Non-technical Loss Detection Using Deep Reinforcement Learning for Feature Cost Efficiency and Imbalanced Dataset

Jiyoung Lee1, Student Member, IEEE, Young Ghyu Sun1, Student Member, IEEE, Isaac Sim1, Student Member, IEEE, Soo Hyun Kim1, Student Member, IEEE, Dong In Kim2, Fellow, IEEE, and Jin Young Kim1, Senior Member, IEEE

1Department of Electronic Convergence Engineering, Kwangwoon University, Seoul 01897, Korea
2Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon 16419, Korea

Corresponding author: Jin Young Kim (e-mail: jinyoung@kw.ac.kr).

This work was supported in part by “Human Resources Program in Energy Technology” of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), granted financial resource from the Ministry of Trade, Industry & Energy, Republic of Korea. (No. 20194010201830), in part by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2022-0-01846) supervised by the IITP (Institute for Information & communications Technology Promotion), and in part by Kwangwoon University in 2021.

ABSTRACT One of the problems of the electricity grid system is electricity loss due to energy theft, which is known as non-technical loss (NTL). The sustainability and stability of the grid system are threatened by the unexpected electricity losses. Energy theft detection based on data analysis is one of the solutions to alleviate the drawbacks of NTL. The main problem of data-based NTL detection is that collected electricity usage dataset is imbalanced. In this paper, we approach the NTL detection problem using deep reinforcement learning (DRL) to solve the data imbalanced problem of NTL. The advantage of the proposed method is that the classification method is adopted to use the partial input features without pre-processing method for input feature selection. Moreover, extra pre-processing steps to balance the dataset are unnecessary to detect NTL compared to the conventional NTL detection algorithms. From the simulation results, the proposed method provides better performances compared to the conventional algorithms under various simulation environments.

INDEX TERMS Deep reinforcement learning, non-technical loss (NTL), energy theft, feature cost efficiency, data imbalanced problem

I. INTRODUCTION

Since the advent of advanced metering infrastructure (AMI) in smart grid (SG), electricity consumption information of the users is analyzed at the data center of power utilities. Malfunction detection, demand forecast, and electricity overload detection are various applications provided by the utilities to grid users. The applications are derived by analysis of the electricity data. Moreover, AMI has enabled the utilities to detect and report non-technical loss (NTL) in their power transmission and distribution (TD) network [1]. NTL is caused by the illegal action of consumers or malfunctions of metering devices unlike technical loss (TL) in the TD process as power transmission loss. The malicious actions to exploit energy from the utilities are known as energy theft or electricity theft. The menace approaches are led to a degradation of the overall quality of distributed power and stability of the grid system. Especially, NTL caused by fraudulent behavior can impose a direct economic loss on the power utilities and legal user. The results have been reported in various countries including Turkey [2], Jamaica [3] and India [4]. A wide range of approaches has been taken to mitigate the NTL problem.

In [6], the solutions for NTL problems are broadly categorized into three groups: theoretical studies, hardware solutions, and non-hardware solutions. Theoretical studies are based on socio-economic and demographic knowledge. Hardware solutions are the approaches of employing metering equipment and constructing different electricity network frameworks or infrastructures. Non-hardware solutions are the methods to solve NTL problems with the collected data from the grid system.

In [7], non-hardware solutions are divided into three subcategories: network-oriented methods, data-oriented
methods, and hybrid methods. The standard of dividing each subcategory is defined by the data to be analyzed.

Network-oriented methods are adopted to observe the presence of deceiving behavior in the TD network with the aid of network configurations or network topologies. The network-oriented methods are implemented by extra network unit to control electricity flows. This is observed in the various studies such as deployment of feeder remote terminal unit with dynamic programming to build network configuration [8], deployment of intermediate monitor meter for load flow control [9]. Data-oriented methods are utilized to examine the personal information of people connected to the TD line of AMI. The electricity data are energy consumption records or client registration information. Hybrid methods are implemented to detect theft by combining data and network methods. The hybrid method is dealt with some studies; observer meter with support vector machine (SVM)-based data analysis [10], NTL detection system adopting voltage sensitivity analysis, power system optimization, and SVM [11].

Adaptation of new processes is continuously necessary for network-oriented methods. This is due to the changes in the infrastructure of the electricity grid system and the amount of electricity demand [12]. On the other hand, the data-oriented methods are more utilizable and flexible approaches to NTL detection. This is because restrictions from the improvement of hardware structure are unconcerned.

Data-oriented methods are implemented by traditional machine learning (ML)-based classification methods [13]-[16] and by artificial neural network (ANN)-based methods [17]-[18]. In recent years, the various ensemble models are introduced to detect NTL [5], [22]. Data imbalanced problem between normal users and illegal thieves is a major problem of the data-oriented methods as mentioned in [5], [13], [18], [19], [28]. However, in the previous articles, data preprocessing steps are required to balance the dataset unless the balanced dataset is dealt with. The problem is solved by the data augmentation method as oversampling or generating synthetic data in most of the previous works as in [5], [13], [18], [28]. Moreover, the previous solutions assume that all the gathered features of AMI are usable at the instant of theft detection without [13]. Feature cost efficiency (FCE) is an important problem in classification tasks as [20].

In this paper, the DRL-based NTL detection method is applied to use partial input features for classification. The proposed method is one of the data-oriented methods to identify the existence of the NTL from the electricity usage dataset. The proposed DRL algorithm is suggested to solve the class imbalance problem of the NTL dataset. The advantage of the proposed classification method is that the imbalance dataset is used without changes. The two datasets used in this paper are gathered from Korea Electric Power Corporation (KEPCO) [27] and Irish Social Science Data Archive [30].

The major contributions and innovations of the proposed method are summarized as follow:

1) The DRL algorithm is applied to classify abnormal users from the imbalance dataset.
2) FCE condition is achieved.
3) One dimensional-convolutional neural network (1D-CNN) is applied as the neural network structure to improve the detection performance of time-series data.

The rest of this paper is organized as follows. Section II describes the proposed system model for NTL detection based on deep Q-learning (DQL), which is one of DRL algorithms. The DQL-based NTL detection method is explained in Section III. The simulation environments and results for the KEPCO dataset are presented in Section IV. In the Section IV, performance metrics and conventional NTL detection algorithms are explained. In Section V, the convergence behaviors of the 1% mixed imbalanced Irish datasets are represented by average rewards of reinforcement learning (RL). Finally, the paper is concluded in Section V.

