Non-technical loss detection in limited-data low-voltage distribution feeders

Author(s): Firoozi, Hooman; Mashhadi, Habib Rajabi

Title: Non-technical loss detection in limited-data low-voltage distribution feeders

Year: 2021

Version: Accepted version

Please cite the original version:

Firoozi, H. & Mashhadi, H. R. (2021). Non-technical loss detection in limited-data low-voltage distribution feeders. *International Journal of Electrical Power & Energy Systems*, 107523. https://doi.org/10.1016/j.ijepes.2021.107523

Non-technical Loss Detection in Limited-data Low-voltage Distribution Feeders

Hooman Firoozi[1,2], Habib Rajabi Mashhadi[1,3,*]

[1] Electrical Engineering Department, Ferdowsi University of Mashhad, Mashhad, Iran
[2] School of Technology and Innovations, University of Vaasa, 65200 Vaasa, Finland
[3] Center of Excellence on Soft Computing and Intelligent Information Processing, Ferdowsi University of Mashhad, Mashhad, Iran
* Corresponding author: h_mashhadi@um.ac.ir

Abstract

Non-technical loss (NTL) is the major source of energy loss in distribution networks. Since the direct calculation of NTL is impossible, it is mostly being obtained by calculating the difference between total loss and technical loss (TL). This is rather not possible without adequate equipment for measuring the TL. However, time-worn distribution networks with high limitations and lack of equipment could identify the regions with a high share of NTL to reduce/eradicate the sources of NTL. This strategy can increase efficiency and reduce costs compared to the one-by-one inspection of customers. This paper presents a novel NTL detection procedure based on load estimation in highly limited distribution networks. First, the low-voltage (LV) distribution network will be divided into a few smaller hypothetical sub-networks through a fuzzy c-means (FCM) clustering technique. Then, a meter placement based on the maximum likelihood criterion is deployed to find the best places for installing the existing meters in each cluster. In the next stage, a linear system of equations is utilized to extract the pattern load profiles (PLPs) related to each class of customers for each hypothetical sub-network separately. These PLPs will be used to estimate the transformers' load in all LV feeders based on their share in the total consumption of the entire network. Finally, NTL possibility per capita will be achieved for each LV feeder by defining an index that estimates the degree of NTL in those areas.

Keywords: non-technical loss detection, electricity theft, energy loss, load estimation, low-voltage feeders, clustering

NOMENCLATURE

Abbreviations
TL : Technical loss
NTL : Non-technical loss
FCM : Fuzzy c-means
PLP : Pattern load profiles
LV : Low voltage
MV : Medium voltage
LC : Loss coefficient

Indices
t : Index of time [hour]
i : Index of cluster
j, k : Index of LV feeders

I. INTRODUCTION

A. Motivation

The problem of NTL, specifically electricity theft, has been prevailing in all countries' distribution networks. Distribution companies and distribution system operators (DSOs) have put their effort into detecting and preventing the sources of NTL in electrical networks. In most developed countries, this issue has been solved completely by utilizing smart metering solutions, imposing penalties, and strict law enforcement. On the other hand, in developing countries, energy loss in distribution networks is still a disputable problem, which mostly originates from non-engineering design and lack of supervision. These shortcomings might be even worse when it comes to looking at distribution systems from an LV point of view since major lack of measurement devices exist at this level. Yet, in some developing countries, there are not any measurement devices as well as equipment at all. For this reason, distribution companies/operators have no choice but to seek a method for detecting/estimating NTL in the areas limited to the LV-level network.

B. Literature Review

A comprehensive literature review was conducted in this paper. Clearly, the amount of NTL cannot be calculated directly. However, it can be detected/estimated by different kinds of methods presented in previous research. One of these methods is to reach the value of NTL by evaluation of TL and then subtracting it of the total loss. In [1], the mean value of TL and NTL were estimated by defining an equivalent operational impedance that was obtained using a power flow solution, transformers load, and customers' bills. This method requires meters to be installed in the LV sector to measure real-time values of voltage, current, and power. Moreover, the equivalent operational impedance introduced in this reference assumes that the average values of voltage, current, and power measured in the entrance of the network are the same for all transformers. This method also requires power flow studies which could not be applicable if there is no information about, e.g. impedance of the branches and loads. Additionally, in [2], the authors introduced an anomaly framework for NTL detection based on the calculation of TL through the loss factor in which the required data received from smart meters. The method of this literature considers a smart grid that requires advanced metering infrastructures (AMI) in the customer-level, LV-level, and distribution system operator-level (DSO-level) for double-way communication. These infrastructures are not available in traditional networks in developing countries. In [3], a procedure was proposed to detect NTL based on utilizing intermediate monitor meters. In this work, a data management center uses smart metering datasets and load flow equations to find malfunctioning and bypassing meters from cheating customers. The methodology of this reference is based on the real-time monitoring of the network by using smart meters located on the distribution transformers as well as the customer-level with the capability of bi-directional communication. The metering infrastructure used in this reference is only available in smart distribution grids. Another method that was proposed in [4] is based on a load estimation procedure. In this work, data received from meters on transformers used in a state estimator to estimate network variables. Then, the suspicious customers were classified by analyzing of variance. The authors of this reference considered a smart distribution network in which every single customer is equipped with the smart energy meter. This assumption however could not be applicable in the third world and developing countries' distribution networks in which the customers have only one simple energy meter.

Mathematical techniques were also believed to be beneficial in NTL detection. For example, in [5], the authors proposed a multi-dimensional approach to detect anomaly in consumption by preprocessing data with multiple time series of demand. This reference requires real-time power measuring devices at the customer level which have the capability of preserving the historic power consumption data. This means the traditional periodic energy meters could not be useful in the proposed method. In [6], authors provided an algorithm in which illegal consumptions were detected through deploying time series and data mining of recorded datasets from smart meters.

