International Journal of Smartcare Home
Vol.1, No.1 (2021), pp.9-22
https://dx.doi.org/10.21742/26531941.1.1.02

Electricity Theft Detection using Fusion DenseNet-RF Model

Tai-hoon Kim

University of Tasmania, Churchill Ave, Hobart TAS 7005, Australia
taihoon.kim@utas.edu.au

Abstract

Aiming at the problems of high cost and low efficiency in power theft detection, this paper proposes a new type of DenseNet-RF model to detect power users theft. The DenseNet and Random Forest (RF) algorithms are fused, where DenseNet is used to automatically extract customer power usage characteristics, and RF is used to classify customer power abnormalities. By introducing the SMOTE algorithm, the imbalance problem of customer power consumption data was solved, the hyperparameters inside DenseNet were adjusted and the network model was pruned to obtain a better preprocessing model. When training the random forest classifier, call the preprocessing model and optimize the classifier parameters through grid search, and then obtain the final fusion algorithm model. The experimental results show that the DenseNet-RF fusion model effectively improves the accuracy of classification. Compared with the ensemble learning algorithm in single machine learning, traditional convolutional neural network, DenseNet, and algorithms of existing research results, the algorithm adopted by this model has better classification accuracy and stability, and the model has good generalization. ability.

Keywords: *Random forest, DenseNet, Electricity stealing detection, Network search algorithm*

1. Introduction

In recent years, with the rapid construction of smart grids, smart grids based on the AMI (Advanced Metering Infrastructure) system [1] can measure and monitor electricity consumption information in real-time or quasi-real-time, which helps reduce the risk of electricity theft. However, the two-way communication network, programmable smart meters and other information technologies introduced in AMI have brought new threats to the power system. The thief may use the new Internet of Things technology to attack the system vulnerabilities of the smart meter and then tamper with the measurement data of the local and remote interaction to achieve the purpose of illegal electricity use. In addition, illegal power use also includes physical methods such as line tapping, zero-breaking, undercurrent, and Undervoltage methods [2]. These deceptive power consumption behaviors have caused hundreds of millions of losses to the power industry in various countries. However, the current power detection methods are mostly based on manual on-site inspections, including manual inspection of unauthorized wiring, comparison of normal meter reading data and vicious meter reading data of the electricity collection system, and inspection of equipment or hardware problems. These methods are costly and time-consuming and labor-intensive. Therefore,

Article history:
Received (March 10, 2021), Review Result (April 7, 2021), Accepted (June 10, 2021)

eISSN: 2653-1941 IJSH

research on fast and accurate power theft detection technology is essential to improve the security and stability of the power grid.

The current research on electricity theft detection methods can be divided into two categories: state detection and artificial intelligence. The artificial intelligence method [3][4] mainly uses machine learning and deep learning technology to train the classifier on the sample data obtained by the smart meter. Then the illegal electricity consumption behavior is judged, and the operational complexity is moderate. However, many machine learning methods require manual feature selection for high-dimensional data, including maximum value, mean value, variance, standard deviation, etc. Manual feature extraction is tedious and time-consuming, and it is more difficult to extract features from two-dimensional data in smart meters. difficult. In contrast, deep learning technology has simple feature extraction operations and high discrimination accuracy. It has greater advantages and development prospects. More and more scholars have introduced it to the detection of electricity theft. Bhat et al. [5] compared the Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) algorithms in the identification of illegal electricity behavior. In contrast, the convolutional neural network is accurate. The higher the rate, it is the algorithm that is more suitable for the detection of electricity theft. In the United States, the loss caused by power theft is as high as US$4.5 billion per year [6]. Although the CNN model has been able to achieve a higher accuracy rate, further improving the accuracy of the detection of electricity theft will help to reduce the loss of national energy to a greater extent. As the network depth increases, the accuracy of the CNN model will reach saturation or even decline. However, the above algorithm does not consider the problems of gradient disappearance and gradient explosion in the deep network, so the classification accuracy of the model cannot be further improved [7].