II. SYSTEM MODEL FOR NTL DETECTION

The overall system model for NTL detection is illustrated in Fig. 1. Load profiles of grid systems are normalized to range from 0 to 1 using min-max scaler. The equation for min-max scaler is given by

\[ x_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \]  

where \( x \) is one sample of the recorded electricity usage data from the dataset \( X \), \( x_i \) is the \( i \)th recorded value of \( x \), \( \min(x) \) is the minimum value of \( x \) and \( \max(x) \) is the maximum value of \( x \).

After the normalization process, a pre-processed dataset is generated with a false data injection (FDI) process. The dataset is divided into two datasets to train and test deep Q-learning network (DQN). The network is to decide an action of the
DRL algorithm. The trained neural network is used to detect NTL from the test dataset.

KEPCO and Irish datasets are used in the simulation process. The two datasets are comprised of legally registered users. Therefore, the FDI process is necessary to generate simulation datasets for NTL detection.

FDI is implemented in various research projects as in [10], [11], [13]-[16], and [19]. The total number of types of FDI is set to 6 in this paper as given in the previous papers. First, the above five types in [10] are used in this paper to inject false data, which are widely adopted FDI methods. The reason for selecting the five types of FDI without type 6 is to generate practical theft patterns mentioned in [13]. In addition to the five types, the FDI type 6 in [10] is replaced with the revised version of cut-off error, which is type 2 in [14], [15]. Each type of FDI used in this paper is mentioned as FDI1, FDI2, FDI3, FDI4, FDI5, and FDI6, respectively.

The details of DQN-based classification method are explained in the following sections.

III. DQN-BASED NTL DETECTION METHOD

A. CLASSIFICATION METHOD USING DQN FOR NTL DETECTION

The main problem of NTL detection is that the number of illegal users is less than legal users in the collected dataset. In this paper, the classification of thieves among users is executed with DQN. The proposed NTL detection method is an improved version of the previous DQN-based FCE classification model [20] for an imbalanced dataset.

The purpose of using the DQN algorithm for NTL detection is to optimize the trade-off relationship of classification error of imbalanced dataset and FCE. The DQN algorithm is suggested to optimize the equation as

$$ \arg\min_{\theta} \frac{1}{B_n} \left[ \sum_{D} (e_{\theta^+} + e_{\theta^-}) + \sum_{f \in \hat{F}} c(f) \right], \quad (2) $$

where the input dataset is denoted as $D(X,Y)$. The size of $D(X,Y)$ is represented as $D_n$. $y \in Y$ is a specific class of $x$ in the class set $Y$. Each time step of $x$ is denoted as $i$th feature, $f_i \in F$, where $F$ is input feature space. $D_n$ is the total number of $D$. $e_{\theta^+}$ means the classification error of a majority class. $e_{\theta^-}$ means the classification error of a minority class. $\hat{F}$ is denoted as a selected feature set for the NTL classification. $c(f)$ is the feature cost function for feature selection applying as a reward of the DRL. The value of $c(f)$ is unified in this NTL problem.

The components of DRL are agent, environment, state $s$, reward $r$, and action $a$. Each load profile of users is represented as an episode of DQN. For each state $s = (x, y, \hat{F})$ of an episode, an agent performs one of the actions in an action space $A = \{ A_f, A_c \}$. In the action space, $A_f$ is a feature-selecting action and $A_c$ is a classification action terminating the episode. The agent observes the partial information of $s$ comprised of $(x_F, \hat{F})$, where $x_F$ represents selected data corresponding to $\hat{F}$. In the return of a specific $a \in A$, the $r$ is given by the environment. In this paper, the reward in [20] is modified to (3) for imbalanced dataset. For each state, $r$ of the $a$ is defined as follows:

$$ r \left( (x, y, \hat{F}), a \right) = \begin{cases} -c(a) & \text{if } a \in A_f \\ -\tau & \text{if } a \in A_c \text{ and } a \neq y_M \\ -1 & \text{if } a \in A_c \text{ and } a \neq y_m \\ \tau & \text{if } a \in A_c \text{ and } a = y_M \\ 1 & \text{if } a \in A_c \text{ and } a = y_m \end{cases}, \quad (3) $$

where $y_M$ denotes the label of the majority class, $y_m$ denotes the label of the minority class. A reward of the majority class, which is abnormality ratio (AR), is denoted as $\tau$. The concept of $\tau$ is same as imbalance ratio in [21]. The AR is represented as follows:

$$ \tau = \frac{\text{the number of minority class samples}}{\text{the number of majority class samples}} \quad (4) $$

The transition function $t(s, a)$ of environment to get next state $s'$ is denoted as follows:

$$ t \left( (x, y, \hat{F}), a \right) = \begin{cases} T & \text{if } a \in A_c \\ (x, y, \hat{F}^r) & \text{if } a \in A_f \end{cases}, \quad (5) $$

where $T$ is a terminal state. $\hat{F}^r$ is containing the newly selected feature based on $\hat{F}$.

The aim of DQN is to find optimal Q-function $Q^*$ with an optimal policy $\pi^*$ which chooses $a$ at $s$ to maximize Q-value using neural network. The Bellman equation with discount factor $\gamma$ is expressed as follows:

$$ Q^*(s, a) = E_{s' \sim t(s,a)} \left[ r(s, a, s') + \gamma \max_{a'} Q^*(s', a') \right], \quad (6) $$

where $r(s, a, s')$ is the received reward.

The Q-function of the neural network is expressed as $Q^\theta(s, a)$ with network parameter $\theta$. Huber loss function is used as a loss function of the neural network to find $\theta$. $Q^\theta$ converges to $Q^*$ when the error between $q$ and $Q^\theta$ decreases. Cost function $C$ of batch $B$ for training the neural network is dictated as follows:

$$ C = \frac{1}{B_n} \sum_{\theta} l(\theta), \quad (7) $$
where $B_n$ is the size of $B$. Huber loss function is used for loss function $l(\theta)$ for NTL detection, which is denoted as:

$$l(\theta) = \begin{cases} 0.5\eta^2 & \text{if } |\eta| < 1 \\ |\eta| - 0.5 & \text{otherwise} \end{cases}$$

(8)

where $\eta = q - Q^\theta(s, a)$. The $q$ is defined as:

$$q = \begin{cases} r & \text{if } s' = T \\ r + \gamma \max_{a'} Q^{\theta'}(s', a') & \text{if } s' \neq T' \end{cases}$$

(9)

where $Q^{\theta'}$ is Q-function from the target network with target network parameter $\theta'$.

