Electricity-theft detection in smart grids based on deep learning

Noor Mahmoud Ibrahim, Sufyan T. Faraj Al-Janabi, Belal Al-Khateeb
College of Computer Science and Information Technology, University of Anbar, Ramadi, Iraq

Article Info

Article history:

Received Feb 16, 2021
Revised May 2, 2021
Accepted Jun 9, 2021

Keywords:

Blue monkey algorithm
electricity consumption
Convolutional neural network
Deep learning
Smart grid

ABSTRACT

Electricity theft is a major concern for utilities. The smart grid (SG) infrastructure generates a massive amount of data, including the power consumption of individual users. Utilizing this data, machine learning, and deep learning techniques can accurately identify electricity theft users. A convolutional neural network (CNN) model for automatic electricity theft detection is presented. This work considers experimentation to find the best configuration of the sequential model (SM) for classifying and identifying electricity theft. The best performance has been obtained in two layers with the first layer consists of 128 nodes and the second layer is 64 nodes. The accuracy reached up to 0.92. This enables the design of high-performance electricity signal classifiers that can be used in several applications. Designing electricity signals classifiers has been achieved using a CNN and the data extracted from the electricity consumption dataset using an SM. In addition, the blue monkey (BM) algorithm is used to reduce the features in the dataset. In this respect, the focusing of this work is to reduce the features in the dataset to obtain high-performance electricity signals classifier models.

Corresponding Author:

Noor Mahmoud Ibrahim
Department of Computer Science
College of Computer Science and Information Technology
University of Anbar, Ramadi, Al-Anbar, Iraq
Email: nnmm78444@uoanbar.edu.iq

1. INTRODUCTION

Smart grid (SG) is the ever-growing dispersion of renewable and divided sources of power, which is intended to attain flexibility, self-healing, effectiveness, and sustainability. The idea of SG is being recognized over the application of pretend infrastructure covering the inheritance power grid [1]. The cyberinfrastructure allows the group and study of data from lots of different dispersed endpoints, for example, units of determination of phasor, smart meters, and breakers of the circuit.

Usually, these grids contain some improvements that will develop the dependability, effectiveness, and delivery of continuous sources of energy to households and industries. In addition, SG contains different resources of renewable energy such as (power of wind, solar, and others), distributed storage (DS), and distributed generation (DG) [2]-[6]. The components of smart grid are supervisory control and data acquisition (SCADA), phasor measurement unit (PMU), flexible AC transmission system (FACTS), advanced conductors devices, electric power generators, electric power substations, transmission and distribution lines, controllers, smart meters, collector nodes, and distribution and transmission control centers [7]. Consequently, grids of smart metering operate with smart sensors permitting companies to run and regulate the SG, supplied with the technology of communication and information [8].

Electrical energy has become essential in the life of a human. Losses of electrical energy regularly happen for the duration of production, distribution, and transition of electrical energy. The losses of electrical energy can be generally classified into non-technical losses (NTLs) and technical losses (TLs) [9]. One of the significant Non-technical losses is electricity theft. There is a large group of investigations on detecting electricity theft. Traditional ways of detection of electricity-theft contain physically examination problematical meter set up or disconfirmation, associating the irregular meter readings with the regular ones and observing a line of the transition of the by-passed power. These ways are ineffective, tremendously time-consuming, and costly. The presence of SGs brings chances in resolving of electricity-theft. SGs are comprised of conventional networks of power, grids of communications linking smart devices (for example, smart sensors and meters) in networks, and calculating services to sense and regulate networks [10]. Information and energy move in smart networks attach companies of service and employers. In this way, smart sensors or meters can assemble various data including status information of networks, using electrical energy, information of financing, and cost of electrical energy [11]. The electricity network is considered as an SG that can smartly mix the actions of whole employers linked to its producers, customers and those that do both to proficiently provide maintainable, financial, and safe sources of electricity [12].

