

Received 14 August 2023, accepted 29 August 2023, date of publication 13 September 2023, date of current version 20 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3314841

RESEARCH ARTICLE

Construction and Application of VS-DBN Anti-Theft Diagnosis Model Based on Neural Architecture Search

HECHUAN ZHANG¹, LEI WANG¹, WENHUI LI¹, YUJUN WANG², AND GE ZHANG³

¹Department of Power Supply Engineering, Baoding Technical College of Electric Power, State Grid Jibei Electric Power Company Limited Skills Training Center, Baoding 071051, China

²Marketing Department, State Grid Jibei Electric Power Company, Beijing 100053, China

³Sales Department, State Grid Tianjin Chengdong Electric Power Company, Tianjin 300400, China

Corresponding author: Hechuan Zhang (zhanghec103@163.com)

ABSTRACT Anti-theft work is important to the normal operation and safety guarantee of power system. But currently, many anti-theft diagnostic methods are no longer able to meet the demand for accurate detection. Therefore, the study introduced deep belief networks and improved them to address their shortcomings, designing a volume sparse deep belief network based on neural architecture search. On this basis, the construction of an anti-theft diagnosis model was studied and implemented. In the results, the hidden layers number obtained by the neural architecture search was all 3, and the fluctuation range of the number of neurons searched was [80100]. The true occurrence rate of the receiver operation characteristic curve after model processing has been effectively improved, and it is infinitely close to 1. At the same time, the region value of this curve is as high as 0.98, which is 0.52 higher than the untreated region value. In addition, the feature recognition accuracy of this model is stable at around 95%, and the highest accuracy reaches 98.26%. And the recognition loss of this model is around 0.2, with a minimum of only 0.11. This indicates that the anti-theft diagnosis model has excellent performance and can be well applied in practical anti-theft diagnosis. This provides a reliable safety guarantee for the operation of the power system.

INDEX TERMS Deep belief network, neural architecture search, anti-theft of electricity, sparse volume, particle swarm optimization.

I. INTRODUCTION

With residential electricity consumption gradually increasing, the coverage rate of the power grid has also increased. Illegal elements began to engage in electricity theft (ET) for profit, which greatly harmed the interests of the power company and caused significant safety hazards. Meanwhile, due to the continuous progress of technology, ET is becoming increasingly intelligent. It has evolved from traditional shunt and voltage sharing power theft to non-invasive power theft using high-frequency jammers. In response to this issue, domestic and foreign researchers have conducted a large amount of anti-theft of electricity (ATE) research. In the past, the proposed low-voltage substation ATE diagnosis method

is mainly based on the spatio-temporal correlation matrix method. This method uses Big data mining technology to mine the high-frequency data of smart meters, and builds the diagnosis model according to the obtained data. Then, the diagnostic model was used to estimate the amount of stolen electricity. This diagnostic method has improved the traditional method of ET diagnosis, mainly analyzing the spatiotemporal distribution characteristics of ET users effectively. At the same time, the starting time, ending time, and electricity consumption capacity of the ET information were judged. However, this diagnostic method has defects such as low diagnostic accuracy [1], [2]. The deep belief network (DBN) algorithm is a deep learning (DL) algorithm that has great advantages in mining deep attributes of data, and can use data analysis to obtain information about ET behind the data. The study introduced a DL algorithm based on neural

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

architecture search (NAS) for DBN to achieve ATE diagnosis in low voltage platform areas, to play a certain control effect on power anti-theft behavior [3]. The innovation of this article lies in the fact that the deep belief network based on volume sparsity (VS-DBN) model anti theft diagnosis model based on NAS mainly utilizes the idea of NAS to achieve accurate diagnosis of anti theft by automatically searching for the optimal neural network structure. At the same time, the model adopts Bayesian optimization to design the neural architecture search framework of the DBN model. During the search process, the model guides the search direction by evaluating the performance of each candidate network, ultimately finding the optimal network structure. In addition, the research will construct an anti theft diagnosis model based on the obtained anti theft data in the low-voltage substation area, in order to achieve anti theft diagnosis in the low-voltage substation area. The contribution of the NAS based VS-DBN anti theft diagnostic model lies in its ability to automatically discover and construct the optimal network structure, thereby improving the efficiency and performance of the model. Secondly, it can accurately identify and locate electricity theft behavior by learning the characteristics and patterns of the power system, and achieve the function of anti electricity theft diagnosis. Furthermore, the model adopts a deep confidence network as the basic structure, which has better expressive and generalization capabilities. Finally, it can help power management departments detect and prevent electricity theft in a timely manner, ensure the stable operation of the power system, and reduce economic losses and resource waste. The research content mainly includes four parts. Firstly, the application of DBN and methods for detecting ET are reviewed. Secondly, a detailed introduction was given to the construction method and application of the ATE diagnostic model based on NAS for VS-DBN model. The first section introduces the composition and structure of the traditional DBN model, as well as the VS-DBN model that introduces NAS on this basis. The second section discusses the construction method and implementation steps of the ATE diagnostic model. Then, experimental analysis was conducted on model's performance, and its effectiveness was verified. Finally, these results were discussed, and this model's superiority was demonstrated. At the same time, a summary of the existing problems in the research was made, and future prospects were proposed.

II. RELATED WORKS

DBN can play a significant role in detecting feature information. Traditional DBN neural networks have the problem of high-level data abstraction limitations. Scholars such as Naskath J proposed combining different neural networks to improve the model performance. This method can effectively solve the constrained problem of data abstraction and can be applied in various fields. This includes natural language processing, wireless networks, remote sensing applications, speech recognition, and more. Researchers such as Hu C H have found that current DL based residual service life

prediction methods can't reflect uncertain prediction results. Therefore, a RUL prediction model on the foundation of diffusion process and DBN is proposed. This model mainly extracts deep hidden features of monitoring signals through DBN, and selects highly inclined features as input data. This method can make RUL's prediction accuracy improved effectively [4], [5]. Wei et al. found that DL can effectively extract features and diagnose faults. Based on this, they introduced a nonlinear process fault diagnosis method based on DBN discarding. This network can extract abstract representations of nonlinear process data and construct a deep network to make real-time monitoring achieved. This model has strong generalization ability [6]. Chu et al. and other researchers believe that medium to long-term runoff forecasting is important to manage water resource, reservoir optimization and operation, etc. Therefore, they proposed a framework that integrates DBN, minimum multiplication contraction and selection operators, and bootstrap to make runoff forecasting stability improved. This framework can effectively improve the accuracy and stability of long-term runoff forecasting [7]. Zhang et al. and his team introduced a classifier based on interpretable fuzzy DBN to predict indoor user motion. This classifier has strong neural representation ability and uncertainty processing ability. And it can use fuzzy clustering algorithms to obtain fuzzy partitions on the training dataset, thereby forming interpretable antecedents for fuzzy rules. This classifier has achieved good classification performance on different datasets [8]. In order to obtain more interesting and compact summaries of automatic cricket videos, researchers such as Shingrakhia and Patel introduced a stacked gated recurrent neural network with attention modules for audio analysis. At the same time, a mixed rotation forest DBN was used to classify video scenes. The proposed machine learning method has an accuracy of up to 96.82% and effectively improves the salient features of video abstracts [9]. To better predict the stock prices, Zhang et al. put forward a method combining DBN and long short-term memory network (LSTM). Among them, DBN mainly learns potential feature representations from stock prices, while LSTM mainly handle long-term relationships in trading history. This model has significant predictive performance and can construct investment portfolios in the financial field [10].