In recent years, deep learning has made shocking breakthroughs in the fields of speech, image, natural language, and data mining, but the good results achieved cannot deny the traditional machine learning theory [8]. In image recognition, commonly used classifiers such as Logistic classifier and Softmax classifier can only solve general recognition and classification problems, and the accuracy of the recognition of complex and easily confusing objects is not high [9], and there are many limitations. The Random Forest (RF) algorithm has a high prediction accuracy, has a good tolerance for outliers and noise, and is not prone to overfitting [10]. Reference [11] draws on the advantages of traditional machine learning methods and proposes a random forest classification method combined with deep learning for image recognition of power equipment, which is better than the average recognition accuracy of conventional convolutional neural network classifiers and traditional random forest classifiers. It is 6.8% and 12. 6% higher. Based on the above background, this paper proposes a power user theft detection algorithm DenseNet-RF, which is a combination of dense convolutional neural network (DenseNet) and Random Forest (RF), to further improve the accuracy and generalization of the discrimination of power theft. ability. DenseNet is used to extract customer electricity consumption behavior characteristics from smart meter data, and RF is used to classify and distinguish the features extracted by DenseNet. The optimal parameters of the model are determined through multiple sets of experiments, and then the DenseNet-RF fusion model is applied to the detection of electricity theft by power users. The experimental results show that the classification accuracy of the DenseNet-RF fusion model in the training set is about 99.97%, and the classification accuracy on the test set is about 96.76%, which is compared with the integrated learning algorithm in a single machine learning. Traditional convolutional neural network, DenseNet model, and the algorithm performance of existing research results are better, which proves the effectiveness of the DenseNet-RF model.

In addition, the AUC (Area Under the Curve) value of the DenseNet-RF model reached 0.99, which shows that the fusion model has good generalization ability.

2. Design of DenseNet-RF fusion algorithm

2.1 DenseNet algorithm

DenseNet is a new network structure proposed by the literature [12]. The basic structure of the network mainly includes two-component modules: Dense Block and Transition Layer. Each layer in the Dense Block is connected to all subsequent layers in a densely connected manner. The Transition Layer is used for two adjacent Dense Blocks [13]. In DenseNet-121, the numbers of the 4 Dense Blocks are 6, 12, 24, and 16, respectively, and the growth rate k of the Feature Map is set. Suppose there are i feature layers in the Dense Block structure. X_i is recorded as the output of the i-th layer, H_i is recorded as the non-linear mapping of the i-th layer. A dense block structure with i layer, the number of dense connections is $(i \times (i+1))/2$,, then:

$$X_i = H([X_{i-1}, \cdots, X_1, X_0]) \tag{1}$$

$[X_{i-1}, \cdots, X_1, X_0]$ A means using the merged cascade layer to merge and cascade the outputs of all the previous feature layers according to the number of channels. The $H_i(\cdot)$ function uses Batch Normalization (BN), REU (Rectified Linear Unit), and the result after Convolution, (Conv). Using the merge cascade to fuse the feature information extracted by the convolution operation or the output layer information, and then increase the number of original features during training, which can improve the utilization of feature layers and is more conducive to improving the classification accuracy of the model. Concatenation is used to connect between layers.

2.2. Random forest (RF) algorithm

Random Forest (RF) is a very representative bagging integrated algorithm proposed by Breiman [14]. The algorithm uses decision trees as the basic classifier and builds multiple decision trees through random re-sampling technology [15] and node random splitting technology [16], and finally combines the prediction results of the decision tree for average or majority voting principles to determine the integrated classifier the result of. The random forest classifier used can effectively solve the problems of a single decision tree prone to overfitting, weak generalization ability, and low classification accuracy, and can reduce the generalization error of the learning system. In addition, the random forest has the advantages of parallel computing and processing high-dimensional features, which has great advantages when training classifiers.

2.3. Establishment of DenseNet-RF model

Although a single DenseNet algorithm can be used to effectively classify data, based on the complexity of smart grid data, it is necessary to further classify the extracted features to improve accuracy and generalization ability. Therefore, this paper combines DenseNet and RF With their respective advantages, the DenseNet-RF fusion model is proposed for the first time. The fusion model first uses DenseNet for automatic feature extraction, and then uses random forest for classification. The structure of the built DenseNet-RF model is shown in Figure 1. Feature extraction is the key factor for the success of the detection model. The structure adopted in

DenseBlock is BN→RELU→Conv(1×1)→BN→RELU→Conv(3×3), this continuous alternating feature extraction architecture. You can learn more advanced features step by step. The classification of electricity theft behavior is the ultimate goal of model detection. When training the RF classifier, the grid search algorithm can be used to optimize the parameters of the RF classifier, which helps to improve the accuracy of the classification.