Lately, matters of privacy and security have been the issues of comprehensive research since the national economy, public security, and safety depend significantly on the grids of energy. Unfortunately, fears of privacy are not continuously completely understood in SG metering, and thus more is required to deal with threats of electricity theft. Some good survey papers have discussed the matters of privacy and security in the field of SGs. They listed the overall cybersecurity challenges including trust models, connectivity, management of security, the privacy of consumers, software vulnerabilities, and human factors. Possible solutions to these challenges were also proposed [13], [14]. Deng and Shukla surveyed the vulnerabilities and countermeasures, especially for the transmission subsystem within SG [15]. They focused on the point of weaknesses of technology of phasor measurement units (PMUs) and wide area measurement system (WAMS). Wang and Lu examined challenges of security in the grid of SG, containing home area networks (HANs), advanced metering infrastructures (AMIs), subsystems of distribution and transmission [16]. They showed the necessities of security and estimated network fears with matter studies. Komninos *et al*, presented SG and smart home safety study [17]. Those authors generally assumed the communication amid the environments of SG and smart home are categorized their hazards of safety. Mohassel *et al,* explained a study on (AMI) advanced metering infrastructure [18]. They studied the main ideas of AMI. They showed the physical and cyber safety challenges containing privacy briefly.

Some literature studied the methods of ETD, which uses consuming data of smart meters to discover deceitful consumers. Observing of load profiles of consumers for marks of electricity-theft in conventional power schemes has attracted the concerns of academics to this point. Angelos *et al,* utilized five characteristics containing maximum consumption, mean consumption, inspection remarks summation, standard deviance, and the mean consumption of the neighborhood to produce a usual form of consumption of power for every consumer. K-means-based fuzzy clustering was achieved for the collection consumers with the same profiles. Customers with ample spaces to the cluster centers were assumed potential cheats. Gathering the consumers and depending on long-period measurements limited the accurateness of this ETD system and produced a long detection delay. Possessing more detailed metering information may deliver a much better ETD with a much shorter delay [19].

Other earlier work has been directed towards deep learning and convolutional neural networks (CNNs). Abdel-Hamid *et al*, explained the CNN innovation in the variability of domains associated with forming appreciation from image treating to voice recognition. The most advantageous feature of CNNs is decreasing the factor number in artificial neural networks (ANNs). This attainment has encouraged both designers and academics to approximate bigger models to resolve difficult jobs, which was not probable with classic ANNs [20]. Indeed, Mallat extended earlier presented tools to produce a mathematical framework to analyze the properties of general CNN architectures. At a significant level, the extension was attained via substituting the requirement of invariants and contractions to translations via contractions along with adaptive collections of local symmetries [21].

Furthermore, an emphasis is given to review some research papers that have been conducted on using CNNs for ETD. In this respect, Krizhevsky *et al*, explored the use of CNNs for the task of theft detection. Motivated via the numerical model method, the periodicity of consecutive data is of considerable significance for the classifier. Therefore, an adequate explanation of the periodicity can be beneficial to develop the accurateness of the detection of electricity theft. Concretely, they suggested adjusting the multi-scale DenseNet, which can automatically capture the short-term and extensive-term periodic characteristics of the consecutive data [22].

Bhat *et al*, investigated three deep learning methods for the detection of electricity-theft, specifically, CNNs, (LSTM) recurrent neural networks (RNNs), and loaded autoencoders. Nevertheless, the working of the detectors was examined by utilizing synthetic data, which did not permit a dependable valuation of the performance of the detector associated with shallow architectures [23]. Finally, Zheng *et al*, proposed a wide and deep CNN model ETD in SGs. They noticed that most of the current approaches have poor accurateness of detection of electricity-theft as they are based on one dimensional (1-D) data of electricity consumption and failed to arrest the electricity consuming periodicity [24].

Therefore, in this paper, the focus is on proposing an efficient technique of electricity-theft detection (ETD) to deal with all worries mentioned above. In specific, we initially suggest a convolutional neural networks (CNN) with a recently proposed nature-inspired metaheuristic optimization algorithm called the blue monkey (BM) algorithm model [25] to recognize the thieves of electricity. The CNN part consists of several convolutional layers, a pooling layer, and a completely connected layer. Principally, the CNN component can capture the periodicity of electricity consumption data. This style mixes the profits of the CNN component and BM algorithm to facilitate efficient ETD. To the best of our knowledge, it is the first research to suggest such a deep algorithm model (mixing CNN with the BM algorithm) and carry it out to study electricity-theft in smart networks. In addition, we have done widespread trials on a huge accurate electricity-consuming dataset.