At present, the ET occurs frequently. Many researchers have studied this problem. Tao and Michailidis and other scholars found that although smart meter technology has more advanced sensor communication capabilities, it has triggered the phenomenon of ET. Therefore, the team proposed a statistical technique that detects power theft attacks and identifies relevant power theft users through higher-order statistical information on power consumption. This statistical method can effectively identify power consumption and user information related to ET [11]. Researchers such as Saeed MS have found that ET and bill fraud can cause significant economic losses to distribution system operators. Therefore, an energy theft detection model has been developed to identify ET and fraudulent users in

distribution systems. Simultaneously, automatic interactive detection decision tree algorithm is used to classify such users. The accuracy of the proposed detection method reached 86.35%, and the area under the curve reached 0.927 [12]. Cui et al. believe that users' ET behavior poses a huge threat to the safe operation and economic benefits of the power company's system. Therefore, a two-step detecting method is put forward to identify ET users. Among them, the first step is to propose a convolutional autoencoder neural network model for identifying electricity stealing users. The second step is to use an improved regression algorithm on the foundation of limit gradient boosting algorithm to predict the potential electricity stealing of electricity stealing users. This strategy has extremely high prediction accuracy [13]. The team of Bhattacharjee et al. believes that the false power consumption data in the damaged meter report has had a negative impact on the normal operation of the smart grid. Therefore, a security detection technology based on fine-grained and coarse-grained anomalies has been introduced. This technology mainly uses anomaly detection to detect deviations and directional changes in the time series for ET identification. This technology can not only be used for identifying theft information, but also has good application effects in other fields [14]. Gupta et al. designed low-cost Tamper resistance smart meters to effectively detect and communicate power theft against the rampant problem of modern power theft. This method not only saves detection costs, but also has significant accuracy in detecting ET [15]. To better ensure the normal operation of power companies, researchers such as Ozcanli et al. used deep neural networks to predict ET in the power system and classify the relevant information of ET. This method effectively improves power system's security and improves power company's economic benefits [16]. The distribution department of the power company has energy fraud issues such as ET, billing errors, or tampering with electricity meters. Scholars such as Poudel S have proposed a machine learning methods for theft information detection, while classifying theft users. This anti-theft method has achieved good practical application results [17].

In summary, domestic and foreign researchers have conducted much research on ET detection and achieved corresponding results. However, few studies have combined DBN for predicting ET behavior. Therefore, DL algorithms based on DBN have been introduced for ATE diagnosis in order to achieve better diagnostic results.

III. CONSTRUCTION AND APPLICATION OF VS-DBN ANTI-THEFT DIAGNOSIS MODEL BASED ON NAS

The construction of ATE diagnostic model is crucial to power system's normal operation. This chapter introduces a NAS based VS-DBNATE diagnostic model based on the existing diagnosis of ET and ATE information. The study introduced the classic DBN model structure and introduced a method based on the combination of volume sparsity and NAS to optimize the DBN model. At the same time, the

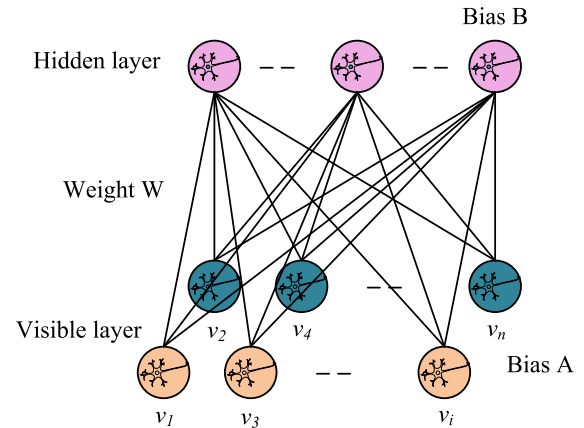


FIGURE 1. RBM model structure.

construction method and implementation steps of the ATE diagnostic model were elaborated in detail.

A. VS-DBN MODEL BASED ON NAS

To achieve ATE diagnosis in the low pressure plateau area, a DBN network was introduced for the construction of a diagnostic model. The core of DBN is the restricted Boltzmann machine (RBM). RBM is an undirected bipartite graph model. Its internal neural architecture and model training method are very suitable for modeling fMRI data, and its modeling effect is significantly better than other models [18]. RBM mainly has visible and hidden layers, where the former is the input layer and the latter is the output layer. The neurons in each layer are independent of each other, and the neurons present Boltzmann distribution, while the neurons between layers are fully connected. This model mainly utilizes the interaction between hidden layer variables and visible variables to express the distribution of input data. Figure 1 is this model's structure.

Equation (1) is the energy function calculation of the updated weight values for each layer of RBM.

$$E(v, h | \theta) = - \sum_{j=1}^n b_j h_j - \sum_{i=1}^m a_i v_i - \sum_{i=1}^m \sum_{j=1}^n v_i h_j W_{ij} \quad (1)$$

In Equation (1), m stands for neurons number in the visual layer. n stands for the hidden layer neurons number. The vectors v_i and h_j are neurons activation states in the visible and hidden layers, respectively. a_i and b_j represent neurons biases in the visible and hidden layers, respectively. $\theta = \{a_i, b_j, W_{ij}\}$ is a set of parameters. RBM is an energy based probability distribution model that can calculate the probability distribution of each neuron through the energy function in Equation (2).

$$P(v, h) = \frac{e^{-E(v, h | \theta)}}{Z} \quad (2)$$

In Equation (2), Z stands for the normalization coefficient, and the total distribution of probability must be 1. Combining

the structural properties of RBM and the obtained probability distribution, given layer neuron's probability distribution, the activation probability of hidden layer neurons can be calculated using Equation (3).

$$P(h|v) = \prod_{j=1}^m P(h_j|v) \tag{3}$$

Neuron's probability distribution in hidden layer is calculated using Equation (4).

$$P(h_j = 1|v) = \delta(\sum_{i=1}^n w_{i,j}v_i + c_j) \tag{4}$$

In Equation (4), δ is the sigmoid activation function. If hidden layer neuron's probability distribution is given, visible layer neuron's activation probability can be calculated using Equation (5).

$$P(v|h) = \prod_{i=1}^n P(v_i|h) \tag{5}$$

Neuron's probability distribution in visual layer is calculated using Equation (6).

$$P(v_i = 1|h) = \delta(\sum_{j=1}^m w_{i,j}h_j) \tag{6}$$

If a set of training samples is given, Equation (7) is its representation method.

$$S = \{v^1, v^2, \dots, v^{n_s}\},$$

$$v^i = (v_1^i, v_2^i, \dots, v_{n_v}^i)^T \quad (i = 1, 2, \dots, n_s) \tag{7}$$

RBM is mainly to make the distribution of fitted output characteristics consistent with S set. So the maximum Likelihood function can be obtained in Formula (8).

$$\Upsilon_{\theta,S} = \prod_{i=1}^{n_s} P(v^i) \tag{8}$$

Because the results of the maximum Likelihood function obtained by the gradient derivation method are too redundant, the Gibbs sampling method is used to approximate the data [19]. This method mainly involves sampling and iterating sequentially according to the probability distribution, and ultimately obtaining the probability distribution of each variable. When the sampling frequency reaches its limit, the variable will converge based on probability. Equation (9) stands for the specific sampling process.

$$h_0 \sim P(h|v_0), v_1 \sim P(v|h_0), h_1 \sim P(h|v_1),$$

$$v_2 \sim P(v|h_1), \dots, v_{k+1} \sim P(v|h_k) \tag{9}$$

If there are too many sampling steps, it will lead to a decrease in training efficiency. Therefore, a study has proposed an algorithm based on Contrast Divergence (CD) to train the data. This algorithm can obtain the maximum approximation of the input sample through only K (K = 1) steps, and Equation (10) is its sampling process.

$$h_{t-1} \sim P(h|v_{t-1}), v_t \sim P(v|h_{t-1}) \tag{10}$$

DBN can be seen as a stack of multiple RBMs, with the top layer being undirected and the remaining layers being directed. In the DBN model, the input layer from the second

RBM to the last can be considered as the hidden layer. It can be understood that the DBN model has one input layer and multiple hidden layers. In addition, the adjacent two layers can be considered as one RBM [20], [21]. In the structural design of the DBN model, the number of nodes in the input and hidden layers has a significant impact on the experimental results. In the research process based on DBN anti theft model, the number of input layer nodes is the number of existing influencing factors. Hidden layer neurons can represent complex nonlinear relationships. If the number of neurons in the hidden layer is too small, it will lead to incomplete Data modeling. If the number of neurons in the hidden layer is too large, it will lead to over fitting problems, which will lead to no ideal experimental results. So, the question of how many neurons should be set in the hidden layer has a significant impact on the experimental results. In this article, experimental methods are proposed to design the optimal structure of DBN. For the input layer, the number of input neurons is fixed to the number of existing influencing factor elements, and factors that are not responsive to the results are removed. For the hidden layer, first consider a DBN model with one hidden layer, and set the number of hidden layer neurons to five different levels: 4, 8, 12, 16, and 20. The reason for doing this is that compared to the number of input nodes, the network prediction effect is more sensitive to changes in the number of hidden layer neurons. Through experimental results, the number of hidden layer neurons corresponding to the highest classification accuracy is found, which is the experimental process for a hidden layer. Afterwards, a new hidden layer is added, and the impact of changes in the number of nodes in the new hidden layer on classification accuracy is examined to determine the optimal number of nodes. As the number of nodes in the hidden layer changes, the number of neurons in the hidden layer corresponding to the highest classification accuracy is found. Calculate the classification accuracy of models with single-layer and multi-layer hidden layers separately, find the direct relationship between the number of hidden layers and the change in classification accuracy, and determine the number of hidden layers. At the same time, the Activation function based on the DBN anti stealing model is the Sigmoid function, which is mainly used to perform nonlinear transformation on the output of the hidden layer to increase the expression ability and fitting ability of the model. Figure 2 is the DBN structure.