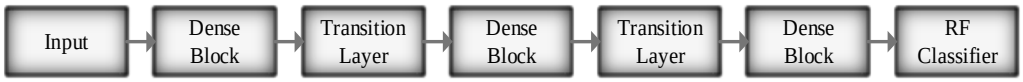


Figure 1. DenseNet-RF model structure

3. Smart meter data set and data preprocessing

3.1. Introduction to the data set

The data set used in this article is a smart meter electricity consumption data set collected by the Sustainable Energy Authority of Ireland (SEAI). The data set contains electricity consumption data from more than 6,000 households, non-residents, and businesses in Ireland and London. All users participating in the Customer Behavior Test (CBT) of smart meters have installed smart meters, and the installed meters record power consumption at a collection frequency of every 30 minutes. In the data set, each user has 534 d of electricity consumption data, and the daily data contains 48 components [17].

3.2. Data preprocessing

The electricity consumption data set of the original smart meter is about 150 million pieces of data. The original data is traversed many times, and the electricity consumption data of each customer is finally converted into two-dimensional data daily, and the two-dimensional data is defined as the actual user's consumption. Electric energy consumption value matrix [d， t], where the two-dimensional data row represents the date as $d\{d = 1, 2, \ldots, 534\}$, and the data column represents the time period $t\{t = 1, 2, \ldots, 48\}$. Then, the data set is cleaned according to the missing data to improve the reliability of the sample. After the cleaning is completed, the sample data set retains all the electricity consumption data in 534d of 6048 users, all of which belong to honest users. Figure 2 shows the daily electricity consumption of a certain user.

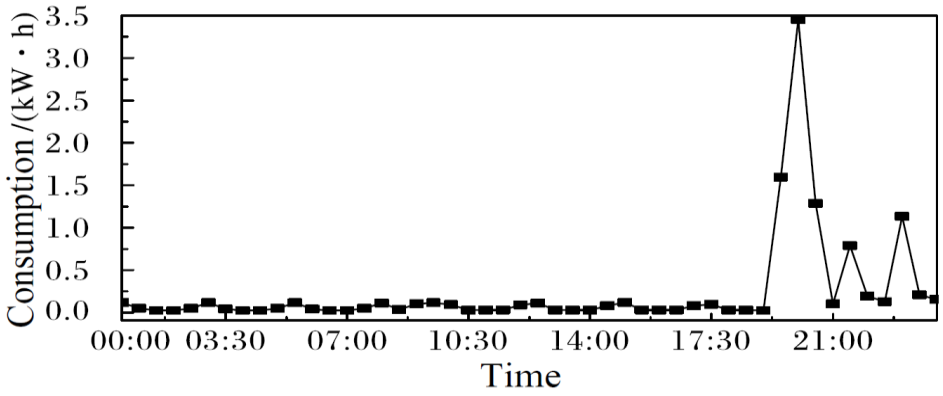


Figure 2. Daily electricity consumption of a user

To solve the above problems, the Synthetic Minority Over Sampling Technique (SMOTE) algorithm [18] is used to make the number of samples of electricity stealing users in the training data set equal to the number of samples of normal electricity users. The basic principles of the algorithm: 1. For each sample m in the minority class M, use the Euclidean distance as the standard to calculate the distance from all samples in the minority class set to obtain its k nearest neighbors. 2. Set one according to the sample imbalance ratio Sampling ratio to determine the sampling magnification s, for each minority sample m, randomly select s neighbors from its k neighbors, assuming that the selected neighbors are $m^{(1)}, m^{(2)}, \cdots m^{(s)}$. 3. For each randomly selected neighbor $m^{(i)}(i = 1,2, \cdots, s)$, construct a new sample according to the following formula: $m_{\text{new}} = m + \text{rand}\ (0,1) \times \left(m^{(i)} - m\right)$. In addition, to enable the network to quickly sign off and avoid numerical problems, the power consumption data is normalized, as shown in equation (2):

$$e_{t(0,1)} = \frac{e_t - \text{m}\ \ (e)}{\text{m}\ \ (e) - \text{m}\ \ (e)} \tag{2}$$

Among them: $\max(e)$ and in (e) respectively represent the maximum and minimum values in the data sample. Through the above-mentioned data processing method, the number of normal samples and power-stealing samples are finally equalized, and there are 12096 data sets in total. Data set label 0 represents normal electricity customers, and 1 represents electricity stealing customers. When dividing the training set and the test set, 10,000 samples are randomly selected for each training in the code and divided into the training set, and the remaining 2,096 samples are divided into the test set.