The proposed method in this literature utilizes the AMI data from customers in order to analyze the consumption behaviors. This method could not be the best choice for NTL detection since installing the AMI meters for all customers in order to detect NTL is not only far-fetched in the developing countries but also it might not be necessary to equip all the customers with such costly devices. Another NTL study that was based on analyzing time series of customers' consumptions. Then, the suspected customers were identified by fuzzy logic sets [7]. This work deals with the fraud detection from household customers by analyzing their energy consumption time-series over more than a decade. The method does not consider the fraud originated from commercial consumers which have a great portion of NTL in distribution systems. Moreover, the study period is required to be long enough (i.e. over a decade) to find a pattern of fraud that could not be precise since the customer might change during such a long time. Some references such as [8], proposed a method to detect the NTL stemming from measuring devices malfunctioning. A state estimator with constant Jacobean matrix (AMB-SE) along with normalized residual technique was employed to detect NTL in distribution grids. This reference requires the installation of power and voltage measurement devices both in the network-side and customer-side as well as the good knowledge about network nodes and branches characteristics in order to run the state estimation program. This method is also not applicable to limited data distribution networks with a considerable lack of supervision. In work [9], the authors presented an NTL estimation by measuring several variables selected randomly. A mathematical theft model proposed by using multiple linear regression models. The proposed method in this work estimates the NTL in LV feeders in a one-by-one manner. This means for every single LV feeder it is required to have the complete information about the LV branches for different phases of the feeder which itself imposes a huge cost and time for investigation. In fact, there is no difference between the proposed method and the traditional method for the investigation of all customers. The proposed method is highly similar to the approach used in [10] in which the authors proposed a method applicable only in smart grids which requires smart metering infrastructure in the customer-level as well as the mutual communication between customers, which is difficult to achieve in run-down and non-smart networks. Another mathematical approach was used in [11] to predict the suspected users to analyze the socio-economic issues of customers in a Brazilian distribution network. The methodology of this literature presents a predictive model to deal with the NTL occurrence in the distribution network using TL calculation process. This was done by running a power flow algorithm that needs the measured data of power as well as the topological knowledge of the network. More importantly, is the social study on customers that needs to be done beforehand.

Some of the presented methods were based on artificial intelligence and learning techniques. Reference [12], for instance, used smart meters infrastructure along with support vector machine (SVM) to analyze NTL and then proposes methods to mitigate the amount of this loss in distribution networks in India. This paper deals with an approach in order to prevent NTL through installing a smart meter in the customer-side as well as AMI in the network which all have the capability of two-way communication the measured data of voltage, current, and power. The proposed method could not be implemented in non-smart distribution networks. In [13], a neural network procedure was used to detect NTL at the customers' level. In this work, smart meters recorded the required data. Then, by using a state estimator and a self-organizing trainer, electricity theft was obtained among customers. The proposed method deals with the smart meters attack detection. In this work, the real-time values of power need to be communicated through phasor measurement units (PMUs) to the distribution company. Another similar work presented in [14] detect NTL in smart grids based on convolutional neural networks by following the consumption patterns of customers. This paper also needs the daily consumption of all customers to be communicated which could not be suitable to use for customers who have simple non-smart energy meters. Another example based on learning methods was presented in [15]. This work used a semi-supervised learning method to find the suspected residential customers in a smart grid. This research also requires the AMI devices as well as the smart meter for all customers. Unsupervised learning based on a clustering technique was utilized to detect abnormal usage of customers. The required data in this study was also achieved from smart meters [16]. The proposed method requires the smart meters on the customers' side which have the ability to communicate the hourly consumption of customers to the electricity service provider. A resampling technique was proposed in [17] so as to tackle the problem of data imbalance in NTL detection procedure through a supervised classifier. This work proposes a method for NTL detection which requires the smart metering data of all commercial and industrial customers over about two years. The method not only is not applicable in traditional networks because of the lack of smart energy meters but also requires a long period of study on metered data which could result in NTL-originated monetary loss for the distribution company over such a long-time span. Also, a mathematical approach, as well as a classifier based on neural networks, was employed in [18] in order to study on consumption patterns of customers to detect NTL in distribution network with distributed generators (DGs). The authors provided a classification-based NTL detection approach in smart grids which requires the recorded data from smart meters every 15 minutes. The availability of smart meters for all LV customers with this resolution of sampling is far-fetched in traditional networks in developing countries.

There have been other methods of NTL detection based on remote sensing techniques. For example, a process based on compressed sparse sensing was presented in [19] to detect the location of electricity thefts known as NTL. The method utilizes smart sensors located on all nodes which are capable of measuring time-continues power. This method could not be useful again in the traditional networks which suffer from lack of smart meters. Another similar methodology was proposed in [20] to monitor TL by using temperature sensors and a meshed wireless network in order to make an improvement in NTL estimations. The proposed method aims at monitoring the real-time value of TL in order to emerge the NTL from the total loss. This method is useful when the target of the study is to calculate the NTL in a specific feeder. Therefore, for the whole distribution network, smart monitoring/metering devices are needed to calculate the NTL of LV feeders separately, which is not at all achievable with the lack/limitation in metering devices. In reference [21], a real-time NTL detection method based on online supervising was proposed in order to detect and prevent NTL in a smart-meter-equipped distribution network. This reference also utilized smart meters for every single customer along with a data management system owned by the distribution company. The current of branches, in general, needs to be measured and monitored in this method. Finally, a comprehensive literature review on different methods of NTL detection such as machine learning approaches, metaheuristics algorithms, load profiling methods, TL calculation methods, etc. was conducted in [22], [23] and [24].

It must be mentioned that all of the above-said methods of detection or estimation of NTL in distribution networks are required to have adequate sets of data and measurements. On the one hand, the methods of NTL estimation based on TL calculation are mostly dependent on having topological information of LV feeders. On the other hand, the approaches of NTL detection at customers' level based on mathematical and learning algorithms might be appropriate for grids with smart metering infrastructure or at least enough measurements in medium voltage (MV) and LV networks. However, this paper aims to detect and estimate NTL in networks in which there is no smart metering devices installed on either MV or LV sectors. In fact, the only measuring devices are traditional periodic energy meters of end-users as well as quite limited number of energy meters on distribution transformers. Therefore, none of the proposed methods in previous works have the capability of detecting/estimating NTL in this condition since the existing methods are not applicable to limited-data networks.

C. Main Contribution

The main contribution of this study can be summarized as a method of selecting the areas in the LV distribution network with high level of NTL in order to be inspected based on an index that estimates the degree of NTL in those areas. The studied system is an urban distribution system with residential and commercial consumers. Firstly, by clustering the LV feeders, the network is divided into some smaller grids. Due to the limited data assumption, only one measurement device can be used in each cluster. Since the center in each cluster is representative of the features of that cluster, the center of each cluster is assumed as the proper place to install measuring equipment. It is obvious that features of the centroid may not match with features of LV feeders, hence the transformer with maximum likely-hood features selected to install measuring equipment based on the Euclidian distance criterion. The purpose of installing measuring equipment on transformers is to find residential and commercial load patterns of customers located in each one of the clusters. Hence, by using output information of the installed meters, information of real-time injected power to distribution network from main substation and network information, PLPs are extracted with a linear method. In the next step, distribution transformers load estimated by using a proposed linear relation between transformers' injected hourly load and number of consumers according to the relation between the number of consumers of each class and transformers' load [25]. Energy loss of each LV feeder is calculated by subtracting the total bills of that feeder of the estimated load over a specific period. TL calculation in such these networks is quite difficult therefore, we can assume that loss has a linear relation with consumed energy and then NTL of each LV feeder can be evaluated. Finally, LV regions were ranked by the maximum possibility of NTL happening and regions located on top of this ranking are chosen for intrusive investigations.