DBN mainly uses CD to train each RBM one by one. The output of each layer will serve as the input of the next layer, which is also known as the greedy layer by layer pre training algorithm. As layer number in this model increases, the error information and noise in each layer will be filtered out, making the features more abundant. Due to the small number of samples used for modeling volume images, each sample has a large dimension. Moreover, the neural architecture of the model set based on experience has shortcomings such as insufficient GPU memory and low reliability of the constructed model. Therefore, the study introduced DBN based on volume sparsity for improvement [22]. Meanwhile,

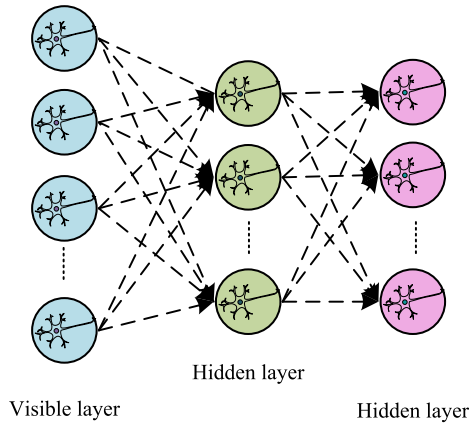


FIGURE 2. Structure of DBN model.

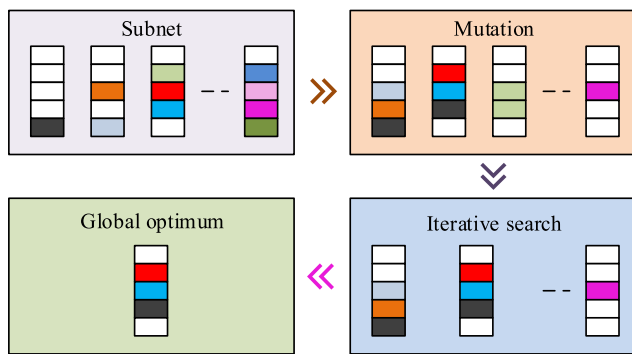


FIGURE 3. Schematic diagram of NAS operation.



FIGURE 4. NAS structure.

traditional DBN models have problems such as complex structure, high manual parameter tuning time, long training time, and unreliable results. Therefore, the study introduced NAS to optimize the architecture of the model, resulting in the generation of a NASbased DBN model. Figure 3 is a schematic diagram of NAS operation.

In Figure 3, NAS mainly searches for the optimal model structure from a specified set range in some way. The composition of NAS mainly includes search strategy, model performance evaluation, and search space. Among them, the search space mainly defines all possible architectures that may exist in the model. It can concatenate different neural architectures in the same DL model and determine the position and parameters of each structure. Search strategy refers to the optimal measure to define the model architecture. It needs to consider factors such as computing resources and search time. The performance evaluation of the model mainly involves dividing the data of the optimal model architecture into test sets and training sets for verification. Figure 4 is a schematic diagram of NAS.

The research adopts Bayesian optimization algorithm to design the neural architecture search framework of DBN model, in order to improve the performance and generalization ability of the model. Compared with Particle swarm optimization algorithm, Bayesian optimization can usually find the global optimal solution with less evaluation times, while Particle swarm optimization algorithm may require more iterations. This is because Bayesian optimization can infer from existing evaluation results and selectively evaluate parameter points with high potential. Particle swarm optimization usually needs to search the whole search space globally until the optimal solution is found. In addition, Bayesian optimization is relatively more effective in dealing with high-dimensional problems, as it can utilize prior information to guide the search process. PSO may face the challenge of dimensional disaster in high-dimensional problems, where increasing the dimensionality of the search space can lead to a significant decrease in search efficiency [23]. Compared to Tree structured Parzen Estimator (TPE), Bayesian optimization has certain advantages in search efficiency and robustness, and is suitable for complex optimization problems. TPE is only suitable for simple optimization problems and small search spaces [24]. The design steps of the DBN neural architecture search framework based on Bayesian optimization are: first, define the neural architecture search space of the DBN, including the number of layers of the network, the number of nodes in each layer, Activation function, etc. The definition of this search space directly affects the efficiency and results of the search process. Next, determine an evaluation metric to measure the performance of each DBN architecture. Common indicators can be classification accuracy, Mean squared error, etc. At the same time, it is necessary to select an appropriate prior distribution for each architecture parameter. After determining the range of parameter values and possible distribution characteristics, Gaussian distribution is used for prior distribution. Then, use Bayesian optimization methods for sampling and evaluation. Select a set of DBN architecture parameters for model training and evaluation based on prior distribution and evaluation indicators. Subsequently, by evaluating the results of the indicators, each parameter group is sorted and compared. Update the posterior distribution of parameters based on both sampling and evaluation results. In this way, the distribution of parameters can be adjusted based on existing observation data to better guide the next sampling process. Finally, the process of sampling, evaluating, and updating the posterior distribution is repeated until the stop condition is reached. The stopping condition can be reaching the maximum number of iterations, reaching a certain performance threshold, etc. By using Bayesian optimization for DBN neural architecture search, the optimal neural network architecture can be found with relatively few training and evaluation times. This can accelerate the process of model design and improve model performance. In addition, the effectiveness of Bayesian optimization mainly depends on the definition of search space and the selection of evaluation indicators, so adjustments and

optimizations need to be made based on specific tasks and datasets.

B. CONSTRUCTION AND APPLICATION OF ATE DIAGNOSIS MODEL BASED ON VS-DBN

To diagnose ATE in the low-voltage substation area, data mining technology should first be used to mine and process ATE data. This technology first analyzes the AC frequency of users who steal electricity, then analyzes voltage fusion, and manages the AC frequency load through power fusion method. From this, the parameter model of ATE data in Equation (11) can be obtained.

$$x_n(t) = \sum_{u=0} n(t) + \sqrt{p_u} g_u d_u(t) \quad (11)$$

In Equation (11), $x_n(t)$ stands for the communication frequency at the moment of power theft t . u is the current of the transmission and distribution network. p_u stands for the voltage distribution parameter. g_u stands for the load parameters of the DC microgrid. d_u stands for the bus capacitance. $n(t)$ is the amount of ET data of user n at time t . The ATE data parameter model can detect and analyze the active power of ET. And it can diagnose bus differences in the transmission and distribution network. Thus, the correlation between the AC/DC bus and three-phase AC current was obtained to achieve differential identification of anti-theft data. In addition, this model can combine bus differences and DC microgrid spread spectrum sequences to obtain the ambiguity parameters of ATE data. In addition, this model can also achieve the fusion of ATE phase difference and bus capacitance. Equation (12) stands for its fusion data.