2. THE PROPOSED ELECTRICITY THEFT DETECTION SYSTEM

This section focuses on the design aspects of the proposed electricity theft detection system. Using statistical examination of the consumption of electrical energy data of both thieves of energy and usual consumers, one can find that the electricity consumption data of energy thieves are typically less non-frequent or frequent, associated with that of usual consumers. This monitoring can facilitate classifying the irregular using of electricity and the periodicity of the electricity consumption.

Nevertheless, it is challenging to examine the periodicity of the electricity-consuming data because of many reasons as: 1) it is problematic to study the periodicity of the electricity-consuming data because it is 1-D time series data with enormous size, 2) The electricity consumption data is frequently incorrect and loud, 3) Several traditional methods of data investigation, e.g., ANN and support vector machine (SVM) cannot be straight carried out to the consumption of electricity data because of the calculation difficulty and the restricted simplification ability. Thus, to face the above challenges, the CNN approach has been adopted in this work.

A realistic electricity consumption dataset released by State Grid Corporation of China is used to train the models. This work is intended to identify electricity theft from the power consumption pattern of users, utilizing CNN-based deep learning and BM techniques. This classifier model is trained to utilize a dataset consisting of daily power consumption data of both normal and fraudulent users in a supervised manner. First, the data is prepared by a data-preprocessing algorithm to train the model. The preprocessing step also involves synthetic data generation for better performance. At the next stage, the proposed model is hyper-tuned and finally, the optimized model is evaluated via the test data. The overall system is depicted in Figure 1.

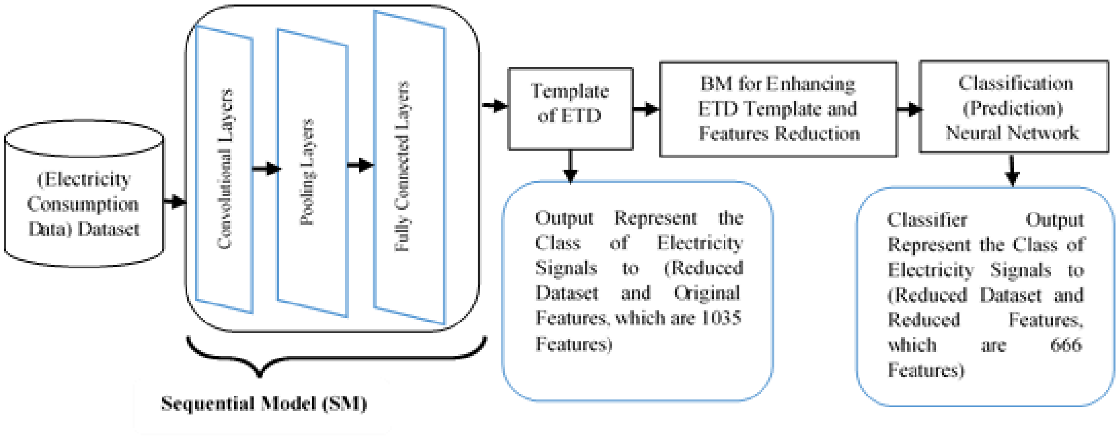
(Electricity Consumption Data) Dataset
Convolutional Layers
Pooling Layers
Fully Connected Layers
Sequential Model (SM)
Template of ETD
BM for Enhancing ETD Template and Features Reduction
Classification (Prediction) Neural Network
Output Represent the Class of Electricity Signals to (Reduced Dataset and Original Features, which are 1035 Features)
Classifier Output Represent the Class of Electricity Signals to (Reduced Dataset and Reduced Features, which are 666 Features)

Figure 1. Architecture of the proposed model (CNN & BM)

2.1. Electricity consumption data

The research is performed on a series of real consumer electricity usage data, made accessible by the State Grid Corporation of China (SGCC). This dataset consists of 42,372 rows and 1,035 columns. The first column includes the costumers' ID, and the second column includes a pointer of prediction called "Flag" while the days' columns start from the third column up to the column (1,035). The Metadata types in the dataset are a set of characters, numbers, and missing or erroneous values called non-numeric (NaN). The numbers and missing or erroneous values represent the amount of electricity consumption (electricity signals) for each consumer over two years.

In addition, the metadata in the flag column is (zero and one) and it is referring to the type of consumers (normal or thief), where the numbers of zeros in the "Flag" column represents the normal consumer of electricity and the total number of them is (38,757). While the number of one in the "Flag" column represents the thieves and the total number of them is (3,615). Finally, this means that the number (42,372) represents electricity consumers' data on electricity usage within 1,035 days (from Jan. 1, 2014, to Oct. 31, 2016).