It is worth mentioning that the current study is in continuation of the previous work done in 2016. With this regard, the idea behind the clustering of LV feeders, meter placement as well as the load estimation stages are driven by using the authors' previous work [25]. The rest of the paper, however, is the contribution to the novelty and originality of this paper.

D. Paper Organization

The rest of this paper consists of the following sections: Section II expresses the problem definition by having a brief overview of the structure of distribution networks in developing countries. Section III proposes the methodology of the paper. Section IV demonstrates the case study presents the simulation results carried out on the case study. Section V argues the possible questions and ambiguity related to the proposed method. Finally, section VI concludes the paper.

II. PROBLEM DEFINITION

Distribution networks mostly consist of two voltage levels: MV as the primary sector and LV as the secondary sector. Primary feeders have a radial topology with a three-phase line structure that starts from a sub-transmission substation (HV/MV) and ends with a distribution transformer (MV/LV). In secondary feeders (i.e. secondary side of distribution transformers), there are single-phase conductors and laterals supplying the end-users. The only measurement devices that can be seen in most of these distribution grids is one meter located at the main substation as well as energy metering devices in end-user's premises.

According to the statistics, the electrical energy loss in countries like Iran is around 11% which is by far higher than the desired values compared to developed countries [26]. A major part of energy loss in power systems occurs in distribution networks, especially in LV sector, due to the enormous number of time-worn equipment and non-engineering topological configuration. Another major source of energy loss in distribution feeders is NTL which stems from many reasons such as electricity theft, metering devices inaccuracy, billing errors, etc. The main part of NTL is electricity theft especially in the urban areas, suburbs, and rural zones. Hereupon, distribution companies are being enforced to reduce their network losses. Therefore, they have recently started to make decisions about energy loss reduction.

The first step for this purpose is calculating the value of energy loss. The main method that electricity distribution networks utilize to calculate the energy loss is presented in (1):

$$\mathcal{L}oss = \mathcal{D}E - \mathcal{S}E \tag{1}$$

Where $\mathcal{L}oss$ is representing the total energy loss. The terms $\mathcal{D}E$ and $\mathcal{S}E$ are determined by main substation's (MV/LV) metering devices and all customers' billing information, respectively. This method yields the total energy loss of the entire distribution network (including MV and LV levels). Although it does not need any specific network information, it does not give a specific sight of the energy loss of each LV feeder. Thus, by using this method, distribution companies cannot be able to study LV feeders in detail for loss reduction.

Methods explained in the previous section such as the loss factor method, the sum of energies, etc are also employed to estimate the distribution network's energy loss. However, in Iranian distribution networks as an example, the results are highly inaccurate for the sake of high range of NTL, error in energy metering devices, non-engineering network planning, etc. On the other side, the lack of enough metering devices and equipment at the distribution level in developing countries that have economical limitations has made the problem even worse. Hence, this paper is about to work on such networks.

The case study of this paper is similar to the networks used in developing countries such as Iran. The understudy distribution network consists of the MV sector as well as the LV sector with radial topology. Each LV feeder is being fed with a distribution transformer along with several LV customers. In order to record the real-time values of the network's parameters, a large number of measuring devices are needed to be installed on every single node. Installing meters in all areas of this network is farfetched due to economic issues and a lack of equipment. However, it is assumed that a limited number of meters might be accessible if needed.

III. PROPOSED METHODOLOGY

As mentioned in the previous section, one of the main problems related to the distribution network is the lack of data due to the absence of enough measurement devices, especially in the LV sector. In order to tackle this circumscription, a novel NTL detection method base on load estimation is presented here. This method is based upon a minimum information condition in a distribution network in which the only existing information is the billing data of consumers in a 30-day billing period. The proposed method in this paper will utilize the billing data along with output information of a very limited number of real-time measurements located at the secondary side of distribution transformers.

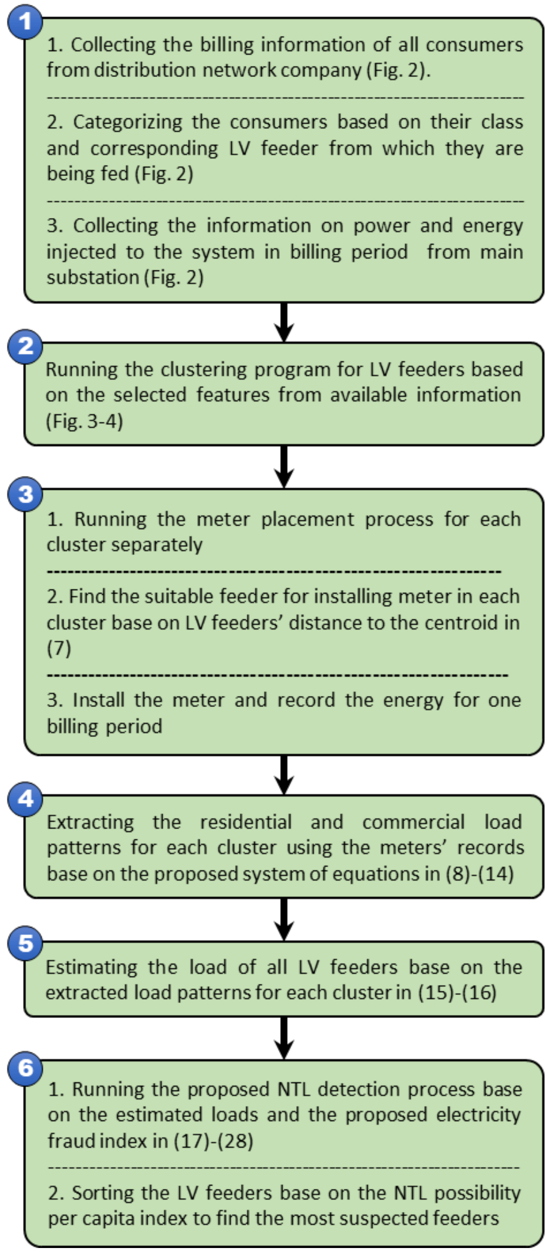

Fig. 1. Overview of the proposed NTL detection procedure

The procedure of this method tries to estimate the load of distribution transformers that are injected into LV feeders. By estimating the power and periodic energy that is injected to each one of LV feeders and then subtracting the sum of consumed bills from these estimated values, the value of energy loss for LV feeders will be estimated. The whole procedure of the proposed method can be found in Fig. 1. The above-mentioned NTL detection steps are explained in detail in the following subsections.

A. Primary Information Preparation

In this method, the required information consists of information about consumers, loads that has direct relevance to the value of energy loss in each LV feeder, and total distribution energy loss which are categorized as follows:

- Consumption billing information (for all LV customers).
- Number of consumers of each class (in each LV feeder).
- Injected hourly power to the whole network from the main substation.
- Injected energy to the whole network from the main substation in the billing period.