**B. Neural network model for Q-function based on 1D-CNN**

The neural network configuration of [20] is replaced with 1D-CNN because KEPCO and Irish datasets consist of time-series electricity usage data. 1D-CNN has an advantage of local pattern of time series data [23]. The 1D-CNN is applied to solve classification problems of time series data as [24], [25].

The experiments are conducted to compare the performance of 1D-CNNs and fully connected layer-based neural networks (FNNs). The dataset for simulations is 1% imbalanced dataset with mixed FDI types. The neural network in [20] is referred to as baseline model. The simulations are conducted 100 times for the two neural network structures to experiment with the efficiency of the 1D-convolutional layer for electricity data. Seven neural networks are constructed for the simulations, which are three FNNs and four 1D-CNNs. The performance metrics and details of simulation environment are explained in Section IV-A.

The baseline model is comprised of 3 fully connected layer-based hidden layers, 1,024 units, and rectified linear unit (ReLU) activation function. The structure of 1D-CNN is comprised of two convolution layers with average pooling and fully connected layers, which is illustrated in Fig. 2. The number of units of fully connected layers in 1D-CNN is equal to 1024 as the baseline model. All 1D-convolution layers output channel size is 128 with kernel size 2 and stride 1 with average pooling of kernel size 2 and stride 2. ReLU activation function is used for all layers. The input channel size of the proposed 1D-CNN is 2. The two input channels are for $x_F$ and $x_F'$, respectively. In the proposed 1D-CNN network, the correlation between $x_F$ and $x_F'$ is considered by two input channels and average pooling process. This is underestimated by the input structure of the baseline. $Q^\theta(s, a)$ is the output of the network.

According to the simulation result listed in Table 1, average F1-scores of 1D-CNNs are higher than FNNs. Moreover, the values are similar for FNNs regardless of increasing the number of hidden layers. The average F1-score is increased at least 4.32% by changing the network structure from FNN to CNN. From the perspective of the average value, increasing the hidden layers of FNN is ineffective for improving the performance of NTL detection. Although the maximum and
average F1-scores of C2F5 are the highest among the neural networks, the standard deviation of the C2F5 is marked higher than C2F4. C2F4 is selected for the neural network structure in the following simulations because of the lowest standard deviation with prominent average F1-score.

Simulated neural network structures are listed below:
1) Baseline: three fully connected layers
2) FNN4: four fully connected layers
3) FNN5: five fully connected layers
4) C2F2: two 1D-convolutional layers with average pooling + two fully connected layers
5) C2F3: two 1D-convolutional layers with average pooling + three fully connected layers
6) C2F4: two 1D-convolutional layers with average pooling + three fully connected layers
7) C2F5: two 1D-convolutional layers with average pooling + five fully connected layers

C. ALGORITHMS OF NTL DETECTION USING DQN
Steps to pre-process energy datasets are summarized in Algorithm 1. First, a min-max scaler is applied to the original dataset to normalize the dataset for the neural network. The normalized data is modified with selected FDI types. The scaled dataset is combined with FDI data sampled with an abnormality ratio. The combined dataset is shuffled to be separated into the training and test datasets. The validation dataset is included in the training dataset.

The DQN-based training algorithm is represented as Algorithm 2. In the initialization stage, all the neural network parameters \( \theta \) and \( \theta' \) are randomly initialized. The parallel environments \( E \) are initialized with new episodes with the initially state \( s_0 \) which is \((x, y, f_0)\). \( f_0 \) is the initial selected feature vector with all zeros. The buffer \( M \) is initialized with the transition \((s, a, r, s', \tau)\) following \( \varepsilon \)-greedy policy \( \pi_\varepsilon \) until the number of transitions reaches buffer size \( N \). For each epoch, the transition is continuously collected by \( \pi_\varepsilon \) and recorded to \( M \) replace the old transition. After saving the transition, batch \( B \) is randomly sampled from \( M \) with the predetermined batch size \( B_N \). \( q \) is calculated by (9) for the \( B \). The neural network parameter \( \theta \) is obtained by (7), (9) using Adam optimizer [26]. The target network parameter is updated with soft replacement.

The mechanism of the step function \( \text{STEP}(s, a) \) is explained in Algorithm 3. For the feature selecting an action, the newly selected feature is added to the previously selected feature vector \( \hat{F} \). The value of each selected feature of \( \hat{F} \) the is turned to 1 from 0. The returns of \( \text{STEP}(s, a) \) are \( s' \) and \( r \). The \( s' \) has three components: a vector for selected data \( x_{F'} \) corresponding to \( \hat{F}, y, \) vector for selected features \( \hat{F}' \). The \( \hat{F}' \) is updated from \( \hat{F} \) to reflect the result of \( a \). The value of \( x_{F'} \) is derived from the inner product between \( x \) and \( \hat{F}' \). The \( r \) of the newly selected feature is \(-c(a)\). Four different rewards are used for each classification action for the imbalanced dataset, whereas two different reward values are needed for the

---

**Algorithm 1. Pre-processing energy dataset**

- Import KEPCO energy usage dataset
- Normalize dataset using min-max scaler
- Select correct type(s) of FDI
- for all selected FDI do
  - FDI modification of normalized data
- end for
- Decide ratio of abnormality
- Sample FDI data
- Concatenate FDI data with original data
- Shuffle concatenated data
- Separate dataset into training and test dataset