2.2. Modifying the dataset

The given dataset of electricity consumption is passed in various stages of modifying to reduce it to be used in building operations of ETD templates using various algorithms. These stages are shown as: 1) Generating new dataset by replacing all null and Nan values in an original dataset with zero to get rid of null or non-numeric values (NaN), because the neural network accepts numbers only, and these values are not defined, so we have converted them into zeros until the neural network understands them, 2) Splitting new dataset into two parts, one part used for training (80%) and the other part used for testing (20%), 3) Reducing new dataset by dropping location and flag columns from a new dataset. The reason is to reduce the complexity and the time as those two attributes will not be used in the proposed system.

3. BUILDING THE ELECTRICITY-THEFT DETECTION MODEL

Our system used for ETD goes through several steps. The first step is our dataset passed in several modified operations to reduce it as discussed above. Then, the sequential model (SM) is built, as will be described in the next subsection. The third step is to build the prediction model (the ETD model), and this can be done by two operations. The first operation is done by using the developed SM, while the second operation is done using the BM algorithm. The input is SM with the reduced dataset and the output is the model of ETD. This algorithm consists of a set of fully connected layers, convolution layers, and softmax layers to train and test the dataset (electricity consumption data).

3.1. The sequential model

The SM is appropriate for a plain stack of layers where each layer has exactly one input tensor and one output tensor. The input to this SM algorithm is the dataset and the output is the reduced dataset. The first step in this algorithm is defining the input shape to be compatible with the SM. After that, there are two cases to use SM: The first case is predicting electricity signals using original fully connected layers of the SM. This can be done by sending the electrical signals to the SM and SM will return the class of these electrical signals. The second case is using an array and will use to build, and train a given dataset.

In addition, there are three types of 2D (3*3) convolution layers. The three convolution layers (3*3) are linked with fully connected layers (activation, dropout), then linking the other fully connected layers (dense, activation, dropout) with each other. Finally, the resulting metrics from the above operations are used to generate the reduced dataset. This approach uses three fully connected layers (dense) which are (layer1, layer2, and output), where the value of (layer 1) is 128 nodes and the value of (layer 2) is 64 nodes, while the (output) value represents the customer's type to be either thief or normal customer depending on the dataset. In addition, applying four fully connected layers (dropout), where the values of those layers have been selected after trying many possible values. It was found that the best values for them are 0.25 for the first two layers and 0.5 for the last two layers.

The input is SM with the reduced dataset and the output is the model of ETD with its accuracy and loss. Where the dataset is divided into two parts, the first part is used for training, which is 80% of the dataset and the second part is used for testing, which is 20% of the dataset. This algorithm is used to test the best configuration of the neural network in terms of several layers and parameters, beginning with two layers and ending with four layers. The maximum dimension of the layer has been 128 nodes and the minimum one has been 16 nodes. The best architecture has been obtained with two layers, where the first layer contains 128 nodes, and the second layer contains 64 nodes.

3.2. The blue monkey algorithm

The BM algorithm has been represented as a function to enhance the ETD template and return the solution of the best location, as shown in Figure 2. The input to this function is the ETD template. The BM algorithmic program mimics the behavior of the BM. BM is a set of solutions for parents and children each one of parents and children has random values. The input to this algorithm is a set of solutions each one represents the template of reducing the dataset and the ETD. The number of solutions used is 10 solutions, each solution has a length of 1035 values generated randomly using zeros and ones.

Now there is a template coming from BM, which has a length of 1035 values (0 or 1). This template will be used in two steps: The first step is to reduce the dataset according to modify function, where the input to this function is the template from BM and the original dataset (in the case of the building model). The output is a new dataset, which is less than the original dataset. Then building a model that has an input shape equal to the size of a new dataset. The second usage is when there is a new electricity signal to classify; it should reduce the values of the electrical signal according to the same template. Hence, the electrical signal can be classified using this model. The important goal of building the BM template is to reduce the number of features in the given dataset.