The values of the above information related to the LV feeders can be found in Table I. The summary of the very first step of the study is depicted in Fig. 2 in which the source of needed information and preprocessing stage are shown.

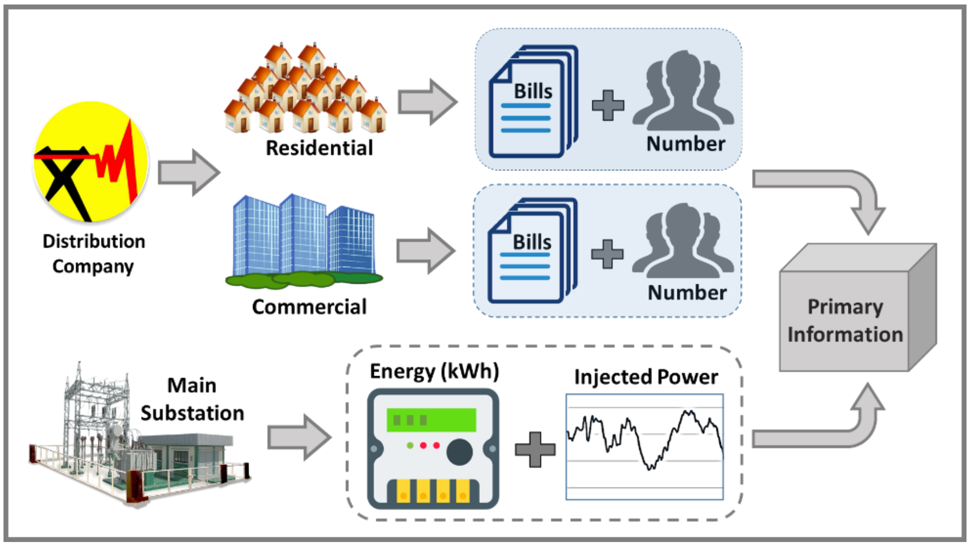

Fig. 2. Primary Information Collection and Preparation Stage

The number of consumers of each class (residential and commercial) should be determined through an experimental investigation in each LV feeder to specify that each one of the consumers is connected to which transformer and what its tariff is. This information is already available and could be collected from the distribution company. Injected power and energy from the main substation are also recorded by the measurement system located at the entrance of the MV-level network. Finally, consumers' billing information is already prepared in the distribution network company database. After collecting and splitting this information, they should be used in the clustering program which is described in the next step.

B. Clustering Low-voltage Feeders

The only way to attain accurate values of all LV feeders' load in the distribution network is by installing measurement devices inside each one of the feeders. In this paper, as mentioned before, the number of measurement devices are limited due to minimum information assumption, and it is necessary to utilize the available measurements optimally. For this purpose, this paper employed a clustering technique. The clustering technique is a computational method that can group several objects into independent clusters through an optimization procedure. The criterion that the clustering technique utilizes to separate objects is the homogeneity of the objects' features.

By using this technique, homogenous LV feeders could be divided into separated groups. In each group, some LV feeders have similar features such as load patterns, the number of consumers of each class, etc. In other words, this stage aims to divide the whole network into smaller hypothetical sub-networks so as to find the LV feeders which are similar in consuming behavior. Accordingly, instead of using one meter for each feeder, this procedure paves the way to use one meter for each group of similar feeders. The obtained hypothetical sub-networks from the clustering stage are obtained from a mathematical procedure which means they are not being clustered physically. This is due to the nature of clustering algorithms which are meant to analyze their objects in a mathematical space.

Regarding the high sensitivity of clustering results to data features, selecting adequate and appropriate features of LV feeders would truly affect the accuracy of estimations. It should be noted that the limited data in this paper refers to the lack of measured data from metering devices at the LV-level network. These data then will be recorded by installing a limited number of meters on the network. However, the information that is utilized as the inputs of the clustering program, which can affect the results of the clustering stage, are considered amongst the available primary information obtained from the first stage of the study (see Section III.A).

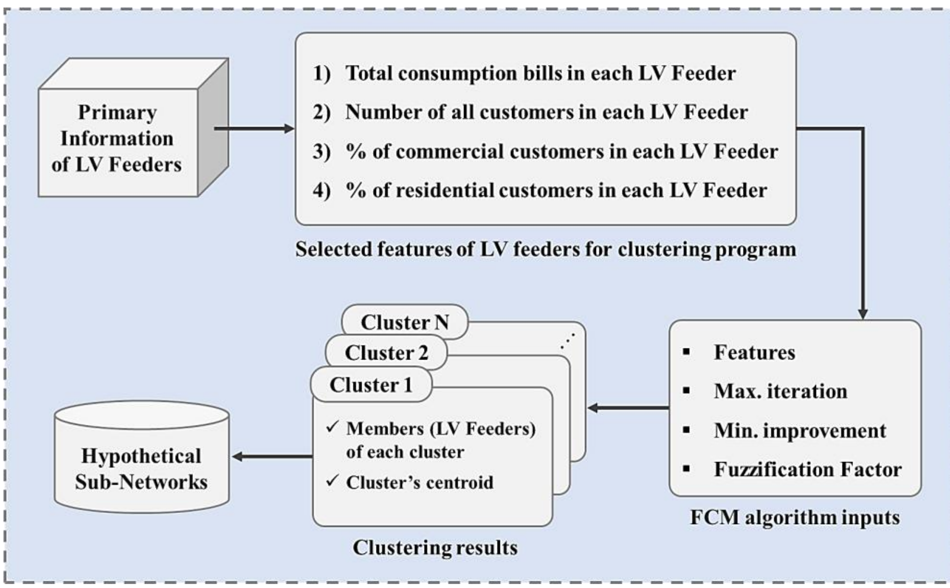

Fig. 3. Infographic of Clustering Stage

The choice of features for every clustering program is essential and directly affects the results of that clustering. However, it is mostly dependent on how and on what purpose the objects are supposed to be clustered. In this paper, the choice of clustering features should be in a way that the features of clustering could effectively describe the similarity of LV feeders in terms of consumption patterns. This could be crucial for estimating the load of LV feeders in which no meter is installed on their entrance.

In this paper, the available features of LV feeders are the number of customers of each class and the billing information of all customers. The features of each LV feeder that could be useful here are selected as in Fig. 3:

- Percent of residential consumers
- Percent of commercial consumers
- Number of all customers
- Total consumed energy (based on bills)

Where percent of residential/commercial consumers of each LV feeder can be obtained by dividing the number of residential/commercial consumers of that feeder by its total number of customers. The procedure of this step is depicted in Fig. 3. The aforesaid features are used as input of the clustering computer program. The method that is used here is fuzzy c-means (FCM) clustering in which the required parameters for the clustering algorithm are tuned as conventional values [27]. This algorithm is employed for clustering because of its suitable performance in the case of LV feeders which their customers' behavior are closely similar together. Therefore, in such cases, other types of clustering algorithms (such as K-means) in which the clusters are being created crispy could not yield the proper results. This means, in FCM, instead of considering a member to only one cluster, members are assumed to be also in other clusters, but with different degrees of membership. Finally, the members of each cluster are achieved by finding the clusters in which members have the highest degree of membership.

