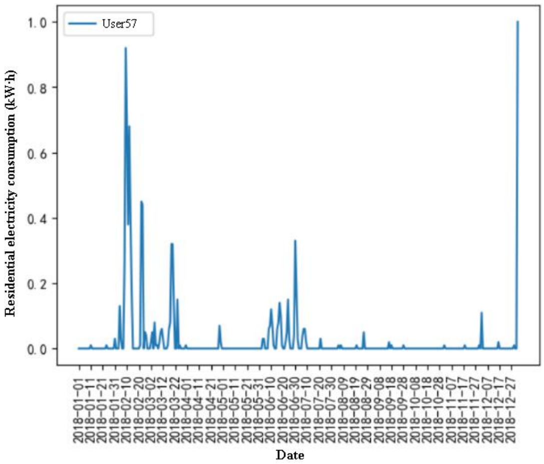
- involves preprocessing actual data related to residential electricity consumption and utilizing it as the model's input.
- contour features are extracted utilizing CNN.

- sequence features are extracted utilizing BiLSTM.
- involves using the Attention structure to weigh and combine the features.
- Proceed through the fully connected layers and classifiers.
- Obtain the outcome of abnormal power usage classification and terminate.

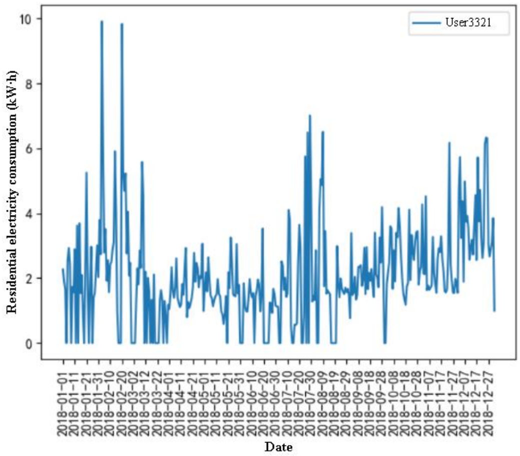
IV. Experiments and Analysis of Results

A. Experimental Data and Pre-processing

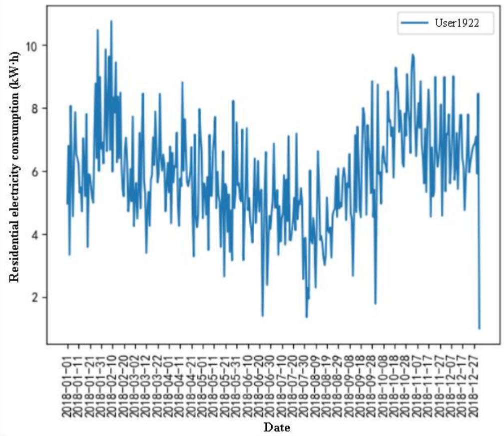
1) Data Cleaning: In this study, we selected real residential electricity consumption data collected by the power marketing business system as our research focus. The data was collected from January 1st, 2018 to December 30th, 2018, with a daily sampling interval, and includes a comprehensive package of 3,400 users' electricity consumption data over the course of 365 days. The data information consists of user ID, daily electricity consumption of users for a period of one year, and labels for abnormal electricity consumption. Randomly selected four users' electricity consumption data for the whole year are plotted to draw the electricity load curve, as shown in Figure 5.