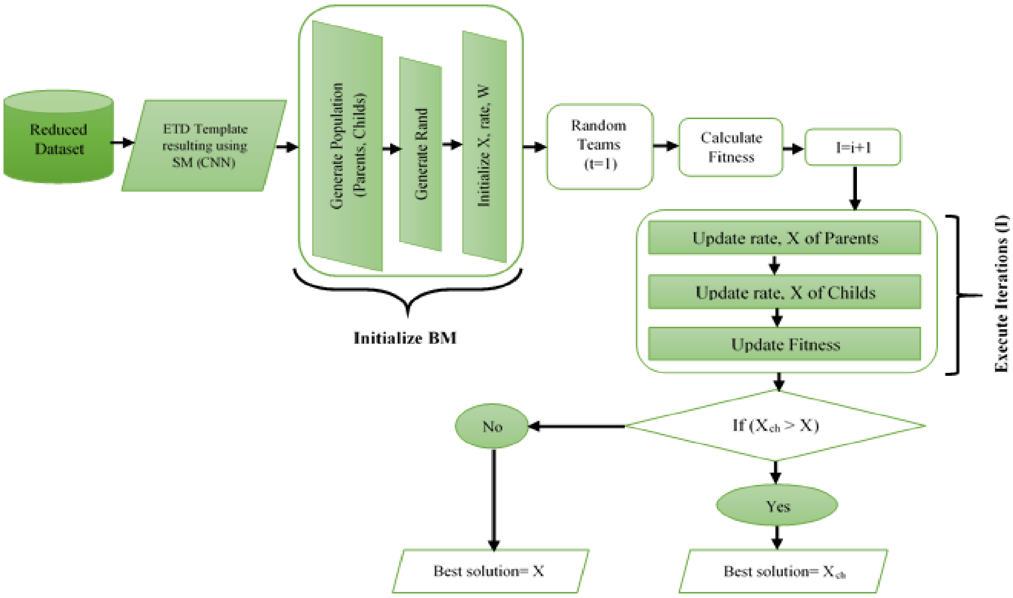

Figure 2. The BM algorithm

4. RESULTS AND DISCUSSION

Electricity signals classification is one of the core problems in the world with a large variety of practical applications. In this section, the proposed system is tested to obtain and discuss the results to indicate the effectiveness of this system. Various experiments have been done using the electricity consumption dataset. These experiments include testing the electricity signals classifier configuration, applying the BM model with the best configuration (of two selected layers), measuring accuracy and loss using the deep CNN and BM model, and finally comparing results of loss and accuracy resulting from CNN and BM model with those obtained by the CNN model alone.

4.1. Configuration of classifier part experiments

The configuration of fully connected layers in terms of many layers and nodes was tested beginning with two layers and ending with four layers. The maximum dimension of each layer is 128 nodes and the minimum one is 16 nodes. In Table 1, each row in these tables represents the complete model configurations and results obtained from this model. The columns of this table represent the model number (the sequence of the model in the experiment), the number of fully connected layers in the model, the number of nodes in each layer (16 -128 nodes), the best accuracy of the architecture, the worst accuracy value, the average accuracy value, the average training loss (This value equals to the difference between a true label and predicted label of the electrical signals inside training, which should be minimized as much as possible), and the time consumed for training, respectively. The selected architecture is that has two layers as it has the best compromise of average accuracy and consumed time.

4.2. Two layers experiments

This subsection explains the results for two layers in detail. In Table 2, the rows represent ten training rounds of the model. The columns comprise the number of nodes in each layer (128 and 64 nodes for the first and the second layer, respectively), the number of layers (2 layers), the sequence of each training round, and the loss, accuracy, and time of each model, respectively. The table also contains the values of the average of loss, accuracy, and time for the ten training rounds, respectively. Finally, it is possible to note that the values of accuracy and time are convergent, and this indicates the presence of stability in the network.