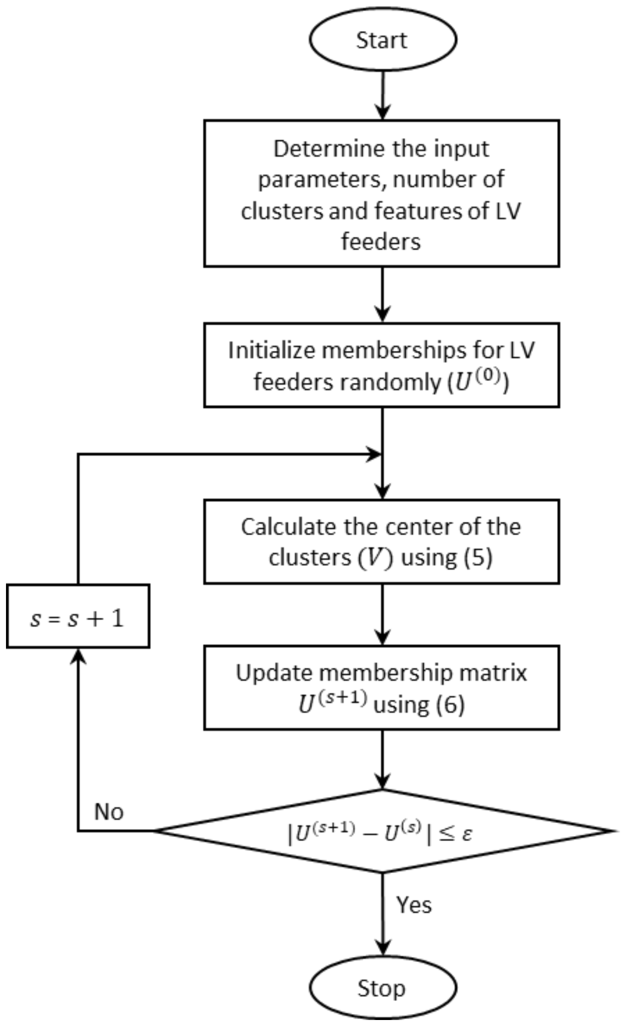

Fig. 4. The flowchart of FCM algorithm

In this algorithm, the first step of the program is to determine the input values. The algorithm will be initialized by making a random assignment for objects (i.e. LV feeders) to which cluster they belong. Accordingly, the optimization with the objective function (2) will be run iteratively in order to minimize the intra-cluster variance. It then calculates the centroid of the clusters as well as the objects' degree of membership to each cluster. This process will continue until the maximum number of iterations is reached or the program is converged. The convergence criterion will be met when the difference between the memberships of two consecutive iterations is nil. The objective function along with the related constraints of the FCM algorithm can be found in (2)-(5):

$$min. \quad J(U,V) = \sum_{j=1}^{F}\sum_{i=1}^{C} u_{ij}^{m} (x_j - v_i)^2 \tag{2}$$

Subject to:

$$0 < u_{ij} < 1 \tag{3}$$

$$\sum_{i=1}^{C} u_{ij} = 1 \tag{4}$$

$$0 < \sum_{j=1}^{F} u_{ij} < F \tag{5}$$

In (2)-(5), F, C and m are the number of features, number of clusters and fuzzification factor, respectively. In these equations, u_{ij} is the membership degree of object j to cluster i, and x_j represents the corresponding data point of the object j (i.e. LV feeder j) within the clustering mathematical space.

Inequality (3) indicates that the membership degree of object j to cluster i should be between 0 and 1. Constraint (4) ensures that the sum of membership degrees of object j to all C clusters must be equal to 1. For instance, in case there are three clusters and the membership of one object to clusters 1 and 2 is 0.7 and 0.2, respectively, then the membership of this object to the third cluster is equal to 0.1. Constraint (5) indicates that the sum of memberships for all features is lower than the number of features. In each iteration of the FCM algorithm, the centroid of cluster i as well as the membership degree of feature j to cluster i are calculated according to (6) and (7), respectively. Note that, $V = [v_1, v_2, \dots, v_C]$ and $U_{C\times F}$ represents the vector of clusters' centroid and membership matrix, respectively. The procedure of the clustering algorithm can be found in Fig. 4.

$$v_i = \frac{\sum_{j=1}^{C} u_{ij}^{m} x_j}{\sum_{j=1}^{C} u_{ij}^{m}} \tag{6}$$

$$u_{ij} = \left(\sum_{k=1}^{C} \left(\frac{x_j - v_i}{x_j - v_k} \right)^{\frac{2}{m-1}} \right)^{-1} \tag{7}$$

Running the clustering computer program showed that this distribution grid should be divided into smaller parts according to its customers' and feeders' characteristics as well as their features. This means we can install the measurement devices in the entrance of LV feeders on the secondary side of their LV distribution transformers. However, the problem of finding the most suitable feeders for installing meters is ambiguous. Hence, a meter placement procedure is proposed in the next stage to overcome this issue.

C. Meter Placement

In this study, the purpose of a meter is to measure the total load of its corresponding LV feeder during a temporal period, which is supposed to be utilized in the next stages. Therefore, in the LV sector, the best place to measure the total load of an LV feeder (including consumption, TL, and NTL) is the secondary side of the transformer located at the entrance of the LV feeder.

According to the limited number of measurement devices, installing meters on all LV feeders is unrealizable. Therefore, these available meters should be placed on specific LV feeders to provide adequate information about other ones. For this reason, it is assumed that in each cluster there exists at least one measurement device to be installed. It is obvious that if selecting feeders for meter placement done randomly, results would not be reasonable. In this case, the only way to find the proper feeders to place the meters is by using an eligibility criterion in each cluster. There is a fundamental definition in clustering topics, named "centroid", which is recognized as cluster's objects representative. Cluster centroid is described as a mathematical median of the cluster's dataset and indicates an average of objects' features.

It could be realized that when there is only one point to choose in a cluster that shows the features of that cluster, the cluster's centroid would be the best choice. Hence, the meters would be installed on the centroids of clusters. The cluster’s centroid is a mathematical concept, feeders' features and clusters’ centroids would not exactly be the same values in a way that overlaps each other. To solve this problem, a criterion must be used to show the level of likelihood of feeders in each cluster to its centroid.