$$y_u = r_1 g_{u1} + r_2 g_{u2} \quad (12)$$

In Equation (12), y_u stands for the mined ATE data. r_1 stands for the output power change of DC microgrid. g_{u1} stands for the change value of ATE data. g_{u2} is the AC side voltage of ATE. r_2 is the ambiguity parameter of ATE data. The parameter model of ATE data can effectively obtain information such as theft time and electricity consumption of users who steal electricity. The study will construct an ATE diagnostic model using the obtained ATE data from low-voltage substation area. Firstly, ATE data will be divided into a training and a sample datasets. Among them, the training dataset is applied to ATE training diagnostic model construction, while the sample dataset is mainly used to establish ATE sample diagnostic model. Finally, it is necessary to combine two diagnostic models to form the final ATE diagnostic model. The study first used DBN to construct an ATE sample diagnostic model. Due to the presence of invalid data in the ATE sample data, it may have a negative impact on ATE sample diagnostic model construction, so it is necessary to remove it. Invalid ATE data is generally located within the smallest number of nodes in each layer. Therefore, it is necessary to filter out the smallest nodes number for deletion. Equation (13) is the ATE sample diagnostic model

constructed using an effective ATE sample dataset.

$$X_\omega = \sum_{s=1}^{\gamma} (A_k + B_l) R_s \quad (13)$$

In Equation (13), X_ω stands for the coefficient of ATE sample diagnostic model. s stands for the nodes number. γ is the total nodes number. A_k stands for ATE sample data on the input layer. B_l stands for ATE sample data on the hidden layer. $R_s r_1 r_1$ stands for ATE sample data on the output layer. By constructing this model, it is possible to effectively diagnose ATE sample dataset. Subsequently, the ATE training dataset was trained using NAS-based VS-DBN. Its main function is to enable the model to fully reflect the behavioral characteristics of ET users. The training problem of the DBN model is actually the problem of training its constituent structure RBM. Figure 5 shows the specific training process.

The ATE data in the training dataset is larger than sample dataset, which results in slower training speed and lower diagnostic accuracy of ATE training diagnostic models. Therefore, a training coefficient was introduced in the study to eliminate errors. Among them, Equation (14) stands for the ATE trained diagnostic model.

$$Y_a = \sum_{a=1}^n \Delta s_a \quad (14)$$

In Equation (14), Y_a stands for the parameters of the ATE training diagnostic model. a stands for the training coefficient. Δs_a stands for the training data of ATE. Finally, the ATE sample diagnostic model was fused with ATE training diagnostic model to form a low-voltage station area ATE diagnostic model, represented by Equation (15).

$$Z = \frac{\sum_{q=1}^n X_e Y_a}{\delta} \quad (15)$$

In Equation (15), δ stands for the model coefficient that combines ATE sample diagnostic model with ATE training diagnostic model. q stands for the constant for ATE diagnosis. Figure 6 shows the diagnostic process of the low-voltage substation area ATE diagnostic model based on NAS VS-DBN.

In Figure 6, in the low-voltage substation ATE diagnosis model based on DBN, due to equipment failures, network interruptions, and other reasons during ATE data transmission, some ATE data may be missing, thereby causing interference to ATE diagnosis. Therefore, firstly, the input ATE data is cleaned with a linear difference method. It mainly calculates the missing values of ATE data through DBN in Equation (16).

$$W = \frac{\sqrt{o_1 - h_1}}{\lambda} \quad (16)$$

In Equation (16), λ stands for the missing coefficient, o_1 stands for the cleaning weight of ATE data, and h_1 stands for the estimated ATE power consumption. By calculating the

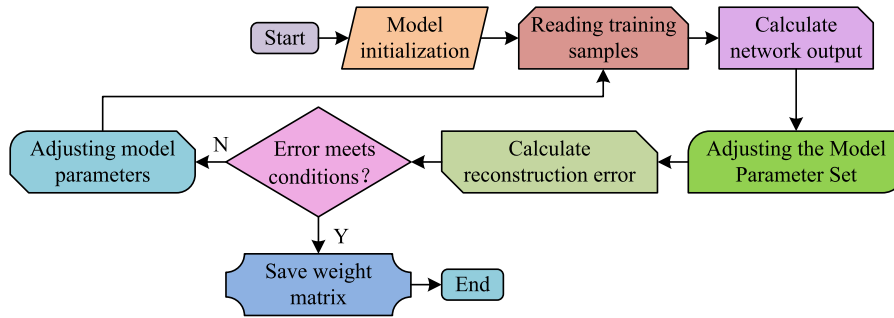


FIGURE 5. The training process of VS-DBN model.

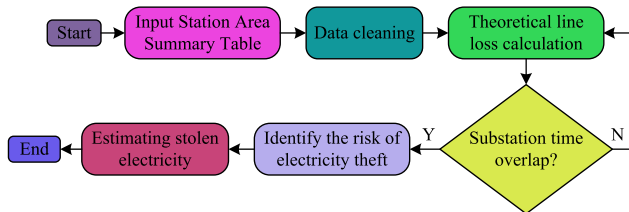


FIGURE 6. The diagnostic process of the anti-theft diagnostic model.

missing values of ATE data, effective cleaning processing can be carried out on ATE data. Subsequently, identification of the risk of ET was carried out. When a user steals electricity in a low-voltage substation area, it will cause an increase in line loss. Therefore, ET in the low-voltage substation area can be determined by detecting changes in line losses. The specific operation is to first record the line loss value before the occurrence of ET in the low-voltage substation area. Then, the line loss value after the occurrence of ET was recorded. If the difference in line loss value exceeds the standard range, the analysis of ET can be carried out based on this difference. When analyzing the ET behavior of users in the low-voltage substation area, the ET behavior can be analyzed based on abnormal changes in the electricity meter voltage, rated power, current, etc. of the electricity stealing user. Therefore, when analyzing users' ET behavior, it is necessary to identify the spatial dimension. Its main purpose is to monitor the voltage value of the low-voltage household meter and analyze the changes in voltage value for ET behavior. After analyzing the user's ET behavior, it is necessary to estimate the amount of electricity stolen by the user. Estimating the user's stolen electricity can ensure the normal operation of ATE diagnosis. The amount of electricity stolen by users is mainly determined by the time and power of the theft. And ET is directly proportional to the theft time and power, and the theft time is calculated from the abnormal voltage value of the theft meter. Equation (17) is the calculation of users stealing electricity.

$$G = \sqrt{3}T_1P_1 \quad (17)$$

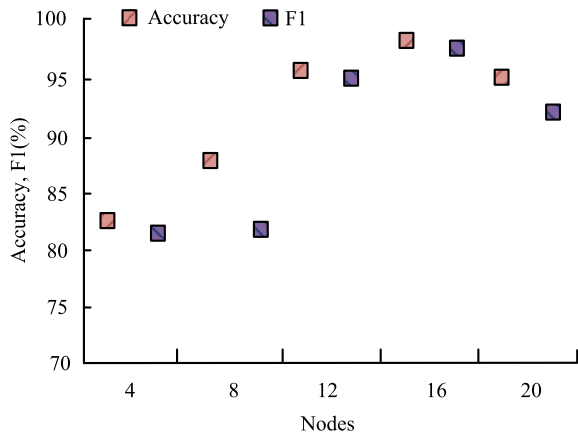
In Equation (17), AA stands for the stolen electricity of low-voltage users. BB stands for the stealing power of the user's

meter. CC stands for the theft time of the user. Through the above diagnostic process, ATE diagnosis within the low-voltage substation area can be achieved.

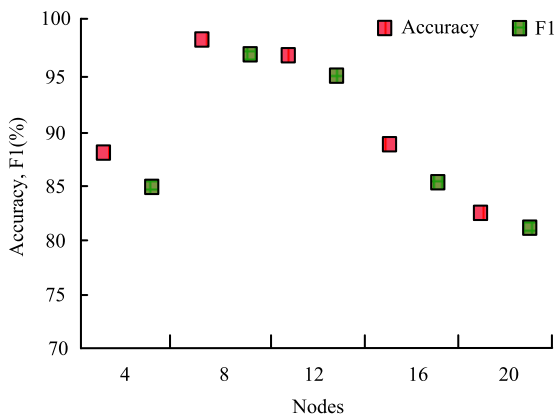
IV. APPLICATION ANALYSIS OF VS-DBNATE DIAGNOSTIC MODEL BASED ON NAS

To obtain the optimal ATE diagnostic model, different hidden layer nodes were used for experimental verification. In the specific experimental process, factors related to ATE need to be input into the model. They include daily, peak, valley, and cumulative electricity consumptions, as well as temperature. Among them, the input neurons is fixed at 8, while the hidden layer neurons is taken as 4, 8, 12, 16, and 20, respectively. Figure 7 show the classifying results of different hidden layer nodes in different hidden layers. When in the first hidden layer's nodes number of DBN is 16, its classifying accuracy and F1 value are the highest, reaching 97.25% and 96.58% respectively. Meanwhile, when the second hidden layer's nodes number is 8, its classifying accuracy and F1 value are the highest, reaching 98.64% and 96.97% respectively. The optimal combination of DBNATE diagnostic models is 16 and 8 nodes in the first and second hidden layers, respectively.