4. User stealing detection method based on DenseNet-RF

4.1. Problem analysis of user stealing detection

Based on the hardware status detection method, on the one hand, special equipment needs to be used to improve the detection accuracy, and the implementation of operation and maintenance costs are high. On the other hand, manual surprise inspections and random tests of on-site electrical equipment are required, which are generally effective and time-consuming, and labor-intensive. When detecting based on artificial intelligence methods, most of the existing methods are based on one-dimensional electricity consumption data, and these methods have poor power theft detection accuracy. The reason lies in the volatility and independence of users' daily electricity consumption. At the same time, electricity consumption is also affected by seasons, lifestyle, weather conditions, and many other uncontrollable factors. Therefore, it is difficult to identify the periodic or non-periodical customer power consumption from the one-dimensional power consumption data [6]. This article aims to design a new type of electricity theft detection method to solve the above problems. The large amount of data used and the long-period smart meter data set provide data guarantee for the use of deep learning methods to study the detection methods of electric power users' theft.

4.2. Principles of user stealing detection

With the application of the AMI system, power supply companies have accumulated a large amount of user historical electricity consumption data, which hides the behavioral characteristics of different users in various periods. Using the electricity theft detection model

can quickly and comprehensively analyze the historical electricity consumption data of a large number of users, to help power supply companies detect suspected electricity theft users.

In this paper, DenseNet is used to automatically extract customer power consumption characteristics from the smart meter data set, and then the RF algorithm is used to train the classifier according to the obtained characteristic parameters. Then use the grid search algorithm to optimize the parameters of the random forest classifier, to obtain the best random forest classifier, and judge whether the user has electricity theft behavior.

4.3. Training process of user stealing detection model based on DenseNet-RF

The specific steps of the user's electricity theft detection model training process are as follows: first, data preprocessing, data cleaning of the smart meter data set, and then use the above 6 types of electricity theft to randomly generate electricity theft samples, and use the SMOTE algorithm to make normal samples and theft The number of electrical samples is equal, and the data set has a total of 12096 positive and negative samples. Use formula (1) to normalize the data samples, and randomly divide 10,000 samples into the training set and 2096 samples into the test set. The power consumption value matrix for each user is (534×48). The second is to configure the network model, using the integrated learning algorithms in machine learning including RF, XGBoost (eXtreme Gradient Boosting), and Gradient Boosting Decision Tree (GBDT) to train their respective user stealing detection models, and then according to the literature [6] And the network configuration information training in [Table 1] [Table 2] includes WDNet [6], CNN and DenseNet user stealing detection models.

Table 1. CNN network parameter configuration

Layer	Output Size	CNN
Convolution 1	266×23	5×5 Conv, stride 2
Convolution2	133×12	3×3 Conv, stride 2
Convolution3	67×6	3×3 Conv, stride 2
Classification Layer	1×1	Full connected, SoftMax

Table 2. DenseNet network parameter configuration information table

Layers	Output Size	DenseNet-121 ($k = 12$)	DenseNet-121 ($k = 32$)	DenseNet-169 ($k = 32$)	DenseNet-201($k = 32$)
Convolution Pooling	91×74 45×37	3×3 Conv，stride 2 2×2 Conv，stride 2			
Dense Block (1)	45×37	$\begin{Bmatrix} 1 \times 1 \text{ Conv} \\ 3 \times 3 \text{ Conv} \end{Bmatrix} \times 6$			
Transition Layer (1)	45×37 22×18	1×1 Conv 2×2 average pool，stride 2			
Dense Block (2)	22×18	$\begin{Bmatrix} 1 \times 1 \text{ Conv} \\ 3 \times 3 \text{ Conv} \end{Bmatrix} \times 12$			