---

**Algorithm 2. DQN training**

- Randomly initialize network parameters \( \theta \) and \( \theta' \)
- Initialize the parallel environments \( E \) with \( s_0 \)
- Initialize the replay buffer \( M \) with size \( N \)
- for Each epoch until converged do
  - for \( e \in E \) do
    - Simulate one step with \( \varepsilon \)-greedy policy \( \pi_\varepsilon \)
    - if \( |M| \geq N \) then
      - Deleting old transition
    - end if
    - Add transition \((s, a, r, s')\) into buffer \( M \)
  - end for
  - Sample a random batch \( B \) from \( M \)
  - for \((s, a, r, s') \in B\) do
    - Calculate \( q \) with (9)
  - end for
  - Find \( \theta \) with Adam optimizer with (7), (9) using soft replacement \( \theta' \leftarrow \rho \theta + (1 - \rho)\theta' \)
- end for

---

**Algorithm 3. DQN Environment for NTL detection**

- for \( \text{STEP}(s, a) \) do
  - if \( a \in A_f \) then
    - \( \hat{F}' \leftarrow \hat{F} + a \)
    - Return \((x_{F'}, y, \hat{F}', -c(a))\)
  - else if \( a \in A_c \) then
    - if dataset = imbalanced
      - \( r \) = \( \tau \) if \( a \in A_c \) and \( a \neq y_M \)
      - \( r \) = \(-1 \) if \( a \in A_c \) and \( a \neq y_m \)
      - \( r \) = \( \tau \) if \( a \in A_c \) and \( a = y_M \)
      - \( r \) = \( 1 \) if \( a \in A_c \) and \( a = y_m \)
    - else
      - \( r \) = \(-1 \) if \( a \in A_c \) and \( a \neq y \)
      - \( r \) = \( 0 \) if \( a \in A_c \) and \( a = y \)
  - end if
  - Reset environment with \( s_0 \) of new episode from dataset
  - Return \((T, r)\)
- end if
balanced dataset. Each reward value of an imbalanced dataset is defined as (3). After the classification action, the environment is reset with a new episode from the electricity dataset. The classification action terminates the episode returning the classification reward.

IV. SIMULATION RESULTS FOR KEPCO DATASET

A. SIMULATION ENVIRONMENT

All simulations are executed in a PC with i5-9400F CPU, GeForce GTX 1650 GPU, and 32-GB RAM. The program for coding is PyCharm 2019 with Python 3.6. KEPCO dataset is comprised of local home and business electricity usage profiles from 2017-2019. The number of grid users of the dataset is 6,782 with the data containing information more than half of the total. The time-step of the data is 36 because the company provides the monthly consumption energy data.

The feature cost of \( c(f) \) of NTL detection based on DQN is set to 1/3600 for all the simulation settings. The cost has an impact on the selected feature number. An increase in feature cost leads to a decrease in selected feature numbers. In this paper, the lowest ratio of abnormal users for simulation is 1%. The total feature cost is decided to be 0.01, which is 100 times smaller than the reward of the minority class. \( \gamma \) is 0.99, \( \rho \) is 0.01. The buffer size is 30,000. The batch size is 2,000.

The training dataset is separated into training and validation datasets. The proportion of training, validation, the test dataset is 3:1:1. More detailed settings are following the parameter settings in [20]. The specification of implemented C2F4 is summarized in Table 2.

Three different experiments are simulated under various FCE conditions. The first simulation is for measuring the performance of the proposed method changing AR. The theft data is made of every FDI types, which is denoted as mixed types. The second simulation is executed with different FDI types for the balanced dataset. The last experiment is for the extreme imbalanced case. The details of data for each simulation are listed in Table 3.

B. PERFORMANCE METRICS AND DIFFERENT DATA ANALYTIC METHODS FOR NTL DETECTION

In this paper, four classification metrics are used to demonstrate the performance of the proposed method compared to conventional algorithms. The metrics are true-positive rate (TPR), false-positive rate (FPR), false omission rate (FOR), and F1-score (F1). The equations are listed as follows:

\[
TPR = \frac{True \ positive}{Condition \ positive}, \quad (10)
\]

\[
FPR = \frac{False \ positive}{Condition \ negative}, \quad (11)
\]

\[
FOR = \frac{False \ negative}{Predicted \ negative}, \quad (12)
\]

\[
F1 - score = 2 \times \frac{Precision \times \ Recall}{Precision + Recall}, \quad (13)
\]

where positive is theft class data. negative is normal class data. Condition means the actual class of the data. Predicted means the classified state of the data. recall is equal to TPR. Precision of F1 is given by

\[
Precision = \frac{True \ positive}{Predicted \ positive} \quad (14)
\]

The direct comparison of our method with other algorithms is impossible due to the difference in the number of input features. Therefore, input features of the conventional
algorithm are selected by feature selection function based on the best score of analysis of variance (ANOVA). The number of selected features is closely selected as the used feature number of our method.

Six different classification methods are selected to compare with the proposed method. Four Conventional methods and two neural network-based models are selected as below:

1) SVM [10]: SVM is a widely used algorithm for classification and used as a ground truth algorithm of NTL detection in [13], [15], [16], [19]. The parameter of SVM is set as $C=50, y_{\text{s SVM}}=0.1$. The radial basis function is used as the SVM kernel.

2) K-nearest neighbors (KNN): KNN is used in [19]. The parameters of KNN are set as follows: the number of neighbors is 5; Euclidean distance is used to implement KNN by setting the parameter $p=2$.

3) Extreme gradient boosting (XGB) [13], [19]: According to [19], the performance of the XGB is the highest on a balanced dataset among 11 algorithms. Therefore, this algorithm is selected as a comparison algorithm. The parameters of XGB are set as [19], which are listed as follows: booster is ‘gbtree’; learning rate is 0.05, maximum depth is 8, lambda for L2 regularization is 2.

4) Random Forest (RF): RF is also listed in [19]. The overall performance of RF is highly ranked. The parameters are the same as follows: the number of estimators is 50; the maximum depth of tree is 6.

5) Simple C2F4: Simple C2F4 is the neural network-based classifier constructed without RL. The detailed structure is same as the proposed C2F4 without the input. The size of the input of the Simple C2F4 is sample to the used input feature number, which is equal to 1 the used input size. This is because the inputs are comprised of the selected input features without $F^r$. The early stopping tolerance is 5 with the total epoch of 1,000.

6) Simple Baseline: Simple Baseline is the neural network-based classifier constructed without RL. The detailed structure is same as the Baseline model without the input size. The details of the implementation are equal to the Simple C2F4.