The experiment continued to train the VS-DBN model, and the framework of the experiment was mainly built on the GTX1080Ti four card server. When using NAS for search, first set the search range of hidden layer neurons to [10200], the search range of hidden layer layers to [2,10], and the number of iterations to 6. After optimizing the model using NAS in the experiment, the optimal model architecture found will be used for the construction of the VS-DBN model. Among them, the parameter of the VS-DBN model is set to weight_Decay = 0.001, batch_Size = 10, and the Learning rate is 0.001. Figure 9 shows the error curves obtained after training the group VS-DBN and CNN. In Figure 9, as training iteration increases, the error curve of VS-DBN shows a downward trend. It tends to stabilize at around 30 iterations, and the error stabilizes around 9.000e-3. The error of CNN only stabilizes after about 40 iterations, and the error is even higher. This indicates that VS-DBN has better performance advantages.



(a) Classification results of the first hidden layer



(b) Classification results of the second hidden layer

FIGURE 7. Classification results of different hidden layer nodes in different hidden layers.

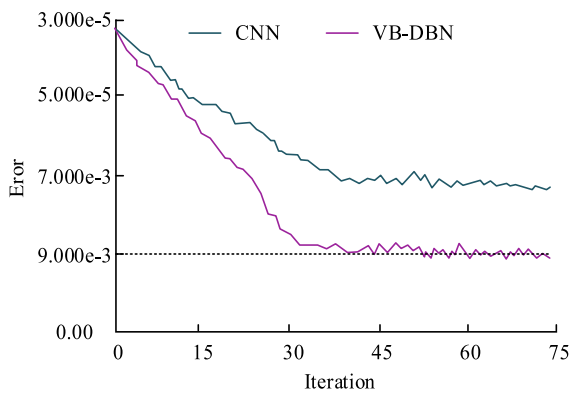
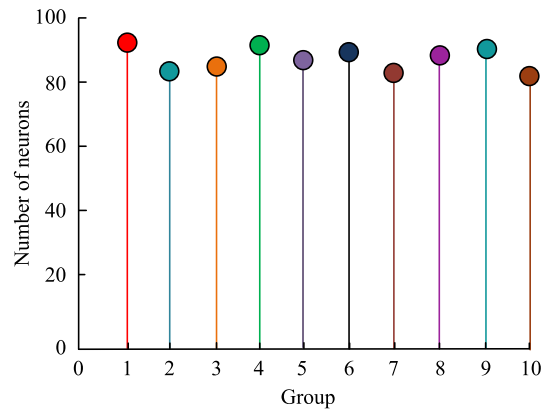
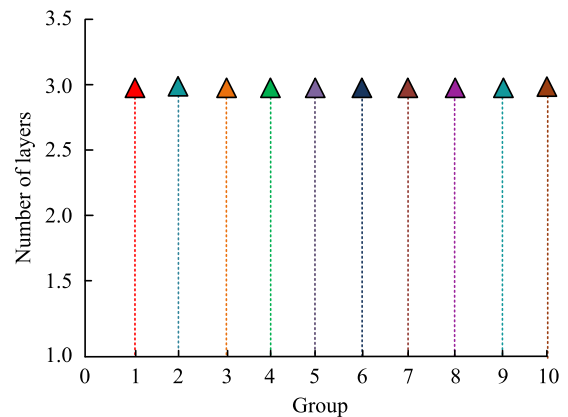


FIGURE 8. The error curve obtained after training.

To make VS-DBN’s optimal architecture obtained, the experiment used NAS to search for it. A total of 10 NAS searches were conducted, each of which took 6 hours. Figure 9 shows the search results. The neurons and layers number in the hidden layer using NAS search maintain high consistency and robustness. The 10 layers found by NAS are all 3, indicating that the optimal architecture for VS-DBN has an implied layer count of 3. Meanwhile, the fluctuation



(a) Number of neurons searched by neural architecture



(b) Number of layers for neural architecture search

FIGURE 9. Sets of neural architecture search results.

range of the number of neurons searched by NAS is [80,100]. The maximum layers number is 98, the minimum layers number is 80, and the standard deviation of the ten data is 5.47. Therefore, the optimal architecture of VS-DBN was ultimately set to three hidden layers, each with 91 neurons. This architecture will be used for ATE diagnosis in this study.

To verify VS-DBN’s effectiveness when extracting ATE data feature, the study first used t-distributed stochastic neighbor embedding (t-SNE) graphs to transform ATE data into visual data, and used scatter plots for distribution analysis. The experimental dataset consists of load data from 1580 electricity users in a certain power company’s electricity metering system from March 2019 to October 2019. It mainly comes from four industries with high electricity theft rates, including civil engineering and construction industry, structural metal product manufacturing industry, gypsum cement and similar product manufacturing industry, plastic parts and other plastic products industry. Figure 10 shows the unprocessed data’s t-SNE diagram in this experiment. The distribution of original ET data is relatively chaotic and cannot distinguish the types of data, so effective feature extraction is needed.

To verify the superiority of NAS based VS-DBN model in data processing, CNN was selected for experimental comparison. At the same time, 60% of the samples were

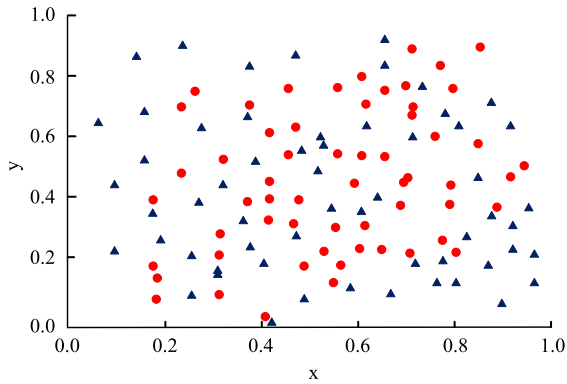
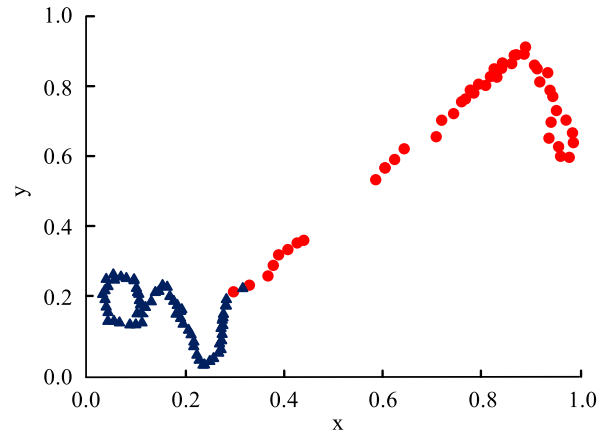
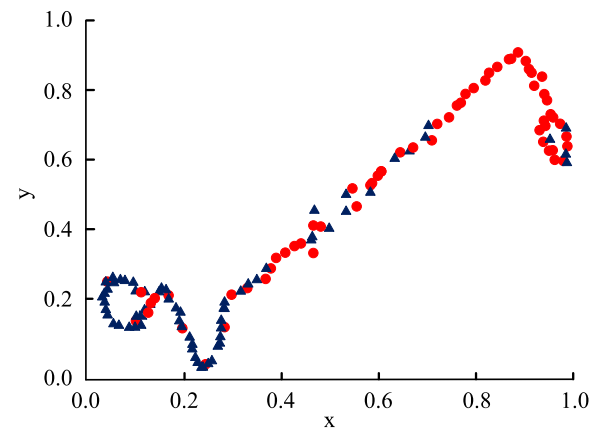


FIGURE 10. Original anti-theft data t-SNE graph.