Table 1: Settings and results of network consists of two-four layers

No.	No. of Layers	No. of Nodes	Best Accuracy	Worst Accuracy	Average Accuracy	Average Loss	Average Time (Second)
1	2	128-128	0.918112094	0.916578174	0.91716815	0.266775871	13.1
2	2	64-64	0.91764009	0.914926231	0.915740407	0.283764297	10
3	2	32-32	0.915634215	0.915162265	0.91520946	0.285755402	10
4	2	16-16	0.915162265	0.915044248	0.915150464	0.295629337	10
5	2	128-64	0.918230116	0.916460156	0.917297935	0.26462948	10
6	2	128-32	0.91764009	0.915988207	0.916896755	0.276600096	10
7	2	128-16	0.916578174	0.915044248	0.915598828	0.277276114	10
8	2	64-32	0.91716814	0.914808273	0.915905619	0.281315494	10.1
9	2	64-16	0.91716814	0.915162265	0.916176987	0.270790696	10
10	2	32-16	0.915162265	0.91492623	0.915126864	0.296932735	10
11	3	128-128-128	0.918938041	0.916106224	0.91719175	0.255808522	10
12	3	64-64-64	0.916578174	0.915162265	0.915882021	0.272322157	10.1
13	3	32-32-32	0.915162265	0.915162265	0.915162265	0.295278683	10
14	3	16-16-16	0.915162265	0.915162265	0.915162265	0.291884622	10
15	3	128-64-32	0.917404115	0.916578174	0.916849566	0.265754673	10.1
16	3	64-32-16	0.915516198	0.915162265	0.915233052	0.283388585	10
17	4	128-128-128-128	0.916932166	0.915516198	0.916318583	0.267135677	10
18	4	64-64-64-64	0.916578174	0.914218307	0.915339243	0.287885103	10.1
19	4	32-32-32-32	0.915162265	0.915162265	0.915162265	0.293846384	10
20	4	16-16-16-16	0.915162265	0.915162265	0.915162265	0.290049273	10
21	4	128-64-32-16	0.91716814	0.915516198	0.916294992	0.271532404	10

Table 2. Two layers model (128-64 nodes).

No. of Nodes	No. of Layers	Training No.	Loss	Accuracy	Time	Loss Average	Accuracy Average	Time Average (Second)
128-64	2	Train 1	0.265759706	0.916460156	10	0.2646295	0.9172979	10
		Train 2	0.254469693	0.91716814	10			
		Train 3	0.260951936	0.916578174	10			
		Train 4	0.253114253	0.917286158	10			
		Train 5	0.260221213	0.918112099	10			
		Train 6	0.261828154	0.91716814	10			
		Train 7	0.267765313	0.917050123	10			
		Train 8	0.305615753	0.916932166	10			
		Train 9	0.258682787	0.917994082	10			
		Train 10	0.257885993	0.918230116	10			

4.3. Applying BM with best configuration of two layers in the CNN model

This subsection describes the results of accuracy and loss using CNN and BM model, after describing the number of iterations, which are ten iterations, and the number of solutions, which are ten solutions. This model consists of two layers, which are 128 and 64 nodes, as shown in Table 3. The loss is equal to 0.265327107, while the accuracy is equal to 0.916932165, and the time equals 10 seconds. The aim of combining the BM algorithm with CNN is to reduce features that will help in the reduction of time and complexity for the execution process.

Table 3. Results of accuracy and loss for two layers model using CNN & BM

No. of Nodes	No. of Layers	No. of Iteration	No. of Solution	Accuracy	Loss	Time (Second)
128-64	2	10	10	0.916932166	0.265327107	10

4.4. Comparing results of loss and accuracy

This subsection presents comparing the results of loss and accuracy that are resulting from the CNN and BM model with results of loss and accuracy resulting from the CNN model without BM, as shown in Table 4. The accuracy ratio obtained in both models is 0.92. The loss ratio obtained from the CNN algorithm is 0.26462948 while the result obtained from the CNN and BM algorithms is 0.265327107. The average time spent in the implementation of operations is 10 seconds for both cases. The most important result is that the number of features using the CNN algorithm is 1035 features, while the result obtained using both the CNN and BM algorithms is 666 features. Therefore, it can be noted that the superiority of the CNN and BM model over the CNN model in terms of reducing the features of the model while the accuracy remaining the same. The most important benefit of reducing features is that it can be useful in terms of reducing time. In addition, if the accuracy is improving, then the feature reduction can be useful in the process of eliminating the contradiction between the features.