### C. SIMULATION RESULTS OF DIFFERENT ABNORMALITY RATIOS

Under the circumstance of mixed FDI types, the ratio difference between normal users and abnormal users has an impact on the results. ARs for each simulation is 0.1, 0.2, 0.3 and 0.5. For each AR, the numbers of used features for classification are 22, 20, 14, and 12. The numbers out of total feature number are listed below each AR, respectively. The numbers are selected automatically by DRL classification algorithms. The simulation results are shown in Table 4.

From TPR values, our method marks the highest TPR values for all ARs as the lowest value is 93.08% for AR=0.1 and the highest value is 98.3% for AR=0.5. TPR value is related to the correctly classified thieves among all thieves. It is demonstrated that the abnormal users are effectively classified by our method compared to other methods.

From FPR values, the FPR values of our method are constantly marked values around 5 to 6.5 for all ARs. In contrast, FPR of other methods normally tends to decrease as the AR decreases. Higher FPR values indicate the missing normal users among all normal users. The classification of normal data is less affected by the AR for our method, where the difference of minimum and maximum value is 1.49. In contrast to the proposed method, the minimum value of difference is 7.71 for Simple Baseline. The simulation result can be interpreted that the incorrectly classified normal users are increased as the number of thieves increases.

From FOR values, the FOR values of our method are below one for all the ratios with the lowest value. Higher FOR value means that the ratio of undetected theft is higher without detection. Therefore, the ratio of undetected energy thieves among predicted normal users is relatively lower than conventional algorithms.

F1 is related to the number of overall correctly classified imbalance data. For F1, our method shows the highest performance for AR=0.3 and 0.5. For the case of AR = 0.1 and 0.2, the total number of correctly classified users of the proposed method is relatively lower than the Simple Baseline model. However, the ratios of correctly classified thieves are

### TABLE IV SIMULATION RESULTS OF NTL DETECTION METHODS FOR DIFFERENT AR.

| Detection Methods | AR=0.1 (# of features: 22/36) | AR=0.2 (# of features: 20/36) | AR=0.3 (# of features: 14/36) | AR=0.5 (# of features: 12/36) |
|-------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
|                   | TPR | FPR | FOR | F1   | TPR | FPR | FOR | F1   | TPR | FPR | FOR | F1   | TPR | FPR | FOR | F1   |
| Proposed method   | 94.62 | 4.99 | 0.55 | 76.64 | 95.93 | 6.48 | 0.86 | 83.95 | 96.54 | 5.64 | 0.97 | 90.3 | 98.3 | 6.17 | 0.95 | 93.66 |
| SVM               | 73.08 | 1.25 | 2.53 | 78.51 | 77.04 | 2.65 | 4.48 | 80.93 | 86.3 | 5.67 | 7.19 | 80.34 | 92.35 | 13.46 | 5.72 | 84.84 |
| KNN               | 88.46 | 4.7  | 1.14 | 74.43 | 92.22 | 10.67 | 1.14 | 75   | 84.44 | 7.36 | 6.26 | 76.25 | 93.06 | 17.14 | 4.26 | 82.59 |
| XGB               | 80.77 | 3.67 | 1.87 | 73.68 | 85.93 | 8.98 | 2.98 | 74.36 | 80.37 | 4.49 | 8.73 | 79.2 | 89.52 | 11.65 | 5.92 | 84.66 |
| RF                | 56.92 | 1.03 | 3.99 | 67.89 | 77.41 | 4.64 | 4.5  | 77.12 | 64.81 | 3.02 | 11.42 | 72.02 | 85.13 | 12.48 | 8.27 | 81.6  |
| Simple C2F4       | 91.04 | 5.52 | 0.93 | 73.72 | 94.81 | 9.05 | 1.12 | 78.89 | 82.49 | 11.35 | 6.06 | 75.93 | 90.93 | 15.41 | 5.38 | 82.68 |
| Simple Baseline   | 89.55 | 2.58 | 1.05 | 83.04 | 86.67 | 1.84 | 2.63 | 88.47 | 85.02 | 6.99 | 4.99 | 82.37 | 87.82 | 9.55 | 6.67 | 85.34 |
most of the data from the negative class is detected as the datasets from performance metrics of Table 4. On the contrary, Therefore, the proposed method is robust to imbalanced direct degrade the performance of the overall grid system. value for NTL detection because undetected abnormal users related to undetected energy theft. FOR is the most important value are defective because the method tends to decide all the users.

For each FDI type, all the simulation results are shown in D. SIMULATION RESULTS OF DIFFERENT FDI TYPES

| Detection Methods | FDI1 (# of features: 15/36) | FDI2 (# of features: 19/36) | FDI3 (# of features: 15/36) |
|-------------------|-----------------------------|-----------------------------|-----------------------------|
|                   | TPR | FPR | FOR | F1   | TPR | FPR | FOR | F1   | TPR | FPR | FOR | F1   |
| Proposed method   | 97.6 | 5.67 | 2.54 | 96.09 | 97.68 | 4.87 | 2.31 | 96.39 | 98.91 | 5.6 | 1.17 | 96.79 |
| SVM               | 98.76 | 16.19 | 1.49 | 92.06 | 88.86 | 10.45 | 11.31 | 89.29 | 98.91 | 10.6 | 1.24 | 94.54 |
| KNN               | 99.2 | 20.45 | 1.02 | 90.53 | 80.86 | 5.97 | 17.27 | 86.63 | 97.96 | 10.07 | 2.27 | 94.29 |
| XGB               | **99.42** | 15.15 | 0.7 | 92.83 | 92.14 | 2.46 | 7.63 | 94.72 | 99.27 | 11.04 | 0.83 | 94.53 |
| RF                | 97.16 | 15.9 | 3.34 | 91.38 | 91.19 | **2.39** | 8.47 | 94.25 | **100** | 13.96 | 0 | 93.63 |
| Simple C2F4       | 96.43 | 5.82 | 3.74 | 95.43 | 88.65 | 8.66 | 11.30 | 89.96 | 99.56 | 9.03 | 0.49 | 95.56 |
| Simple Baseline   | 96.07 | **5.60** | 4.09 | 95.34 | 87.34 | 6.49 | 12.19 | 90.19 | 98.84 | 6.12 | 1.26 | 96.52 |

the highest for the proposed method as shown in TPR and FOR. From the results, the proposed method has strength in detecting theft data at the expense of misclassified registered users.