Transition Layer (2)	22 × 18 11 × 9	1×1 Conv2×2 average pool，stride 2		
Dense Block （3）	11×9	$\begin{Bmatrix}1\times1\ \text{Conv}\\3\times3\ \text{Conv}\end{Bmatrix}\times 24$	$\begin{Bmatrix}1\times1\ \text{Conv}\\3\times3\ \text{Conv}\end{Bmatrix}\times 32$	$\begin{Bmatrix}1\times1\ \text{Conv}\\3\times3\ \text{Conv}\end{Bmatrix}\times 48$
Transition Layer（3）	11×9 5×4	1×1 Conv 2×2 average pool，stride 2		
Dense Block （4）	5×4	$\begin{Bmatrix}1\times1\ \text{Conv}\\3\times3\ \text{Conv}\end{Bmatrix}\times 16$	$\begin{Bmatrix}1\times1\ \text{Conv}\\3\times3\ \text{Conv}\end{Bmatrix}\times 32$	
Classification Layer	1×1	7×7 global average pool Full connected，Softmax OR Random Forest		

The third is the training process. Through matrix transformation, the user power consumption value matrix meets the characteristic input of various algorithms. For example, when training a machine learning model, call the algorithm function in scikit-learn and use a grid search algorithm to determine the optimal parameters, and then obtain the corresponding model. When training CNN, the dimension of the matrix needs to be increased to (1×534×48), then the network is constructed according to Table 1, and finally, the fully connected layer is used for classification processing, and the classification results are normalized by the vector to the Softmax layer to output specific categories. According to the DenseNet-121, DenseNet-169, and DenseNet-201 (k=32) constructed according to the literature [9], to meet the input specifications of DenseNet, the matrix is transformed into (1×178×144), and the model is constructed first. The layer convolution operation is used to extract more features and a pooling operation is used to compress the layer size. Then construct the Dense Block structure and the transition layer according to the configuration information in Table 2. The Dense Block structure facilitates the feature reuse of information. The 1×1 convolutional layer in the Transition Layer layer is used for dimensionality reduction and 2×2 average pooling. The layer further compresses the size to reduce network parameters. Then after the pooling operation, it is input to the fully connected layer for classification processing, and the classification result is processed by the vector normalization Softmax and then the specific category is output. In addition, each function generates k feature maps, where k is the number of channels in the input layer. Change the hyperparameter k of DenseNet to compare the performance of the model. The fourth is to save the training results and compare model performance: input the test set into the trained model for prediction, and evaluate the model performance based on the classification results generated by the model and the true labels of the test set. The fifth is the fusion model: from the above training results, based on the optimal DenseNet, the model parameters are pruned and retained, and then the classification model is retrained on the extracted features through the random forest algorithm, and finally, the DenseNet-RF model is obtained. The sixth is to output the classification results to determine the users suspected of stealing electricity: According to the results of the classification, the corresponding users suspected of stealing electricity can be found. Figure 3 shows the training process of the user stealing detection model. The input of the model is a smart meter data set, and the output is a user suspected of stealing electricity.

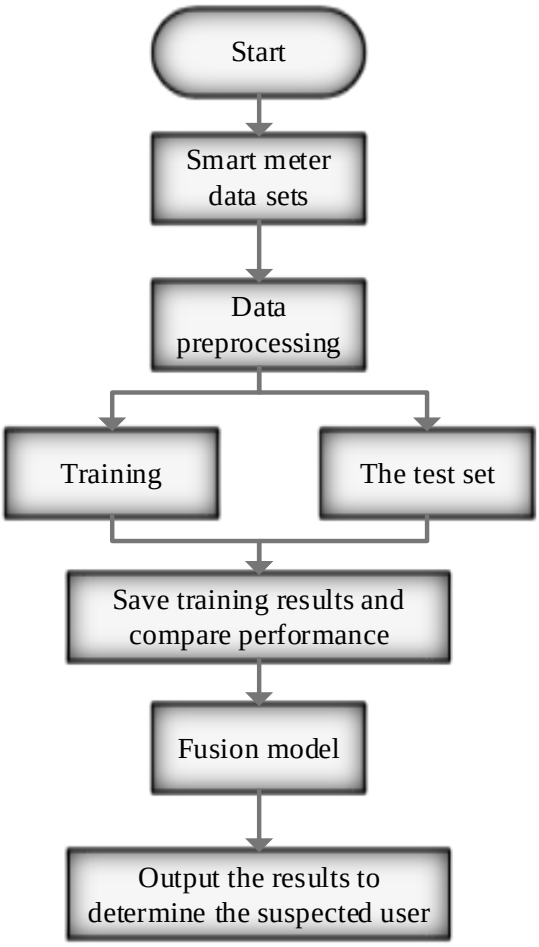


Figure 3. User theft detection model training process

5. Implementation of the detection method for power theft by power users

To verify the performance of the DenseNet-RF model, the data set is used to study the power theft detection of power users. The experimental platform is configured for Windows 10 Professional Edition, the CPU is AMD Ryzen 5 1400, the graphics card is RTX 2070, and the memory is 8 GB; programming and experiments are based on the Pytorch deep learning framework and the scikit-learn machine learning framework.