By drawing features of feeders in features space, it seems that geometric distance is a proper criterion to present the likelihood of objects of a cluster. Many geometric distances were explained before. This paper would like to use "Euclidian distance" as a conventional criterion [28]. As it can be seen in (8), the Euclidian distance between p and q in an n-d space of features would be:

$$d(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \tag{8}$$

Where n is the number of dimensions of feature space. Thus, the approach of this procedure is to calculate the minimum Euclidian distance of all feeders located in a cluster to its centroid. The minimum value of calculated distances in each cluster demonstrates the nearest feeder to the cluster’s centroid which indicates the maximum likelihood of the selected feeder to the centroid. It should be mentioned that the clustering stage is the requirement of the proposed meter placement approach, and both target at having acceptable results in load estimation stage.

The chosen LV feeder in each cluster would be the representative of installing a measurement device. Note that, measurement devices are supposed to be installed on the secondary side of the distribution transformer of the representative feeder of each cluster. These measurement devices in each cluster would record the injected hourly active power from the MV/LV transformer connected to the candidate feeder. The sampling period is based on the billing period. The output data of these meters are about to use in the next step.

The measured data in the billing period might have some outliers or missing values at some data points due to unpredicted reasons. These data points as well as the corresponding data points in the other feeders have been considered as outlier and excluded from the main dataset. For this purpose, this paper deployed a simple method in which the outlier data points were obtained by comparing the value of each data point to the mean value of its corresponding vector. This means that each data points that is quite far from the average is more likely to be an outlier and can be excluded from the dataset.

D. Load Patterns Extraction

A method for minimal meter placement in selected LV feeders was presented in the previous section. After installing measurement devices on representative feeders, it is time to extract PLPs for each cluster of LV feeders. PLP in one cluster is a profile that represents the pattern of load consumption in that cluster which is different for each consumption class (i.e. residential, commercial). Indeed, residential/commercial consumers have their own load pattern profiles which are aggregated in the total transformer's load profile of representative feeders. This stage helps to extract/disaggregate the PLPs of residential and commercial consumption patterns from the total load profile measured by the installed meters.

It is accepted that having the knowledge of energy consumption patterns in distribution networks operation and control studies is essential for the system operator. At the distribution level -especially in urban areas which mostly consist of residential and commercial customers- consumption patterns are highly correlated. Therefore, this paper decides to extract load pattern profiles of consumers in each cluster to estimate the delivered power and energy in a specific period. Then, by using the estimated load profiles, the energy loss of every single LV feeder can be obtained separately. In this regard, in each cluster, the load pattern will be obtained according to data that is collected from meters, billing data, and measurement system located on the main bus. Clustering LV feeders, which was explained in previous sections, is used to divide the whole distribution network into four hypothetical sub-networks in which their features are different from others but similar to each other. Hence, by using this approach, the results of load pattern extraction for each cluster would be more accurate.

In this stage, a linear method is proposed to extract load patterns for each consumption class by using available data. According to the case study, the input dataset of the pattern extraction program for each cluster of LV feeders are as follows:

- Number of residential consumers
- Number of commercial consumers
- Metered power of representative LV feeder
- Metered power and energy of the main substation
- Billing data of all customers

The procedure of this stage can be seen in Fig. 5. In this figure, the information about clusters as well as the recorded data from the installed meters on their representative LV feeder from the previous stage are in hand. This information is supposed to be utilized in load pattern extraction procedure. The depicted information in Fig. 5 will be deployed as the parameters for the hourly equations system, which are denoted by $A_{2\times2}$ and $B_{2\times1}$. The matrix A consist of the number of residential and commercial customers in the representative LV feeder as well as the total number of residential and commercial customers of the cluster. Moreover, B consist of the recorded hourly power from the installed meter on representative LV feeder as well as the weighted value of hourly injected power by the main substation. The variables of these equations system are residential and commercial load patterns of customers of corresponding cluster denoted by $X_{2\times1}$. Accordingly, the mathematical expressions in Fig. 5 are as follows:

$$A_{2\times2} = \begin{bmatrix} \delta_j^i & \varphi_j^i \\ \Delta^i & \Phi^i \end{bmatrix} \qquad \forall i, \forall j \qquad (9)$$

$$B_{2\times1}^{(t)} = \begin{bmatrix} P^{i,t} \\ LAC^i \times P^{m,t} \end{bmatrix} \qquad \forall i, \forall t \qquad (10)$$

$$X_{2\times1}^{(t)} = \begin{bmatrix} P_R^{i,t} \\ P_C^{i,t} \end{bmatrix} \qquad \forall i, \forall t \qquad (11)$$

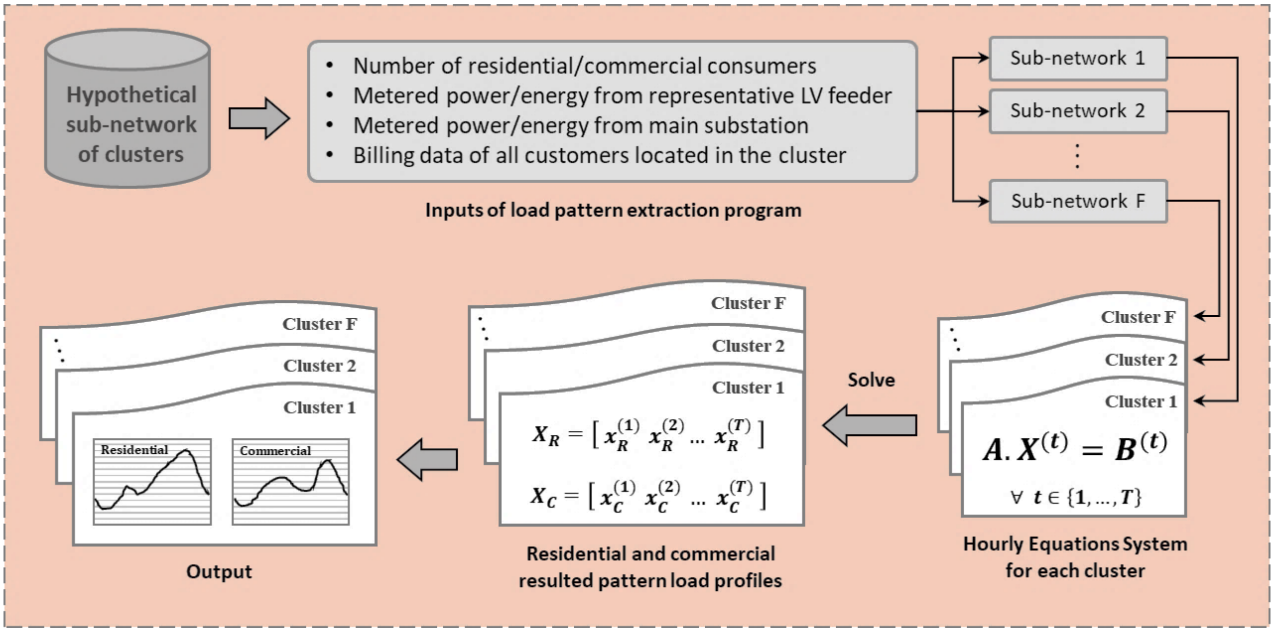

Fig. 5. Load Patterns Extraction Stage

In order to extract load patterns, the above-said data should be collected and used in the following system of (12):

$$\begin{cases} \delta_j^i \times \boldsymbol{P_R^{i,t}} + \varphi_j^i \times \boldsymbol{P_C^{i,t}} = P^{i,t} \\ \Delta^i \times \boldsymbol{P_R^{i,t}} + \Phi^i \times \boldsymbol{P_C^{i,t}} = LAC^i \times P^{m,t} \end{cases} \quad \forall i, \forall t \tag{12}$$