Algorithms performing a low FPR value with a low TPR value are defective because the method tends to decide all the inputs as a majority class. Moreover, the high FOR value is related to undetected energy theft. FOR is the most important value for NTL detection because undetected abnormal users directly degrade the performance of the overall grid system. Therefore, the proposed method is robust to imbalanced datasets from performance metrics of Table 4. On the contrary, most of the data from the negative class is detected as the positive class by other methods. In the point of FCE condition, the number of used features is increased as AR decreases.

D. SIMULATION RESULTS OF DIFFERENT FDI TYPES FOR BALANCED DATASET

For each FDI type, all the simulation results are shown in Table 5. The used feature numbers for each FDI type are 15, 19, 15, 9, 9, and 22. The numbers out of total feature number are listed below each FDI type, respectively. Each method has strength in each FDI type. Our method is robust to balanced dataset compared to conventional methods due to the highest F1 for all FDI types. The minimum value of difference between the proposed method and the other methods is 0.33 occurred in FDI4 compared with C2F4. The maximum value is 11.49 occurred in FDI6 compared with KNN.

For FDI1, the XGB method effectively detects thieves among all methods from TPR = 99.42 and FOR = 0.7. However, the normal users are incorrectly determined as illegal users by the XGB, which is demonstrated by FPR = 15.15. For the proposed method, the number of correctly classified normal users is relatively higher than the XGB because the FPR value of our method is 5.67.

For FDI2, the thefts are effectively detected by the proposed method from the highest TPR (97.68) and FOR (2.31). The normal users are misclassified by the RF with the lowest FPR (2.39). The minimum difference of FOR values between the proposed method and the conventional methods is equal to 5.32, which is the highest value among all FDI types. This can be interpreted as the ratio of undetected theft is higher with the conventional methods for FDI2.

For FDI3 and FDI4, the number of properly classified normal users is the highest by the proposed method among all classification methods. This is because F1=96.79 for FDI3 and F1=99.82 for FDI4. From the TPR and FOR, all thieves are correctly classified by the RF for FDI3. Form the two metrics, The SVM, KNN, and Simple Baseline methods detect all illegal users for FDI4. For the two FDI types, the proposed method efficiently detects legally registered user with the expense of thieves, which is demonstrated by the lowest FPR.
value. The two FPR values of the proposed method are 5.6 and 0.15 for FDI3 and FDI4.

For FDI5, all values of the performance metrics are the best for the proposed method. This means that the proposed method is efficient for FDI5 compared to the other methods.

All classification methods have difficulty in FDI6, which is shown by the relatively low F1 values. The number of correctly detected thefts is the highest by the XGB for FDI6 from TPR. However, the number of the misclassified normal users by the XGB is higher than the proposed method. This is shown by the FPR value and the difference of FPR values is 10.38. From the FPR value, the normal users are effectively classified by the Simple Baseline with lots of misclassified thefts. This is because the FPR value of Simple Baseline is marked the lowest value of 4.1 with the highest FOR value of 21.36.

The used feature number for FDI4 and FDI5 is relatively lower than other types with better performance of all four metrics. For our method, F1 values are 99.89% and 98.63% for FDI4 and FDI5, respectively. This can be interpreted from the perspective of FCE. For the two FDI types, the expense of using all input features to detect NTL is relatively critical to the grid system compared to other FDI types because the proposed method uses 25% of total input features to detect NTL.

E. RESULTS OF THE EXTREME IMBALANCED NTL CASE

In the extreme case as [19], the AR of the extreme simulation case is set to 1% mixed FDI types. From the proposed method, the used feature number of the proposed method is 3, which is 8.3% of the total input features. For the simulation environment, the performance of the proposed method is compared with SVM, which is relatively robust to the imbalanced dataset in the previous section.

Fig. 3(a) is a confusion matrix of the proposed DQN-based detection method, and Fig. 3(b) is the result of SVM. From the confusion matrix of our method, about 85% of thieves are detected by the DRL method. Moreover, only 7.9% of the normal users are incorrectly determined as abnormal users by our method. Although the total number of incorrectly classified data by SVM is lower than our method, every theft data in the test dataset is classified as normal by SVM. From the result of confusion matrices, the SVM-based NTL detection method is inadequate for an extremely imbalanced dataset compared to the proposed method.

F. CONVERGENCE BEHAVIOR FOR 1% IMBALANCED DATASET

The convergence behavior is an important aspect of ANN. In this paper, the performance metric adopted as convergence behavior is an average reward. The average reward is a performance metric for the overall performance of the RL algorithm. The tests are executed under a 1% imbalanced KEPCO dataset of mixed FDI types to investigate the behavior of the proposed 1D-CNN structures compared to the FNN structures with the baseline model in Section III. B. For each structure, the experiments are executed ten times as in [20]. The mean of the simulated average rewards over training epochs is shown in Fig. 4. The summaries of simulations are listed in Table 6. The average reward is calculated every 200 epochs in the simulations.

In Fig. 4, the solid line is dedicated to 1D-CNN models and the dashed line is dedicated to FNN models. As mentioned in [20], the performances of the networks are improved as the training process. From Table 6, the overall average rewards of the 1D-CNNs are marked higher than the FNNs except for C2F5. The highest average reward is achieved from C2F4, which is the main network structure in this paper. The average

FIGURE 3. Confusion matrices for SVM and the proposed method. (a) the proposed DQN-based detection method. (b) SVM.

FIGURE 4. Average rewards of the neural network structures for 1% imbalanced KEPCO dataset.
V. CONVERGENCE BEHAVIOR FOR IRISH DATASET

In Section V, the performances of the proposed DQN method are verified by the Irish smart meter dataset [32] for various AR conditions. The Irish dataset consists of daily electricity consumption recordings of more than 5,000 consumers from 2009 to 2010. From the total dataset, 100 consumers are randomly selected to prepare five datasets for simulation. The daily consumption record for each user is one sample of the dataset. The consumption amounts are recorded every 30 minutes, which means that the size of the total input feature is 48. The details of each dataset are listed in Table 7. The daily consumption records of consumers are removed when null values occupy more than 3/4 of the total features. The rest of the empty data are filled with zero. The samples are randomly selected to apply FDI according to AR = 0.1. The Irish dataset has been proposed in this paper to detect energy theft without the data augmentation method. The strength of our method is that the pre-processing step to select input features is automatically done in the classification step. In addition, average rewards of 1D-CNN are converged to higher values compared to values of FNNs for 1% extreme case with KEPCO dataset.