5.1. DenseNet and related grid model training

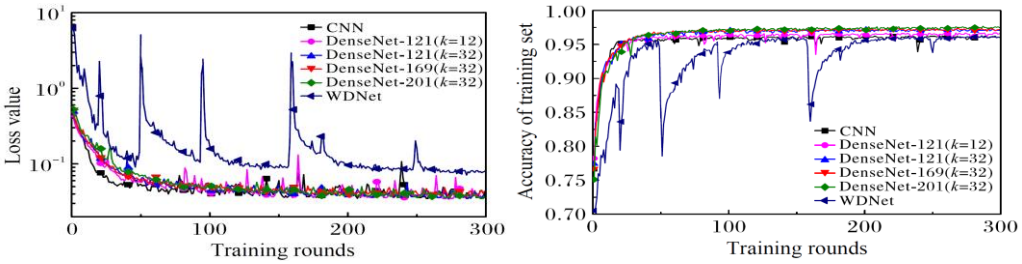
Before model training, set the number of convolution layers, convolution kernel size, number of channels, optimizer Adam, Batch size, etc. according to the network configuration information in [Tables 1] [Table 2]. To make full use of the graphics card memory and computing power to achieve the best time during the training process, different batch sizes are selected. Table 3 shows the indicators of the model when the number of training rounds is 300. It can be seen from [Table 3] that reducing the number of channels k in the network structure of the same depth in DenseNet can greatly reduce the total number of network parameters, and the size of the model can also be greatly reduced.

Table 3. Various indicators of the training model

Model	Total parameter	Batch size	Training time /min	Model size /MB	Tends to converge /epoch
CNN	532040	128	274	2.03	195
DenseNet-121 ($k = 12$)	1007620	128	424	3.84	140
DenseNet-121 ($k = 32$)	6839490	64	566	26.09	160
DenseNet-169 ($k = 32$)	12352706	48	656	47.12	165
DenseNet-201 ($k = 32$)	17940930	32	987	68.44	153
DenseNet-121 ($k = 12$) − RF	1006848	128	425	3.80	140
WDNet	1096674	144	413	4.18	210

5.2. Comparison of classification accuracy of multiple network models

To find a network model with higher classification accuracy, fewer parameters, and smaller size, RF, XGBoost, GBDT, CNN, WDNet, and DenseNet with different network configurations were applied to the test set of the smart meter data set, and the classification was accurate Rate comparison. Among them, the training results of RF, XGBoost, and GBDT can be read directly by calling the algorithm function of scikit-learn. CNN, WDNet, and DenseNet with different network configurations are obtained through multiple rounds of training. The training process is shown in [Figure 4], and the classification results of each network model on the test set are shown in [Table 4].



(A) Change curve of loss value (B) Accuracy rate change curve of the training set

(C) Test accuracy rate change curve

Figure 4. The curve of training process

Table 4. Classification accuracy of different network configuration models in the test set

Network model	Test set accuracy rate/%	Network model	Test set accuracy rate/%
			94.76
XGBoost	89.60	DenseNet-201 ($k = 32$)	94.84
GBDT	88.67	DenseNet-121 ($k = 12$)	95.18
CNN	91.44	WDNet [6]	92.73
DenseNet-121 ($k = 32$)	94.62	-	-

Based on the above results, the following conclusions can be drawn:

(1) It can be seen from Figure 4 that the CNN, WDNet, and DenseNet models have good accuracy on the training set, and the model prediction accuracy of DenseNet in the electricity data test set is significantly better than that of CNN and WDNet, while WDNet the accuracy rate in the test set is slightly higher than that of CNN, but the loss value loss of the WDNet model oscillates during the training process, and the model does not converge quickly.

(2) From Table 4, it can be concluded that the network models constructed by convolutional neural networks such as CNN, WDNet, and DenseNet have significantly higher classification accuracy than the integrated algorithms of RF, XGBoost, and GBDT in the electricity data test set.