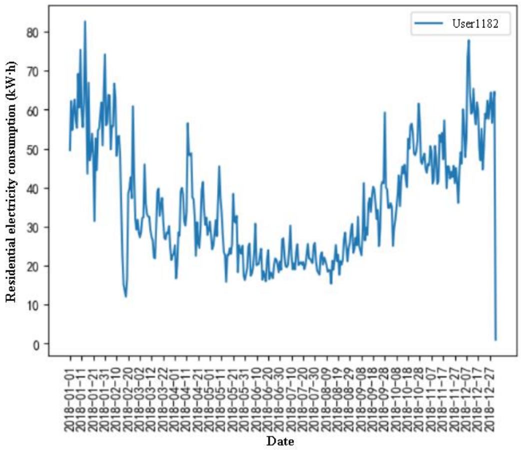
(a) User 57



(b) User 3321



(c) User 1922

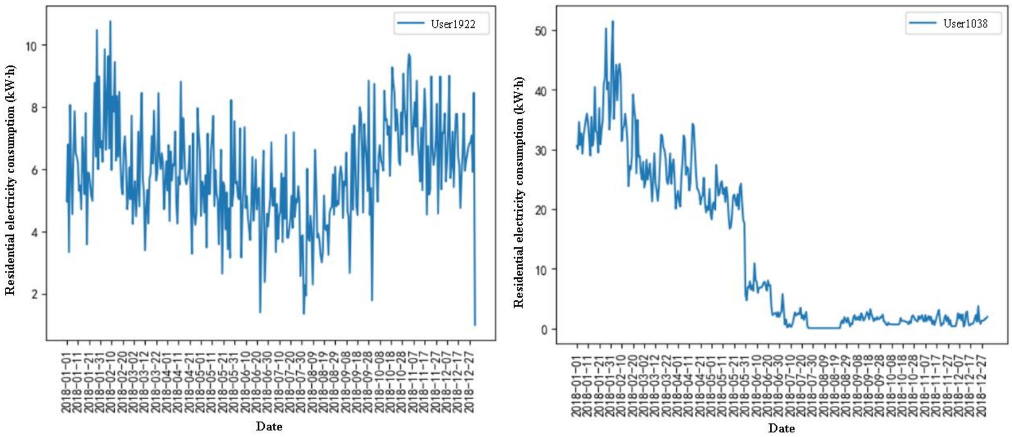


(d) User 1182

Figure 5: Residential Electricity Consumption

By visualizing residential electricity consumption data, it becomes apparent that electricity consumption behavior among users is mostly regular. However, there are exceptions such as user 57 (Figure 5(a)) and user 3321 (Figure 5(b)), There are fluctuations in electricity consumption data around the 0 scale, and it is evident that some users have missing electricity consumption data. There are two approaches to handling these missing values[13]. If the user has over 50% of their data missing, they will be removed. If the amount of missing data is less than 50%, the data will be supplemented using the Lagrangian interpolation method, such as with user 3321. After eliminating users with missing data exceeding 50%, such as user 57, the remaining users with less than 50% missing data will have missing values filled in using the Lagrangian interpolation method, such as user 3321. As a result of the data cleaning process, we obtained electricity consumption information from a total of 3,323 users with complete data. This data will be used as the experimental data for subsequent analysis.

2) Analysis of Abnormal Electricity Consumption data: Figure 6 displays a visualization of electricity user data, differentiating between normal and abnormal cases.



(a)Normal Consumer Data (b)Abnormal consumer data

Figure 6: Display of Normal Electricity User Data and Abnormal Electricity User Data

Figure 6(a) displays the yearly electricity usage of client number 1922, labeled 1 in the anomaly table, indicating a normal electricity customer. Figure 6(b) exhibits the annual electricity consumption of customer number 1038, assigned a value of 2 in the anomaly label table, establishing the client as an abnormal electricity user. Through comparison, it is evident that the electricity consumption trend of regular users remains consistent, whereas during the halfway mark, the electricity consumption of anomalous user 1038 abruptly drops to zero. This unequivocally points towards electricity theft by the user.

The overall sample size is 3323, comprising 2743 normal data and 681 abnormal data. The portion of abnormal electricity usage samples in the actual data is 20.49%, which constitutes a small percentage. The experimental sample is significantly unbalanced, which may lead to suboptimal model training. Therefore, this paper utilizes the SMOTE (Synthetic Minority Oversampling Technique)[14] in conjunction with undersampling to address the data imbalance issue. Specifically, abnormal electricity consumption residential data is interpolated to create new samples and increase the number of abnormal instances. Meanwhile, normal residential data undergoes simple random undersampling, resulting in a more balanced ratio of positive and negative samples that enhances model training effectiveness. After balancing the data, there are a total of 4000 samples, with 2400 samples of standard electricity users and 1600 samples of non-standard electricity users.

Due to variations in electricity consumption patterns across industries, the data obtained through the user's smart meter shows significant differences in order of magnitude. As illustrated in Figure 7, it is evident that electricity consumption by industrial and commercial users is substantially higher than that of residential users. User 1182 and User 1922's electricity consumption curves offer further insight into this trend.