VI. CONCLUSIONS

The DRL-based NTL detection method for the imbalanced dataset has been proposed in this paper to detect energy theft and simulate under the environment of costly features. It was demonstrated that the proposed method is robust to the NTL dataset compared to various detection methods from the simulation results. This was accomplished by the different rewards for each classification action of the positive and negative classes. For the imbalanced dataset, the proposed method can effectively detect energy theft without the data augmentation method. The strength of our method is that the pre-processing step to select input features is automatically done in the classification step. In addition, average rewards of 1D-CNN are converged to higher values compared to values of FNNs for 1% extreme case with KEPCO dataset.

Estimation bias is incurred by overestimations of DQN as mentioned in [29]. The bias is the cause of the large standard deviation values of the proposed method in Table 2. Therefore, advanced DQNs such as Double DQN will be applied to the DQN-based NTL detection method to improve the stability in future works. Our proposed method will also be applied to a similar classification problem of an imbalanced dataset for detecting malfunctions of electrical appliances or electrical devices, whose functions are provided by a home management system or building management system in SG.

REFERENCES

[1] S. McLaughlin, B. Holbert, A. Fawaz, R. Berthier and S. Zonouz, “A multi-sensor energy theft detection framework for advanced metering infrastructures,” IEEE J. Sel. Areas Commun., vol. 31, no. 7, pp. 1319-1330, July 2013.
[2] Ç. Yurtseven, “The causes of electricity theft: an econometric analysis of the case of Turkey,” Util. Policy, vol. 37, pp. 70–78, Dec. 2015.
[3] F. B. Lewis, “Costly throw-ups: Electricity theft and power disruptions,” Electr. J., vol. 28, no. 7, pp. 118 –135, Aug. -Sep. 2015.
[4] F. B. Smith, “Electricity theft: a comparative analysis,” Energy Policy, vol. 32, pp. 2067–2076, Dec. 2004.
[5] A. Aldeghishaem, M. Anwar, N. Javaid, N. Alrajeh, M. Shafiq and H. Ahmed, "Towards sustainable energy efficiency with intelligent...
electricity theft detection in smart grids emphasising enhanced neural networks," IEEE Access, vol. 9, pp. 25036-25061, Feb. 2021.

[6] J. L. Viegas, P. R. Esteves, R. Melício, V. Mendes, and S. M. Vieira, “Solutions for detection of non-technical losses in the electricity grid: A review,” Renewable Sustainable Energy Rev., vol. 80, pp. 1256-1268, Dec. 2017.

[7] G. M. Messinis and N. D. Hatziargyrrou, “Review of non-technical loss detection methods,” Electric Power Research, vol. 158, pp. 250–266, May 2018.

[8] Y. Zhou, X. Chen, A. Y. Zomaya, L. Wang and S. Hu, “A dynamic programming algorithm for leveraging probabilistic detection of energy theft in smart home,” IEEE Trans. Emerg. Topics Comput., vol. 3, no. 4, pp. 502-513, Dec. 2015.

[9] J. Y. Kim, Y. M. Hwang, Y. G. Sun, I. Sim, D. J. Kim and X. Wang, “Detection for non-technical loss by smart energy theft with intermediate monitor meter in smart grid,” IEEE Access, vol. 7, pp. 129043-129053, Sep. 2019.

[10] P. Jokar, N. Arianpoo, and V. C. Leung, “Electricity theft detection in AMI using customers’ consumption patterns,” IEEE Trans. Smart Grid, vol. 7, no. 1, pp. 216–226, Jan. 2016.

[11] G. M. Messinis, A. E. Rigas and N. D. Hatziargyrrou, “A hybrid method for non-technical loss detection in smart distribution grids,” IEEE Trans. Smart Grid, vol. 10, no. 6, pp. 6080-6091, Nov. 2019.

[12] P. Glauner, J. A. Meira, P. Valtchev, R. State, and F. Bettinger, “The challenge of non-technical loss detection using artificial intelligence: A survey,” arXiv preprint arXiv:1606.00626, July 2017.

[13] R. Punnmiya and S. Choe, “Energy theft detection using gradient boosting theft detector with feature engineering-based pre-processing,” IEEE Trans. Smart Grid, vol. 10, no. 2, pp. 2326–2329, Mar. 2019.

[14] K. Zheng, Q. Chen, Y. Wang, C. Kang and Q. Xia, “A novel combined data-driven approach for electricity theft detection,” IEEE Trans. Ind. Informat., vol. 15, no. 3, pp. 1809-1819, Mar. 2019.

[15] M. Zanetti, E. Jambour, M. Pellenz, M. Penna, V. Zambenedetti and I. Chueiri, “A tunable fraud detection system for advanced metering infrastructure using short-lived patterns,” IEEE Trans. Smart Grid, vol. 10, no. 1, pp. 830-840, Jan. 2019.

[16] R. Razavi, A. Gharipour, M. Fleury and J. J. Akpan, “A practical feature-engineering framework for electricity theft detection in smart grids,” Applied Energy, vol. 238, pp. 481-494, Mar. 2019.

[17] Z. Zheng, Y. Yang, X. Niu, H. Dai and Y. Zhou, “Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids,” IEEE Trans. Ind. Informat., vol. 14, no. 4, pp. 1606-1615, Apr. 2018.

[18] M. N. Hasan, R. N. Toma, A.-A. Nahid, M. M. M. Islam, and J.-M. Kim, “Electricity theft detection in smart grid systems: A CNN-LSTM based approach,” Energies, vol. 12, no. 17, pp. 1-18, Aug. 2019.

[19] Z. Yan and H. Wen, “Electricity theft detection base on extreme gradient boosting in AMI,” IEEE Trans. Instrum. Meas., vol. 70, pp. 1-9, Jan. 2021.