(3) Combining Table 3 and Figure 4, it can be seen that although DenseNet-169 and DenseNet-201 gradually increase in the number of layers and network complexity of Dense Block during the training process, the classification accuracy of the training set is only slightly improved. And the classification accuracy rate in the test set is roughly the same as DenseNet-121, but it greatly increases the total network parameters and training time, and the network model is too complex and redundant.

(4) Synthesizing Tables 3~4 and Figure 4, it can be concluded that although the DenseNet-121 (k=12) network model performs slightly inferior to the number of channels k (k=32) in the training set, and the number of network layers is deeper Model, but the classification accuracy rate in the test set did not decrease but slightly improved. The classification accuracy rate reached the highest 95.18% at the peak, and the training time was shorter, and the number of rounds required to converge was less.

5.3. Comparison of classification accuracy of DenseNet-RF fusion model

To obtain a better model, the performance of various network models in Table 4 is combined, and the above-mentioned optimal DenseNet (DenseNet-121 (k=12)) model is retained based on the last classification layer. All parameters are saved as a new DenseNet network model, which extracts the power consumption characteristics of the final power user, and uses the new model at this time as the pre-training model for constructing the RF classifier. Use the scikit-learn machine learning framework to train the RF classifier on the training set, and use the grid search algorithm to optimize the RF classifier parameters, and finally get the DenseNet-RF fusion model. Finally, input the trained fusion model into the test set for prediction. Using the above methods, the DenseNet-GBDT and DenseNet-XGBoost fusion models are obtained by comparative training and compared with DenseNet-RF, DenseNet, and other typical models of existing research results. The experimental results are shown in [Table 5].

Table 5. Accuracy comparison between the fusion model and different models Unit: %

Network model	Training set	Test set
DenseNet-RF	99.97	96.76
DenseNet-GBDT	99.24	96.23
DenseNet-XGBoost	99.08	95.88
DenseNet-121 ($k = 12$)	96.61	95.18
WDNet [6]	96.04	92.73
ELM [3]	78.66	70.69
SVM [4]	83.72	71.45

From [Table 5], the following conclusions can be drawn: the classification accuracy of the DenseNet-RF fusion model in the training set is about 99.97%, and the classification accuracy on the test set is about 96.76%. Compared with DenseNet-GBDT and DenseNet- XGBoost, the single DenseNet model and existing research results models have better performance, which proves the effectiveness of the DenseNet-RF model.

5.4. Performance evaluation of DenseNet-RF fusion model on new samples

The user stealing electricity detection model can be considered as a discrete two-classification task. The purpose is to classify each electricity customer into one of two categories: abnormal or normal. In this paper, the positive category 1 is regarded as the stealing user, and the negative category 0 is regarded as the normal honest user. The purpose of machine learning is to apply the learned model to new samples. The ability of the model to perform well on previously unpredicted inputs is called generalization ability. Generally, the generalization ability of the model is evaluated by measuring the performance of the model on the test set, and the verification result is compared with the ROC (Receiver Operating Characteristic) and Area Under the ROC curve (AUC) for presentation.

Table 6. AUC comparison of various models in the test set

Network model	AUC
DenseNet-RF	0. 99
DenseNet-GBDT	0. 98
DenseNet-XGBoost	0. 96
DenseNet	0. 95
WDNet [6]	0. 93
SVM [4]	0. 77
CNN	0. 93
RF	0. 91

GBDT	0. 77
XGBoost	0. 63
ELM [3]	0. 75

The following conclusions can be drawn from [Table 6]: The DenseNet-RF model performs very well on the test set, with an AUC of 0.99, and has good generalization ability, that is, the model has good judgment accuracy on whether new users are stealing electricity. And applicability.

6. Conclusion

This paper proposes a new type of DenseNet-RF model to detect electricity theft by power users. The DenseNet and Random Forest (RF) algorithms are fused, where DenseNet is used to automatically extract customer power usage characteristics, and RF is used to classify customer power abnormalities. The SMOTE algorithm was introduced to solve the problem of imbalance in customer power consumption data. The hyperparameters inside DenseNet were adjusted and the network model was pruned to obtain a better preprocessing model. The preprocessing model was called and passed through the grid when training the random forest classifier. Search and optimize the classifier parameters, and then obtain the final fusion algorithm model. In addition, to compare the detection effects of different network models on users' electricity theft behavior, several experiments were carried out, and the model parameters were determined from them; finally, the DenseNet-RF fusion model was applied to power users' electricity theft detection. The experimental results show that the DenseNet-RF fusion model effectively improves the accuracy of classification. Compared with the ensemble learning algorithm in single machine learning, traditional convolutional neural network, DenseNet, and algorithms of existing research results, the algorithm adopted by this model has better classification accuracy and stability, and the model has good generalization. ability. This model has good performance in detecting comprehensive electricity theft types, and can accurately determine whether users have electricity theft behaviors. By comparing the existing methods, the method proposed in this paper has very good practical application value for the detection of power theft in the power grid.