(a) T-SNE diagram processed by VS-DBN model



(b) T-SNE diagram processed by CNN model

selected from the dataset as the training set for the model, and 40% of the samples were selected as the testing set. Figure 11 shows the t-SNE diagram of data from different models. The processed data's t-SNE graph by this model was used to classify different types of ATE data. Among them, the VS-DBN based on NAS has the best classifying performance, with a classifying accuracy of up to 98.42%. However, CNN has relatively poor classifying performance, with a classification accuracy of only 75.46%, a decrease of 22.96% compared to VS-DBN. Explanation: VS-DBN based on NAS can effectively process ATE data.

The study continued to compare the performance of commonly used support vector regression machine (SVRM) [25], general regression neural network (GRNN) [26], and abstract data restoration network (PDRN) [27] with VS-DBN models based on neural architecture search. The experiment tests the accuracy of identifying theft data for these three models, with evaluation indicators including accuracy, recall, and F1 value. The experimental results are shown in the figure 12. From the graph 12, it can be seen that the accuracy, recall, and F1 value of the method used in the study are significantly superior to other methods. Among them, the VS-DBN method based on neural architecture search has an accuracy of 95.68%, a recall rate of 94.37%, and an F1 value of 96.06%. The accuracy, recall, and F1 values of the SVRM method are 89.35%, 83.69%, and 84.31%, respectively. The three indicators of the GRNN method are 83.58%, 74.09%, and 77.43%, respectively. The three indicators of the PDRN method are 75.91%, 70.14%, and 81.07%, respectively. The VS-DBN method based on neural architecture search has significant advantages in identifying theft data, with higher recognition accuracy and better application in practical anti theft detection.

The study further validated the model's feature extraction and recognition capabilities using Receiver Operating Characteristic (ROC) graphs. Among them, Figure 13 shows the curve before and after data processing using NAS based VS-DBN. Compared with the unprocessed data curve, the true occurrence rate of ROC curve after model processing has been effectively improved, and is infinitely close to 1. At the same time, the regional value of ROC curve is as high as 0.98,

FIGURE 11. T-SNE plots of data from different models.

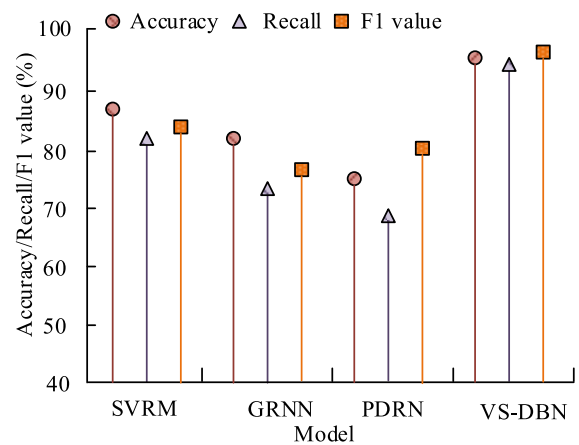


FIGURE 12. Accuracy, recall, and F1 values of different methods.

an increase of 0.52. This indicates that NAS-based VS-DBN can effectively extract feature information from ATE data.

This research continues to validate the performance advantages of the NAS based VS-DBN model, using accuracy maps and loss maps for result validation. The experiment selected

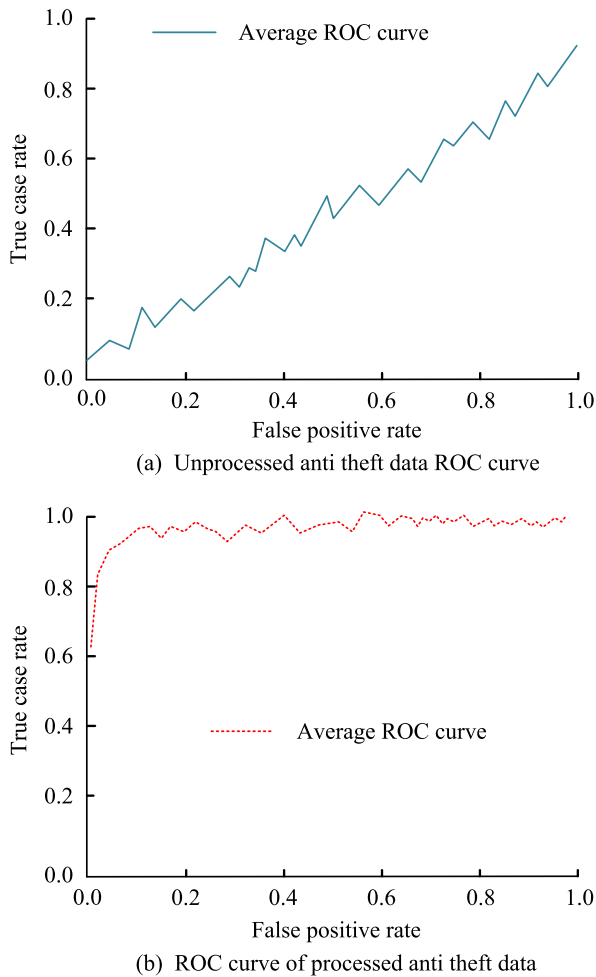


FIGURE 13. ROC curve of the model before and after data processing.

CNN for comparison, and Figure 14 shows the accuracy and loss maps of anti-theft data recognition after model processing. The feature recognition accuracy of this model is stable at around 95%, and the highest accuracy reaches 98.26%, which is 26.14% higher than CNN. Meanwhile, the recognition loss of this model is around 0.2, with a minimum of only 0.11, which is 0.32 lower than CNN. This indicates that VS-DBN in NAS has higher feature recognition accuracy.

To verify the NAS effectiveness based VS-DBN ATE diagnostic model, this study compared it with traditional low-voltage substation ATE diagnostic models based on big data mining and low-voltage substation ATE diagnostic models based on spatiotemporal correlation matrix. Firstly, identification and analysis were conducted on suspicion of ET among users. A power supply analysis was conducted on two low-voltage substation areas of a certain power supply company, with a total of 10 users in each substation area. And taking into account factors such as current, voltage, and power of user meters, the fluctuation threshold of line loss is set to 13. Those who exceed this threshold range are considered theft users. According to the actual investigation, the number of ET users in Zone 1 is 4, and the ET users number

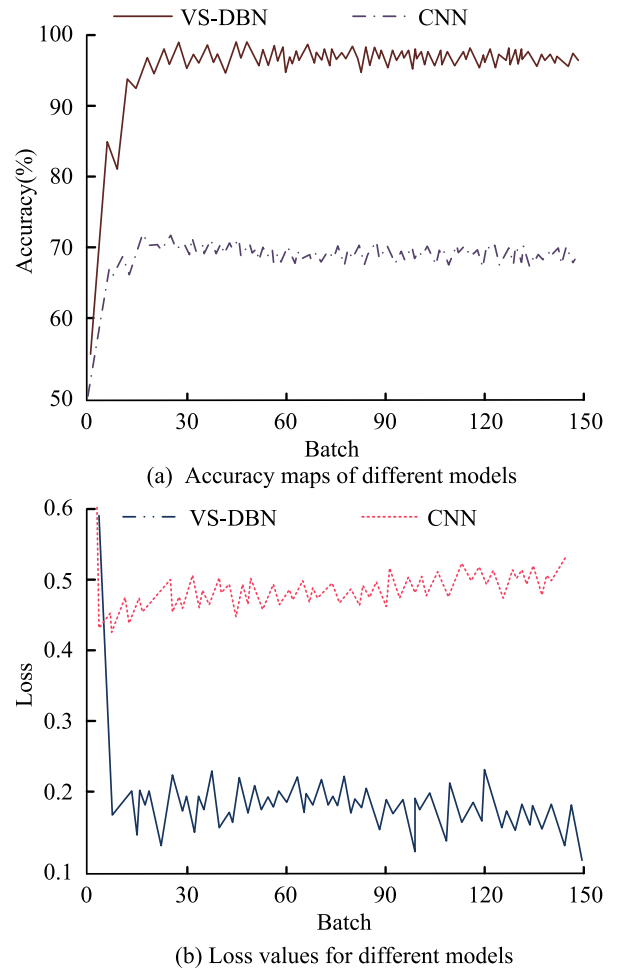


FIGURE 14. Accuracy diagram and loss diagram of anti-theft data recognition after model processing.