[20] J. Jaromr, T. Pevny, and V. Lisy, “Classification with costly features using deep reinforcement learning,” in Proc. AAAI Conf. A.I., Honolulu, Hawaii, USA, July 2019, vol. 33, no. 1, pp. 3959-3966.

[21] L. Lin, Q. Chen, and X. Qi, “Deep reinforcement learning for imbalanced classification,” Applied Intelligence, vol. 50, pp. 2488–2502, Mar. 2020.

[22] S. K. Gunthi and D. Sarkar, “Ensemble machine learning models for the detection of energy theft,” Electric Power Systems Research, vol. 192, Mar. 2021.

[23] B. Zhao, H. Lu, S. Chen, J. Liu and D. Wu, “Convolutional neural networks for time series classification,” J. Syst. Eng. Electron., vol. 28, no. 1, pp. 162-169, Feb. 2017.

[24] W. Tang, G. Long, L. Liu, T. Zhou, J. Jiang, and M. Blumenstein, “Rethinking 1D-CNN for time series classification: A stronger baseline,” arXiv preprint arXiv:2002.10061, Feb. 2020.

[25] Z. Wang, W. Yan and T. Oates, “Time series classification from scratch with deep neural networks: A strong baseline,” in Proc. IJCNN, Anchorage, AK, USA, July 2017, pp. 1578-1585.

[26] D. P. Kingma and J. Ba “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, Jan. 2017.

[27] Korea Electric Power Corporation [Online]. Available: https://home.kepco.co.kr/kepco/main.do

[28] Z. Aslam, N. Javaid, A. Ahmad, A. Ahmed, and S. M. Gulfam, “A combined deep learning and ensemble learning methodology to avoid electricity theft in smart grids,” Energies, vol. 13, no. 21, pp. 1-24, Oct. 2020.

[29] H. van Hasselt, A. Guez, and D. Silver, “Deep Reinforcement Learning with Double Q-Learning,” AAAI, vol. 30, no. 1, Mar. 2016.

[30] Commission for Energy Regulation (CER) Smart Metering Project-Electricity Customer Behaviour Trial, 2009-2010, Irish Social Science Data Archive. [Online]. Available: https://www.ucd.ie/issda/data/commissionforenergyregulation/

BIOGRAPHY

Jiyoung Lee (S’20) received the B.Sc. (summa cum laude) and M.Sc. degrees from the Department of Electronic Convergence Engineering, Kwangwoon University, Seoul, Korea, in 2020. She is currently pursuing the M.Sc. degree at Kwangwoon University. Her research interests include internet of energy, artificial intelligence, and wireless communication systems, etc.

Young Ghyu Sun (S’18) received the B.Sc. (summa cum laude) and M.Sc. degrees from the Department of Electronic Convergence Engineering, Kwangwoon University, Seoul, Korea, in 2018 and 2020, respectively, where he is currently pursuing the Ph.D. degree. He was a recipient of the IEEE Student Paper Gold Award in 2020. His research interests include wireless communication systems, deep learning, and internet of energy, etc.

Isaac Sim (S’16) received the B.Sc. (magna cum laude) and M.Sc. degrees from the Department of Wireless Communications Engineering, Kwangwoon University, Seoul, Korea, in 2016 and 2018, respectively. He is currently working toward the Ph.D. degree at Kwangwoon University. His research interests include wireless energy harvesting, wireless sensor networks, and artificial intelligence, etc.

Soo Hyun Kim (S’19) received the B.Sc. and M.Sc. degrees from the Department of Electronic Convergence Engineering, Kwangwoon University, Seoul, Korea, in 2019 and 2021, respectively, where he is currently pursuing the Ph.D. degree. His research interests include wireless communication systems, and machine learning applications for internet of energy, etc.
Dong In Kim (S’89–M’91–SM’02–F’19) received the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 1990. He was a Tenured Professor with the School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada. Since 2007, he has been a SKKU-Fellowship Professor with the College of Information and Communication Engineering, Sungkyunkwan University (SKKU), Suwon, South Korea.

Prof. Kim has been a first recipient of the NRF of Korea Engineering Research Center in Wireless Communications for RF Energy Harvesting since 2014. He was selected the 2019 recipient of the IEEE Communications Society Joseph LoCicero Award for Exemplary Service to Publications. From 2001 to 2019, he served as an Editor and an Editor at Large for Wireless Communications I for the IEEE Transactions on Communications. From 2002 to 2011, he also served as an Editor and a Founding Area Editor of Cross-Layer Design and Optimization for the IEEE Transactions on Wireless Communications. From 2008 to 2011, he served as the Co-Editor-in-Chief of the IEEE/KICS Journal of Communications and Networks. He served as the Founding Editor-in-Chief of the IEEE Wireless Communications Letters, from 2012 to 2015. He is a Fellow of the Korean Academy of Science and Technology and a Member of the National Academy of Engineering of Korea. He is the General Chair for IEEE ICC 2022 in Seoul.

Jin Young Kim (M’95 –SM’08) received the B.S., M.S., and Ph. D. degrees from the School of Electrical Engineering, Seoul National University (SNU), Seoul, Korea, in 1991, 1993, and 1998, respectively. He was Member of Research Staff at the Institute of New Media and Communications (INMC) and at the Inter-university Semiconductor Research Center (ISRC) of the SNU from 1994 to 1998. He was Postdoctoral Research Fellow at the Department of Electrical Engineering, Princeton University, NJ, U.S.A, from 1998 to 2000. He was currently Full Professor at the School of Electronics Engineering, Kwangwoon University, Seoul, Korea. He had his sabbatical leave as Visiting Scientist at the LIDS (Laboratory of Information and Decision Systems), Massachusetts Institute of Technology (M.I.T), MA, U.S.A from 2009 to 2010.

His research interests include artificial intelligence, design and implementation of wireline/wireless multimedia communication systems with basis on modulation/demodulation, synchronization, and detection/estimation theory. He received the Best Paper Awards from several academic conferences and societies including Jack Neubauer Best Systems Paper Award from IEEE VT Society (2001), the Award of Prime Minister of Korea Government (2011), He is now Senior Member of IEEE, Regular Member of IET and IEICE.