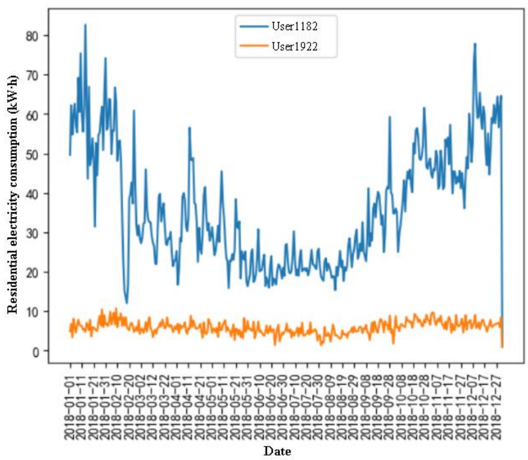


Figure 7: Display of Electricity Consumption Data for Users with Different Levels of Electricity Consumption

Thus, it is necessary to normalize data preprocessing for electricity consumption data processing to eliminate the impact of the magnitude gap on model training effectiveness[15]. In this study, maximum-minimum normalization was utilized to convert data into values ranging from 0 to 1 before model training. The normalization formula is presented below:

$$\text{x\_normalized}=\frac{x-x_{min}}{x_{max}-x_{min}} \tag{10}$$

After normalizing the data, the model's convergence is faster and the training of the network is improved.

B. Evaluation Criteria

For a binary classification problem, that is, classifying instances into positive (positive) or negative (negative) classes, the following four cases occur in actual classification: if an instance is positive and is predicted to be positive, that is, it is a True Positive (TP); if an instance is positive but is predicted to be negative, that is, it is a False Negative (FN); if an instance is negative but is predicted to be positive, that is, it is a False Positive (FP); and if an instance is positive but is predicted to be positive, it is a False Negative (FP)(False Negative); if an instance is a negative class but is predicted to be a positive class, it is a False Positive class FP (False Positive); if an instance is a negative class and is predicted to be a negative class, it is a True Negative class TN (True Negative)[16].

Accuracy refers to the percentage of correctly predicted samples out of the total number, as defined in the following formula:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{11}$$

Precision, also referred to as accuracy rate, serves as an evaluation metric for prediction outcomes. In the model prediction for the results of positive samples, the percentage of truly positive samples, the specific formula is as follows:

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% \tag{12}$$

Recall, also known as the retrieval rate, is an objective measure used to evaluate the original sample. It is calculated as the percentage of positive samples predicted correctly in the actual positive sample. The formula for Recall is as follows:

$$Recall = \frac{TP}{TP+FN} \times 100\% \tag{13}$$

The F1-Score is derived from the fact that precision and recall often pose a trade-off: high precision results in fewer false positives, yet a larger number of false negatives may occur; high recall results in fewer false negatives, yet a larger number of false positives. Therefore, it serves as an indicator of the proposed trade-off. The formula for Recall is as follows:

$$F1 = 2 \times \frac{Precision \times \text{Recall}}{Precision + \text{Recall}} \times 100\% \tag{14}$$

C. Results

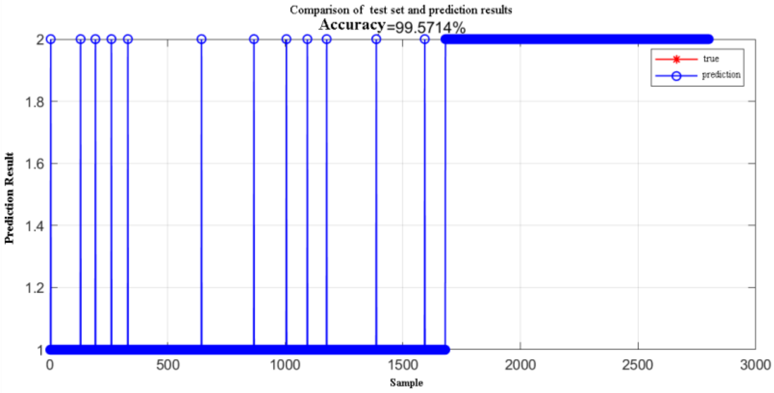
The pre-processed samples are put into the designed CNN-BiLSTM-Attention model for training, and the experimental results in the figure below are obtained.

70% of the dataset was selected as the training set to train the model. Figure 8(a) displays the model's predictions. The remaining 30% of the dataset was used as the test set to validate the model. Figure 8(b) shows the model's predictions, where '1' indicates normal electricity consumption and '2' indicates abnormal electricity consumption. The balance between positive and negative sample sizes is apparent, and the predicted labels are circular, indicating a minimal disparity between the predicted and true labels.