References

[1] S. Mclaughlin, D. Podkuiko, and P. McDaniel, “Energy theft in the advanced metering infrastructure,” Proceedings of the 2009 International Workshop on Critical Information Infrastructures Security. Cham: Springer, pp.176-187, **(2009)**

[2] H. K. Lee and K. Nam, “On-line overshoot suppression method for EV propulsion motor considering cross-coupled inductance,” IEEE Transactions on Industrial Electronics, pp.9255-9265, **(2018)**

[3] A. H. Nizar, Z. Y. Dong, and Y. Wang, “Power utility non-technical loss analysis with extreme learning machine method,” IEEE Transactions on Power Systems, vol.23, no.3, pp.946-955, **(2008)**

[4] J. Nagi, K. S. Yap., and S. K. Tong, “Nontechnical loss detection for metered customers in power utility using support vector machines,” IEEE Transactions on Power Delivery, vol.25, no.2, pp.1162-1171, **(2009)**

[5] R. R. Bhat, R. D. Trevizan, and R. Sengupta, “Identifying non-technical power loss via spatial and temporal deep learning,” Proceedings of the 2016 IEEE International Conference on Machine Learning and Applications. Piscataway: IEEE, pp.272-279, **(2016)**

[6] P. McDaniel and S. McLaughlin, “Security and privacy challenges in the smart grid,” IEEE Security and Privacy, vol.7, no.3, pp.75-77, **(2009)**
[7] G. Pleiss, D. Chen, and G. Huang, “Memory-efficient implementation of DenseNets,” Computer Vision and Pattern Recognition, vol.2, no.1, pp.47-49, **(2017)**
[8] D. Anguita, A. Ghio, and N. Greco, “Model selection for support vector machines: Advantages and disadvantages of the machine learning theory,” Neural Networks (IJCNN), The 2010 International Joint Conference on. IEEE, **(2010)**
[9] Z. Khandezamin, M. Naderan, and M. J. Rashti, “Detection and classification of breast cancer using logistic regression feature selection and GMDH classifier,” Journal of Biomedical Informatics, **(2020)**
[10] M. Balachandran, T. H. Shin, and M. O. Kim, “AIPpred: Sequence-based prediction of anti-inflammatory peptides using random forest,” Frontiers in Pharmacology, vol.9, pp.276, **(2018)**
[11] M. Oprescu, V. Syrgkanis, and Z. S. Wu, “Orthogonal random forest for heterogeneous treatment effect estimation,” Vasilis Syrgkanis, **(2018)**
[12] F. Iandola, M. Moskewicz, and S. Karayev, “DenseNet: Implementing efficient ConvNet descriptor pyramids,” Eprint Arxiv, **(2014)**
[13] B. Fielding and Z. Li, “Evolving deep DenseBlock architecture ensembles for image classification,” Electronics, vol.9, no.11, pp.1880, **(2020)**
[14] L. Breiman, “Random forests,” Machine Learning, vol.45, no.1, pp.5-32, **(2001)**
[15] L. Breiman, “Bagging predictors,” Machine Learning, vol.24, no.2, pp.123-140, **(1996)**
[16] D. Kushary, “Bootstrap methods and their application,” Technometrics, vol.42, no.2, pp.216-217, **(2000)**
[17] P. Jokar, N. Arianpoo, and V. C. M. Leung, “Electricity theft detection in AMI using customers’ consumption patterns,” IEEE Transactions on Smart Grid, vol.7, no.1, pp.216-226, **(2015)**
[18] N. V. Chawla, K. W. Bowyer, and L. O. Hall, “SMOTE: Synthetic minority over-sampling technique,” Journal of Artificial Intelligence Research, vol.16, no.1, pp.321-357, **(2002)**

This page is empty by intention.