in Zone 2 is 8. Figure 15 shows the line loss determination results for each substation area. The ATE diagnostic method based on VS-DBN can accurately diagnose the electricity stealing users in each substation area. The maximum line loss fluctuation rate diagnosed by this model for station area 1 is 25%, and the corresponding user for this value is user 4. At the same time, the maximum line loss fluctuation rate diagnosed by the model for station area 2 is 22%, and the corresponding user for this value is user 8. In addition, the fluctuation rate of line loss detected by the model for other non electricity stealing users' meters is lower than the set threshold, which is completely consistent with the actual situation. Based on the spatiotemporal correlation matrix, the ATE diagnostic model for low-voltage substation area has listed three electricity users as ET users in the diagnostic data of substation area 1. A diagnostic error occurred in the diagnostic data of station area 2. The data diagnosed by the ATE diagnostic model based on big data mining in the low-voltage substation area are all below the set threshold, and it fails to diagnose the actual ET users. The ATE diagnostic method based on VS-DBN is significantly superior to the other two

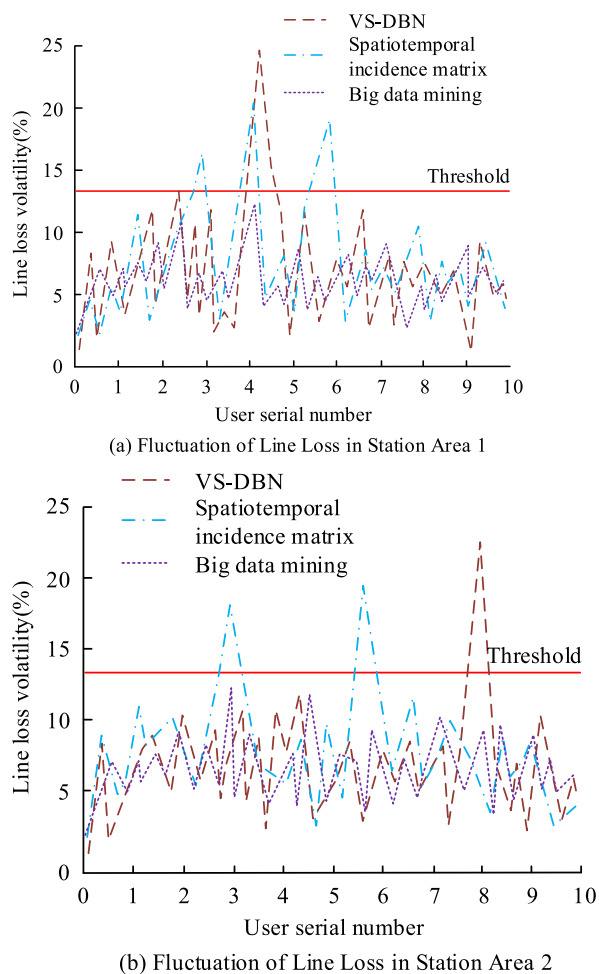


FIGURE 15. Line loss determination results for each substation area.

diagnostic methods, and can accurately diagnose ET users in each substation area.

V. CONCLUSION

ATE diagnosis is significant to maintain power system’s normal operation. The traditional ATE diagnostic method has problems of computational complexity and low diagnostic accuracy. Therefore, the study improved on the basis of DBN and designed a NAS based volume sparse DBN. At the same time, an ATE diagnostic model was constructed using this network. The optimal combination of ATE diagnostic models is 16 nodes in the first hidden layer and 8 nodes in the second hidden layer. As the training iterations increase, the error curve of this model shows a downward trend and tends to stabilize after about 30 iterations. In the t-SNE graph, the classifying accuracy of NAS based VS-DBN model is as high as 98.42%, which is 22.96% higher than CNN. In actual ATE diagnosis, the maximum line loss fluctuation rate detected by this model for station area 1 is 25%, and the corresponding user for this value is user 4. At the same time, the maximum line loss fluctuation rate diagnosed by the model for station area 2 is 22%, and the corresponding user for this value is

user 8. In addition, the fluctuation rate of line loss detected by the model for other non electricity stealing users’ meters is lower than the set threshold, which is completely same as actual situation. ATE diagnostic model constructed by NAS based VS-DBN has good practical application results. But the adaptability of the methods proposed by the research institute is not enough. In some complex environments, such as extremely harsh weather, its diagnostic accuracy is easily affected. Therefore, further improvement is needed to enhance the diagnostic stability of the model.

REFERENCES

- [1] Z. Su, J. Yang, P. Li, J. Jing, and H. Zhang, “A precise method of color space conversion in the digital printing process based on PSO-DBN,” *Textile Res. J.*, vol. 92, nos. 9–10, pp. 1673–1681, May 2022.
- [2] P. Lu, Z. Xu, T. Hong, Y. Wu, and D. Li, “Reliability evaluation of concrete-filled steel tube arch bridge based on DBN-PSOSA hybrid algorithm,” *Structural Concrete*, vol. 22, no. 4, pp. 2309–2321, Aug. 2021.
- [3] X. He and J. Ma, “Weak fault diagnosis of rolling bearing based on FRFT and DBN,” *Syst. Sci. Control Eng.*, vol. 8, no. 1, pp. 57–66, Jan. 2020.
- [4] J. Naskath, G. Sivakamasundari, and A. A. S. Begum, “A study on different deep learning algorithms used in deep neural nets: MLP SOM and DBN,” *Wireless Pers. Commun.*, vol. 128, no. 4, pp. 2913–2936, Feb. 2023.
- [5] C.-H. Hu, H. Pei, X.-S. Si, D.-B. Du, Z.-N. Pang, and X. Wang, “A prognostic model based on DBN and diffusion process for degrading bearing,” *IEEE Trans. Ind. Electron.*, vol. 67, no. 10, pp. 8767–8777, Oct. 2019.
- [6] Y. Wei and Z. Weng, “Research on TE process fault diagnosis method based on DBN and dropout,” *Can. J. Chem. Eng.*, vol. 98, no. 6, pp. 1293–1306, Jun. 2020.
- [7] H. Chu, J. Wei, and Y. Jiang, “Middle- and long-term streamflow forecasting and uncertainty analysis using lasso-DBN-bootstrap model,” *Water Resour. Manage.*, vol. 35, no. 8, pp. 2617–2632, Jun. 2021.
- [8] X. Zhang, F.-L. Chung, and S. Wang, “An interpretable fuzzy DBN-based classifier for indoor user movement prediction in ambient assisted living applications,” *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 42–53, Jan. 2020.
- [9] H. Shingrakhia and H. Patel, “SGRNN-AM and HRF-DBN: A hybrid machine learning model for cricket video summarization,” *Vis. Comput.*, vol. 38, no. 7, pp. 2285–2301, Jul. 2022.
- [10] X. Zhang, N. Gu, J. Chang, and H. Ye, “Predicting stock price movement using a DBN-RNN,” *Appl. Artif. Intell.*, vol. 35, no. 12, pp. 876–892, Oct. 2021.
- [11] J. Tao and G. Michailidis, “A statistical framework for detecting electricity theft activities in smart grid distribution networks,” *IEEE J. Sel. Areas Commun.*, vol. 38, no. 1, pp. 205–216, Jan. 2020.
- [12] M. S. Saeed, M. W. B. Mustafa, U. U. Sheikh, A. Khidrani, and M. N. H. Mohd, “Electricity theft detection in power utilities using bagged CHAID-based classification trees,” *J. Optim. Ind. Eng.*, vol. 15, no. 2, pp. 67–73, 2022.
- [13] X. Cui, S. Liu, Z. Lin, J. Ma, F. Wen, Y. Ding, L. Yang, W. Guo, and X. Feng, “Two-step electricity theft detection strategy considering economic return based on convolutional autoencoder and improved regression algorithm,” *IEEE Trans. Power Syst.*, vol. 37, no. 3, pp. 2346–2359, May 2022.
- [14] S. Bhattacharjee, V. P. K. Madhavarapu, S. Silvestri, and S. K. Das, “Attack context embedded data driven trust diagnostics in smart metering infrastructure,” *ACM Trans. Privacy Secur.*, vol. 24, no. 2, pp. 1–36, May 2021.
- [15] A. K. Gupta, A. Routray, and V. A. Naikan, “Detection of power theft in low voltage distribution systems: A review from the Indian perspective,” *IETE J. Res.*, vol. 68, no. 6, pp. 4180–4197, Nov. 2022.
- [16] A. K. Ozcanli, F. Yaprakdal, and M. Baysal, “Deep learning methods and applications for electrical power systems: A comprehensive review,” *Int. J. Energy Res.*, vol. 44, no. 9, pp. 7136–7157, Jul. 2020.
- [17] S. Poudel and U. R. Dhungana, “Artificial intelligence for energy fraud detection: A review,” *Int. J. Appl. Power Eng.*, vol. 11, no. 2, pp. 109–119, 2022.
- [18] Y. Zhang, N. Qin, D. Huang, B. Wu, and Z. Liu, “High-accuracy and adaptive fault diagnosis of high-speed train bogie using dense-squeeze network,” *IEEE Trans. Veh. Technol.*, vol. 71, no. 3, pp. 2501–2510, Mar. 2022.