Where,
δ_j^i , φ_j^i : Number of residential and commercial customers of feeder j in cluster i, respectively
Δ^i , Φ^i : Number of total residential and commercial customers of cluster i, respectively
$P_R^{i,t}$, $P_C^{i,t}$: Residential and commercial load patterns of cluster i, respectively [kW]
$P^{i,t}$: Hourly measured power by installed meter in cluster i [kW]
$P^{m,t}$: Hourly injected power by main substation to whole radial network [kW]
LAC^i : Load allocation coefficient for cluster i

Eq. (12) mainly states that the total consumption of an LV feeder could be defined as the composition of load consumed by all types of customers in that feeder. The first equation in (12) demonstrates that the measured injected power by LV transformer to its secondary-side connected feeder is linearly equal to the summation of the total consumption of all residential and commercial customers. It must be said that the impact of loss in feeders is considered integrated into the consumed load. In the second line of the (12), it is assumed that the summation of total consumed load by residential and commercial customers in each cluster linearly follows the pattern of injected power by the main bus. Eq. (13) indicates the total energy consumption in cluster i is equal to the summation of billings of all feeders in cluster i, and (14) states that the total injected energy by the main substation is equal to the summation of all recorded samples of related injected power in a sampling period. Since each class of customers (residential, commercial, etc.) has its own consumption pattern, this paper seeks a method to extract PLPs for each class of customers separately. The value of LAC for each cluster, which is defined in (15), is considered as a proportion of total energy consumption of clusters out of total injected energy by the main bus to the whole radial network.

$$EC_i = \sum_{j=1}^{N_i} B_j \qquad \forall\, i \tag{13}$$

$$E_m = \sum_t P_m^t \tag{14}$$

$$LAC_i = \frac{EC_i}{E_m} \qquad \forall\, i \tag{15}$$

Where,

EC_i : Total energy consumption in cluster i [kWh]

B_j : Billing data of customers in a period in feeder j [kWh]

N_i : Number of LV feeders in cluster i

E_m , P_m : Injected energy and power by main substation, respectively [kWh] [kW]

LAC^i : Load allocation coefficient for cluster i

By solving the (12), the approximated hourly pattern loads for residential and commercial customers of each cluster could be obtained separately. Finally, the PLPs will be achieved as a result of integrating these hourly values for 30 days.

E. Load Estimation

This part of the paper presents a linear load estimation procedure over all LV feeders by utilizing the obtained PLPs from the previous section. Based on this method, the injected power by LV transformers to their connected feeders is equal to the sum of total consumption caused by residential customers and total consumption caused by commercial customers. The energy loss is considered aggregated in customer's load.

$$\delta_k^i \times P_R^{i,t} \; + \; \varphi_k^i \times P_C^{i,t} \; = \; P_{k,est}^{i,t} \qquad \forall\, i, k, t \tag{16a}$$

$$P_{k,est}^i = \begin{bmatrix} \delta_k^i & \varphi_k^i \end{bmatrix} \times \begin{bmatrix} P_{R,i}^1 & P_{R,i}^2 & \dots & P_{R,i}^{720} \\ P_{C,i}^1 & P_{C,i}^2 & \dots & P_{C,i}^{720} \end{bmatrix} = \begin{bmatrix} P_{k,est}^1 & P_{k,est}^2 & \dots & P_{k,est}^{720} \end{bmatrix} \tag{16b}$$

$$E_{k,est}^i = \textstyle\sum_t P_{k,est}^{i,t} \qquad \forall\, i, k \tag{17}$$

Where,

δ_k^i , φ_k^i : Number of residential and commercial customers of feeder k in cluster i, respectively

$P_R^{i,t}$, $P_C^{i,t}$: Residential and commercial load patterns of cluster i, respectively [kW]

$P_{k,est}^{i,t}$: Estimated hourly power injected to feeder k in cluster i [kW]

$E_{k,est}^i$: Estimated periodic energy injected to feeder k in cluster i [kWh]

Eq. (16) is defined to be utilized for load estimation in LV feeders. Similar to definition proposed for PLPs extraction in previous section, by using (16), hourly power injected by LV transformers to their connected feeders can be achieved. It is assumed that in this equation, the total load of an unmonitored LV feeder (which does not equip with any meter) could be estimated based on the summation of the total consumption of all customers (of different types) by using the obtained related load patterns (achieved from (12)). Eq. (17) refers to periodic energy which is injected to LV feeders by their connected LV transformers at the entrance of the feeders. The results of this load estimation procedure are crucial in the accuracy of the following loss-related parts. Therefore, it is necessary to compare the results so far not only with actual values of LV transformers load but also with other estimation methods used in previous works. A comparison between estimated values of transformers' load, actual transformers' load, and load allocation method provided in [30] has been done in order to assess the accuracy of presented load estimation method.

F. Non-technical Loss Evaluation

NTL calculation is impossible without having TL and total energy loss in the network. The only way to calculate the NTL in distribution grids is by deducting the TL of total energy loss. However, we must consider the fact that the amount of NTL in every network is closely intertwined with the amount of TL since any leakage of energy or theft can cause extra current flow in network lines, which means extra technical loss. Besides, due to the topological configuration of the network in this study and lack of meters, calculation of TL is quite difficult.