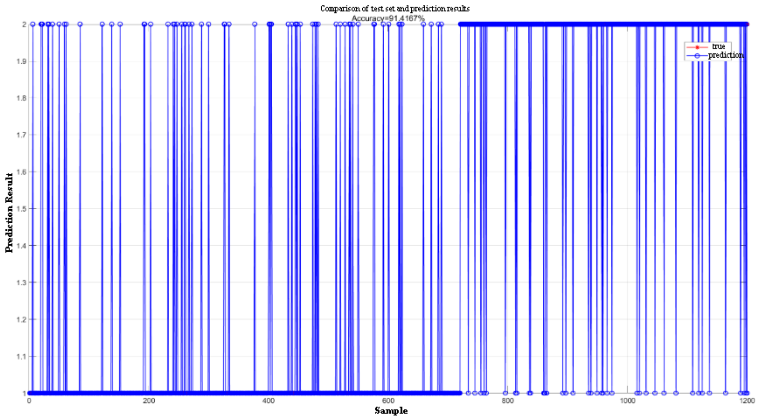
The following is a thorough assessment of the model's evaluation metrics. According to Figure 9, the confusion matrix displays the results of the test set. The model's accuracy rate is 91.42%, which demonstrates its capability of classifying samples. This high accuracy rate reflects the model's effectiveness in classifying all the samples. The model's precision is approximately 94.4%, denoting the proportion of correctly classified positive samples out of all positive samples. The high precision reflects the model's low likelihood of misclassifying positive samples, while the recall, at approximately 91.1%, describes the model's capability to accurately predict and capture positive samples. The high recall indicates fewer missed judgements in detecting positive samples. In summary, the model exhibits high accuracy and impressive overall performance. It excels at identifying normal electricity consumption samples, but its capacity to distinguish abnormal electricity consumption samples is relatively weak. A possible approach to enhancing the classification accuracy is to consider further optimization of the algorithm.

Comparing the test results of other models to those of the model presented in this paper on the same dataset reveals that the CNN-LSTM model outperforms the CNN model.

This finding indicates that the LSTM network can effectively extract temporal data features, and that the hybrid model is better suited to capturing inter-data features, thus improving the predictive capacity of the model. The CNN-BiLSTM model outperforms the CNN-LSTM model, demonstrating that the bidirectional LSTM coding can comprehensively capture the context features in between; the CNN-BiLSTM-Attention model proposed in this study yields the most optimal results compared to other models. The introduction of the Attention model leads to a more reasonable feature weighting, resulting in better prediction performance. Comparison of test results of each model are presented in table 1. The F1-Score reaches 92.72%, which combines precision and recall rates, confirms the effectiveness of the proposed model in this paper.



(a) Display of Prediction Effect of Training Set



(b) Display of Prediction Effect of Prediction Set

Figure 8: Display of Prediction Effect of Dataset

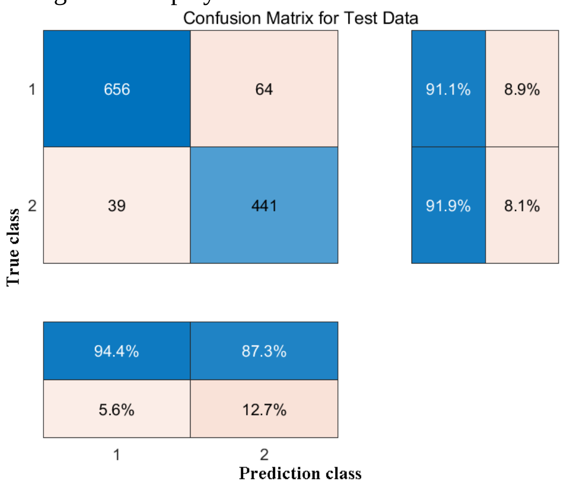


Figure 9: Display of Test Set Result

Table 1: Comparison of Test Results of Each Model

modelling	Accuracy	Precision	Recall	F1-Score
CNN	83.98%	86.90%	83.31%	85.07%
CNN-LSTM	87.46%	89.61%	86.64%	88.10%
CNN-BiLSTM	88.73%	92.75%	88.42%	90.53%
CNN-BiLSTM-Attention	**91.42%**	**94.39%**	**91.11%**	**92.72%**

V. Concluding Remarks

To address the issue of inefficient and costly detection of abnormal residential electricity consumption, this study proposes a hybrid CNN-BiLSTM-Attention model to extract real data from smart meters for identifying consumption anomalies. Experimental results demonstrate the model performs well with an accuracy rate, precision rate, recall rate, and F1-Score all exceeding 91%. However, in the future, individuals who steal electricity through abnormal means may become more difficult to detect. Therefore, the method of detecting abnormal electricity usage needs to be continuously updated. Conducting a cluster analysis of user electricity consumption behavior can help reveal patterns and improve both abnormality detection and the electric power system's service quality. Additionally, abnormal power consumption behavior can be correlated with climate data, users' power consumption behavior types, and other relevant information for analysis to enhance the detection strategy.