Table 4. Comparing results of loss and accuracy of two layers model

	No. of Nodes	No. of Layers	Accuracy	Loss	Time (seconds)	Features
CNN	128-64	2	0.92	0.26462948	10	1035
CNN & BM	128-64	2	0.92	0.265327107	10	666

5. CONCLUSION

The most important conclusions of this paper are that the supervised learning techniques are better than other techniques because there are labeled data that makes training of models has high performance. In addition, pre-trained models have high power in addressing electricity consumption data because these models are trained using big datasets and powerful computers. When extracting the data of the dataset using normal CNN, the accuracy can be low comparing to addressing electricity consumption data using the SM. In this work, the dataset has been reduced before building the models to increase the performance of building models and classifying new electricity signals. Indeed, using an optimization algorithm (the BM algorithm) leads to reducing the extracted features to speed up the performance of the designed system. Outcomes from these experiments explain that our deep algorithm model (based on CNN and BM) can overtake several other current methods.

REFERENCES

[1] S. Tan, D. De, W. Song, J. Yang, S. K. Das, “Survey of Security Advances in Smart Grid: A Data Driven Approach,” in *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 397-422, Firstquarter 2017, doi: 10.1109/COMST.2016.2616442.

[2] D. Alahakoon, X. Yu, “Smart Electricity Meter Data Intelligence for Future Energy Systems: A Survey,” in *IEEE Transactions on Industrial Informatics*, vol. 12, no. 1, pp. 425-436, Feb. 2016, doi: 10.1109/TII.2015.2414355.

[3] M. Ehsani, Y. Gao, S. Longo, K. Ebrahimi, “Modern electric, hybrid electric, and fuel cell vehicles,” CRC press, 2018.

[4] M. H. Rehmani, M. Reisslein, A. Rachedi, M. Erol-Kantarci, M. Radenkovic, “Integrating Renewable Energy Resources Into the Smart Grid: Recent Developments in Information and Communication Technologies,” in *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 2814-2825, July 2018, doi: 10.1109/TII.2018.2819169.

[5] J. Wu, S. Rangan, H. Zhang, “Green communications: theoretical fundamentals, algorithms, and applications,” CRC press, 2016.

[6] K. Hamedani, L. Liu, R. Atat, J. Wu, Y. Yi, “Reservoir Computing Meets Smart Grids: Attack Detection Using Delayed Feedback Networks,” in *IEEE Transactions on Industrial Informatics*, vol. 14, no. 2, pp. 734-743, Feb. 2018, doi: 10.1109/TII.2017.2769106.

[7] S. Rahman, “The Smart Grid and Its Impact on the Integration of Distributed Energy Resources,” Southeast University Nanjing, 2009.

[8] J. Wu, S. Guo, H. Huang, W. Liu, Y. Xiang, “Information and Communications Technologies for Sustainable Development Goals: State-of-the-Art, Needs and Perspectives,” in *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2389-2406, thirdquarter 2018, doi: 10.1109/COMST.2018.2812301.

[9] R. Jiang, R. Lu, Y. Wang, J. Luo, C. Shen, X. Shen, “Energy-theft detection issues for advanced metering infrastructure in smart grid,” in *Tsinghua Science and Technology*, vol. 19, no. 2, pp. 105-120, April 2014, doi: 10.1109/TST.2014.6787363.

[10] H. Jiang, K. Wang, Y. Wang, M. Gao, Y. Zhang, “Energy big data: A survey,” in *IEEE Access*, vol. 4, pp. 3844-3861, 2016, doi: 10.1109/ACCESS.2016.2580581.

[11] X. Yu, Y. Xue, “Smart Grids: A Cyber–Physical Systems Perspective,” in *Proceedings of the IEEE*, vol. 104, no. 5, pp. 1058-1070, May 2016, doi: 10.1109/JPROC.2015.2503119.

[12] E. T. P. SmartGrids, “European technology platform smart grids: strategic deployment document for Europe’s electricity networks of the future,” *European Technology Platform (ETP) Smartgrids*, p. 69, 2010.

[13] M. B. Line, I. A. Tøndel, M. G. Jaatun, “Cyber security challenges in Smart Grids,” *2nd IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies*, 2011, pp. 1-8, doi: 10.1109/ISGTEurope.2011.6162695.

[14] Z. A. Baig, A.-R. Amoudi, “An analysis of smart grid attacks and countermeasures,” *Journal of Communications*, vol. 8, no. 8, pp. 473-479, 2013.

[15] Y. Deng, S. Shukla, “Vulnerabilities and Countermeasures–A Survey on the Cyber Security Issues in the Transmission Subsystem of a Smart Grid,” *Journal of Cyber Security and Mobility*, vol. 1, no. 2, pp. 251-276, 2012, doi: https://doi.org/10.13052/jcsm2245-1439.1236.