- [19] Y. Wang, J. Du, Z. Yan, Y. Song, and D. Hua, "Atmospheric visibility prediction by using the DBN deep learning model and principal component analysis," *Appl. Opt.*, vol. 61, no. 10, pp. 2657–2666, 2022.
- [20] X. Ye, S. Lan, S. Xiao, and Y. Yuan, "Single pole-to-ground fault location method for MMC-HVDC system using wavelet decomposition and DBN," *IEEE Trans. Electr. Electron. Eng.*, vol. 16, no. 2, pp. 238–247, Feb. 2021.
- [21] F. Masood, J. Masood, H. Zahir, K. Driss, N. Mehmood, and H. Farooq, "Novel approach to evaluate classification algorithms and feature selection filter algorithms using medical data," *J. Comput. Cognit. Eng.*, vol. 2, no. 1, pp. 57–67, 2023.
- [22] A. B. Slama, H. Sahli, A. Mouelhi, J. Marrakchi, M. Sayadi, and H. Trabelsi, "DBN-DNN: Discrimination and classification of VNG sequence using deep neural network framework in the EMD domain," *Comput. Methods Biomechanics Biomed. Eng., Imag. Visualizat.*, vol. 8, no. 6, pp. 681–690, Nov. 2020.
- [23] N. Calik, F. Güneş, S. Kozziel, A. Pietrenko-Dabrowska, M. A. Belen, and P. Mahouti, "Deep-learning-based precise characterization of microwave transistors using fully-automated regression surrogates," *Sci. Rep.*, vol. 13, no. 1, p. 1445, Jan. 2023.
- [24] Ö. İnik, "Classification of scenes in aerial images with deep learning models," *Türk Doğa ve Fen Dergisi*, vol. 12, no. 1, pp. 37–43, 2023.
- [25] U. Gupta and D. Gupta, "On regularization based twin support vector regression with Huber loss," *Neural Process. Lett.*, vol. 53, no. 1, pp. 459–515, Feb. 2021.
- [26] D. T. S. Kumar, "Video based traffic forecasting using convolution neural network model and transfer learning techniques," *J. Innov. Image Process.*, vol. 2, no. 3, pp. 128–134, Jun. 2020.
- [27] M. Anderson, A. M. Devlin, L. Pickering, M. McCann, and D. Wight, "It's not 9 to 5 recovery': The role of a recovery community in producing social bonds that support recovery," *Drugs, Educ., Prevention Policy*, vol. 28, no. 5, pp. 475–485, Sep. 2021.



HECHUAN ZHANG was born in Baoding, Hebei, in September 1989. He received the bachelor's degree in electrical engineering and automation from the Langfang Teachers College, in 2012, and the master's degree in electrical theory and new technology majoring in power marketing research from North China Electric Power University, in 2015.

Since 2015, he has been a Trainer with the Baoding Technical College of Electric Power (State Grid Jibei Electric Power Company Limited Skills Training Center). He has published six academic articles and participated in two research projects.

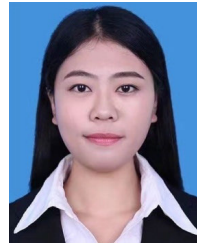
He has published six academic articles and participated in two research projects.



LEI WANG was born in Langfang, Hebei, in October 1982. He received the bachelor's degree in e-commerce majoring in power marketing research from the Hebei University of Science and Technology, in 2007.

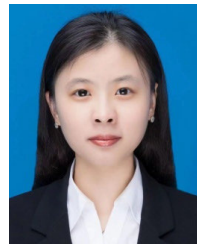
From March 2008 to July 2016, he was a Special Engineer with State Grid Langfang Power Supply Company. From July 2016 to February 2019, he was a Marketing Network Information Chief Security Officer of the Marketing Department, State Grid Jibei Electric Power Company Ltd. From February 2019 to August 2019, he was with the Institute of Measurement, Chinese Academy of Electric Power Sciences, assisted in the management of line loss in the same period in the substation area. He was with the Inspection Center of Langfang Power Supply Company, State Grid Jibei Electric Power Company Ltd., from August 2019 to October 2019. From October 2019 to March 2020, he was with the Marketing Department (Customer Service Center), State Grid Jibei Xianghe County Power Supply Company Ltd. Since March 2020, he has been an Intermediate Trainer with the Baoding Technical College of Electric Power (State Grid Jibei Electric Power Company Limited Skills Training Center). He has published three academic articles and participated in four research projects.

He has published three academic articles and participated in four research projects.



WENHUI LI was born in Zhangjiakou, Hebei, in June 1995. She received the bachelor's degree in electrical engineering and automation from Shijiazhuang Tiedao University, in 2018, and the master's degree in electrical engineering and automation majoring in power marketing research from Chongqing University, in June 2021.

Since 2021, she has been a Trainer with the Baoding Technical College of Electric Power (State Grid Jibei Electric Power Company Limited Skills Training Center) the Baoding Technical College of Electric Power (State Grid Jibei Electric Power Company Limited Skills Training Center).



YUJUN WANG was born in Xicheng, Beijing, in October 1989. She received the bachelor's degree in electrical engineering and automation from the Department of Electrical Engineering and Applied Electronic Technology, Tsinghua University, in 2012, and the master's degree in electrical engineering and automation majoring in power marketing research from the Department of Electrical Engineering and Applied Electronic Technology, Tsinghua University, in August 2015.

From December 2015 to December 2019, she was an Engineer with the State Grid Hebei Electric Power Research Institute. She was an Engineer with State Grid Jibei Marketing Service Center, from January 2021 to December 2021. Since November 2021, she has been an Engineer with State Grid Jibei Electric Power Company Ltd. She has published two academic articles and participated in two research projects.



GE ZHANG was born in Tianjin, in August 1971. He received the bachelor's degree in industrial electrical and automation from the Taiyuan Heavy Machinery College, in 1995. From 1995 to 2000, he was a Staff Member with Tianjin Beichen Power Supply Branch. From 2000 to 2009, he was the Deputy Director and the Director of the Power Supply Office of Tianjin Beichen Power Supply Branch. From 2009 to 2012, he was the Chief of the Electricity Department, State Grid Tianjin Chengdong Power Supply Company. From 2012 to 2021, he was the Deputy Director and the Secretary of the Marketing Department, State Grid Tianjin Chengdong Power Supply Company. Since 2021, he has been a Senior Expert of State Grid Tianjin Chengdong Power Supply Company. He has published four academic articles, participated in two scientific research projects, obtained six utility model patents, and won two provincial and ministerial level science and technology innovation awards. His research interests include power marketing, line loss anti theft management, and energy metering.

He has published four academic articles, participated in two scientific research projects, obtained six utility model patents, and won two provincial and ministerial level science and technology innovation awards. His research interests include power marketing, line loss anti theft management, and energy metering.

...