ACKNOWLEDGMENT

This paper is supported by the Science and Technology Project of the State Grid Shanxi Electric Power Company under grant, named Research and Application of Full Link Control and Tracking Technology for Sensitive Business Data Based on Fluorescent Labeling, No. 52051L230005.

References

[1] S. Meng, C. Li, C. Tian, W. Peng, and C. Tian, “Transfer learning based graph convolutional network with self-attention mechanism for abnormal electricity consumption detection,” Energy Reports, pp. 5647-5658.

[2] S. Y. Zhang, Q. L. Niu, and W. Liu, “Deep neural network detection method for abnormal electricity consumption by power users,” Power demand side management, vol. 25, pp. 82-87.

[3] J. Pan, L. W. Xia, and L. Li, “Abnormal Power Consumption Detection Based on Associative Power Flow Sensing and Gaussian Mixture Model,” Electric Power Construction, vol. 44, pp. 138-148.

[4] G. Q. Zhang, Y. Su, and N. W. Guo, “Abnormal Electricity Detection Method Based on Multi-Dimensional Unsupervised Learning,” Information Technology & Standardization, vol. 8, pp. 80-86.

[5] X. Q. You, and P. Gao, “Electricity consumption behavior detection method based on Hadoop and local anomaly factor,” Technology Innovation and Application, vol. 13, pp. 73-76, doi: 10.19981/j.CN23-1581/G3.2023.21.017.(in Chinese)

[6] H. Tao, and Y. Y. Zhuang, “Convolutional Neural Network-Based Anomaly Detection Method for Residential Power Consumption,” Scientific and Technological Innovation, pp. 92-98.

[7] K. Simonyan, and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” Computer Science, doi: 10.48550/arXiv.1409.1556.

[8] C. H. Chen, “Research on Anomaly Detection Method of Time Series Data in Cloud Data Center Based on Deep Learning,” Zhengzhou University, 10.27466/d.cnki.gzzdu.2021.000449.

[9] G. Wu, “Study on power data anomaly detection algorithm based on data intrinsic property and LSTM,” Wireless Internet Science and Technology, vol. 16, pp. 96-99+112.

[10] Y. F. Hu, Z. Y. Zhang, and T. Jin, “High Voltage Cable Ffault Diagnosis Method Based on UHF and CNN-LSTM-Attention Algorithm,” Computing Technology and Automation, vol. 42, pp. 31-38, doi: 10.16339/j.cnki.jsjsyzdh.202302007.

[11] X. Yang, “An Overview of the Attention Mechanisms in Computer Vision,” Journal of Physics : Conference Series, vol. 1693, pp. 012173.

[12] C. H. Wang, E. D. Wang, and X. F. Gao, “ANOMALY DETECTION ALGORITHM BASED ON CNN-BILSTM FEATURE FUSION,” Computer Applications and Software, vol. 39, pp. 272-277+340.

[13] B. A. Zhang, “Research on Anomaly Detection Method for User Power Data,” Zhejiang University, 2022, doi: 10.27461/d.cnki.gzjdx.2022.002625.

[14] J. Luengo, A. Fernández, and S. García, “Addressing data complexity for imbalanced data sets: analysis of SMOTE-based oversampling and evolutionary undersampling,” Soft Computing, vol. 15, pp. 1909-1936.

[15] M. Zhao, “Research on feature extraction and anomaly detection of user side power consumption data,” Kunming University of Science and Technology, doi: 10.27200/d.cnki.gkmlu.2020.000988.(in Chinese)

[16] T. Q. Chu, X. M. Chen, and X. C. Wang, “Text Classification Method of LSTM-Attention and CNN Fusion,” Journal of Anqing Normal University(Natural Science Edition), vol. 29, pp. 89-94, doi: 10.13757/j.cnki.cn34-1328/n.2023.03.016.