[16] W. Wang, Z. Lu, “Cyber security in the smart grid: Survey and challenges,” *Computer Networks*, vol. 57, no. 5, pp. 1344-1371, 2013, doi: https://doi.org/10.1016/j.comnet.2012.12.017.

[17] N. Komninos, E. Philippou, A. Pitsillides, “Survey in Smart Grid and Smart Home Security: Issues, Challenges and Countermeasures,” in *IEEE Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 1933-1954, Fourthquarter 2014, doi: 10.1109/COMST.2014.2320093.

[18] R. R. Mohassel, A. Fung, F. Mohammadi, K. Raahemifar, “A survey on advanced metering infrastructure,” *International Journal of Electrical Power & Energy Systems*, vol. 63, pp. 473-484, 2014, doi: https://doi.org/10.1016/j.ijepes.2014.06.025.

[19] E. W. S. Angelos, O. R. Saavedra, O. A. C. Cortés, A. N. de Souza, “Detection and Identification of Abnormalities in Customer Consumptions in Power Distribution Systems,” in *IEEE Transactions on Power Delivery*, vol. 26, no. 4, pp. 2436-2442, Oct. 2011, doi: 10.1109/TPWRD.2011.2161621.

[20] O. Abdel-Hamid, L. Deng, D. Yu, “Exploring convolutional neural network structures and optimization techniques for speech recognition,” in *Interspeech*, vol. 11, pp. 3366-3370, 2013.

[21] S. Mallat, “Understanding deep convolutional networks,” *Philosophical Transactions A*, vol. 374, no. 2065, p. 20150203, 2016.

[22] Krizhevsky, I. Sutskever, G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[23] R. R. Bhat, R. D. Trevizan, R. Sengupta, X. Li, A. Bretas, “Identifying Nontechnical Power Loss via Spatial and Temporal Deep Learning,” *15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 272-279, 2016, doi: 10.1109/ICMLA.2016.0052.

[24] Z. Zheng, Y. Yang, X. Niu, H. Dai, Y. Zhou, “Wide and Deep Convolutional Neural Networks for Electricity-Theft Detection to Secure Smart Grids,” in *IEEE Transactions on Industrial Informatics*, vol. 14, no. 4, pp. 1606-1615, April 2018, doi: 10.1109/TII.2017.2785963.

[25] M. Mahmood, B. Al-Khateeb, “The blue monkey: A new nature inspired metaheuristic optimization algorithm,” *Periodicals of Engineering and Natural Sciences*, vol. 7, no. 3, pp. 1054-1066, 2019, doi: http://dx.doi.org/10.21533/pen.v7i3.621.

BIOGRAPHIES OF AUTHORS

Noor Mahmoud Ibrahim received the B.Sc. degree from the Department of computer science of Al-Ma'aref University College in 2016. She is now working for the M.Sc. at the College of Computer Science and Information Technology, University of Anbar. She is interested in deep learning and information security and conducting her master's thesis in this area.

Prof. Sufyan T. Faraj Al-Janabi obtained his B.Sc. (1992), M.Sc. (1995), and Ph.D. (1999) in Electronic and Communications Engineering from the College of Engineering, Al-Nahrain University in Baghdad. His research interest includes internet protocols, information security, and quantum cryptography. Prof. Al-Janabi is the winner of the 1st Award for the Best Research Paper in Information Security from the Association of Arab Universities (AARU), Jordan, 2003. He is a member of ACM, ASEE, IACR, and IEEE.

Prof. Belal Al-Khateeb received the B.Sc. (honors) (first class) degree in computer science from Al-Nahrain University, Baghdad, IRAQ, in 2000, and the M.Sc. degree in computer science from Al-Nahrain University, Baghdad, IRAQ, in 2003, and the Ph.D. degree from the School of Computer Science, University of Nottingham, Nottingham, U.K., in 2011. He is currently a professor at the College of Computer Science and Information Technology, University of Anbar. He has published over 64 refereed journal and conference papers. His current research interests include evolutionary and adaptive learning particularly in computer games, expert systems, and heuristics and meta/hyper-heuristics. He has a particular interest in computer games programming. Prof. Al-Khateeb is a reviewer of fifteen international journals (including three IEEE Transactions) and twenty conferences.