Clearly, the amount of TL caused by legal current flow in a network is a part of total energy loss in that network (LV feeder) which can be approximated by a linear constant coefficient. The valid range of this coefficient could be investigated previously by running a numerical analyze on the historical data of injected energy and total consumption bills of customers so as to have the idea on the relation between energy consumption and energy loss (please see APPENDIX). Let us name this coefficient loss coefficient (LC) denoted by α. The LC parameter is clearly between 0 and 1 since the amount of energy loss in a network/feeder could not be greater or equal to energy consumption and could not be zero. Due to the mentioned fact, this coefficient could be defined as in (18):

$$E_{TL}^{legal} = \alpha \times E_c \qquad (0 < \alpha < 1) \qquad (18)$$

Where,
E_{TL}^{legal}: Technical energy loss caused by legal consumption [kWh]
α : Energy loss coefficient [%]
E_c : Energy consumption [kWh]

Eq. (18) demonstrates technical energy loss caused by legal energy consumption. Similarly, there is another hidden proportion of total TL which stems from NTL current flow in grid's branches such as illegal energy consumption. This part of TL can simply be divided by defining the following (19):

$$E_{TL}^{theft} = \alpha \times E_{NTL} \qquad (19)$$

Where,
E_{TL}^{theft}: Technical energy loss caused by illegal consumption [kWh]
α : Energy loss coefficient [%]
E_{NTL}: Non-technical energy loss [kWh]

Although the value of α could not be the same for all types of loss (i.e. TL, NTL, or total loss), it could be helpful later in the procedure to find the most suspected feeders in terms of electricity theft. We adopted the presented TL decomposition idea in (18)-(19) in order to consider the effect of non-technical loss in the calculations.

By using (18) and (19), total technical energy loss which is defined as (20a) can be rewritten as in (20b):

$$E_{TL} = E_{TL}^{legal} + E_{TL}^{theft} \qquad (20a)$$

$$E_{TL} = \alpha \times (E_c + E_{NTL}) \quad (20b)$$

Where,

E_{TL}: Total technical energy loss [kWh]

E_{TL}^{legal}: Technical energy loss caused by legal consumption [kWh]

E_{TL}^{theft}: Technical energy loss caused by illegal consumption [kWh]

α : Energy loss coefficient [%]

E_c : Energy consumption [kWh]

E_{NTL}: Non-technical energy loss [kWh]

Clearly, the total energy loss in any feeder can be considered either as the difference between injected energy and consumption in that feeder ((21a)), or the summation of total technical and non-technical loss in that feeder (as (21b)). Therefore, by pasting (20b) and (21a) into (21b), the non-technical energy loss can be obtained as in (22):

$$E_{Loss} = E_{inj} - E_c \quad (21a)$$

$$E_{Loss} = E_{TL} + E_{NTL} \quad (21b)$$

$$E_{NTL} = \frac{E_{inj}}{\alpha+1} - E_c \quad \geq 0 \quad (22)$$

Where,

E_{Loss} : Total Energy loss [kWh]

E_{inj} : Delivered energy into a feeder [kWh]

E_c : Energy consumption [kWh]

E_{TL}: Total technical energy loss [kWh]

E_{NTL}: Non-technical energy loss [kWh]

α : Energy loss coefficient [%]

According to (22), the amount of non-technical energy loss can be roughly obtained without having the TL information, and the problem of data limitation for calculating TL can be covered. As mentioned before, α has a value between 0 and 1, and is related to the total percentage of loss in the network. It is impossible to consider a specific value for α since it is different for each LV feeder. Nevertheless, it could become more limited by finding the overlap between the valid values of α in (18) and (22). Thus, the new values of α for LV feeders could be written as in (23):

$$0 < \alpha \leq \frac{E_{inj}}{E_c} - 1 \quad (23)$$

By inserting the information of injected energy (which will be estimated later) and energy consumption of customers in (23) for every single LV feeder, 33 different intervals would be obtained for α. Then, by calculating the intersection of all sets, a new interval for α has been found which is $0 < \alpha \leq 0.02$. This final interval is considered equal for all feeders due to their correlation.

Since in the MV level there is not a considerable amount of non-technical energy loss like electricity theft, the total energy loss stemming from NTL in the whole network can be considered equal to the summation of NTLs in LV feeders. Therefore, for the main feeder (24) can be rewritten as:

$$E_{NTL}^{m} = E_{NTL}^{LV} = \sum_j E_{NTL}^{j} = \frac{1}{\alpha+1}\sum_j E_{inj}^{j} - \sum_j E_c^{j} \qquad (24)$$

Where,

E_{inj}^{j}: Injected energy into feeder j [kWh]

E_c^{j}: Total billings of feeder j [kWh]

E_{NTL}^{LV}: Total non-technical energy loss of LV sector [kWh]

E_{NTL}^{m}: Total non-technical energy loss of whole network [kWh]

Eq. (24) presents the total amount of NTL in the network. Moreover, we know the fact that the total injected energy by LV feeders can be calculated by deducting customers' billings from injected energy by MV substation (25):

$$\sum_j E_{inj}^{j} = E_{inj}^{m} - E_{TL}^{MV} = E_{inj}^{m} - \alpha E_c^{MV} = E_{inj}^{m} - \alpha \sum_j E_{inj}^{j} \qquad (25)$$

Where,

E_{inj}^{m} : Injected energy by main substation to whole network [kWh]

E_{TL}^{MV} : Technical energy loss in MV sector [kWh]

E_c^{MV}: Energy consumption in MV sector [kWh]

Then, (25) can be rewritten as (26):

$$\sum_j E_{inj}^{j} = \frac{E_{inj}^{m}}{\alpha+1} \qquad (26)$$

By inserting the above-mentioned (26) in (24), the total amount of non-technical energy loss can be estimated as (27):

$$E_{NTL}^{m} = \frac{E_{inj}^{m}}{(\alpha+1)^2} - \sum_j E_c^{j} \qquad (27)$$

Now, the share of non-technical energy loss in each LV feeder can be simply obtained as in (28):

$$\varphi_{NTL}^{j} = \frac{E_{NTL}^{j}}{E_{NTL}^{m}} = \frac{\frac{E_{inj}^{j}}{(\alpha+1)} - E_c^{j}}{\frac{E_{inj}^{m}}{(\alpha+1)^2} - \sum_j E_c^{j}} \qquad (28)$$

Eq. (28) indicates that how the NTL in the network is distributed across the LV feeders. By having the share of NTL for each feeder and dividing it to the number of its consumers, a new index can be obtained according to (29):

$$\mathrm{R}_j = \frac{\varphi_{NTL}^{j}}{NC_j} \qquad (29)$$

Where,

φ_{NTL}^{j} : Non-technical loss shares for feeder j [%]

R_j : Non-technical loss possibility per capita for feeder j

NC_j : Number of consumers in feeder j

The value of the defined index in (29) demonstrates that which feeders have the highest rate of non-technical energy loss probability based on the behavior of their consumers. In other words, the aforesaid value gives the fact that how non-technical energy loss is distributed in the questioned grid.

IV. NUMERICAL STUDIES

A. *Case Study*

In order to assess the performance of the proposed method in the previous section, an urban 33-bus distribution system is chosen. This network is a typical radial network operating at 20 kV level. There is a 132 kV substation located at the entrance of the network as the main substation. At each bus, a distribution transformer (20/0.4 kV) is connected to a radial LV feeder supplying a specific number of single-phase consumers. Two classes of consumers exist in this network. This case study contains 33 LV feeders providing service to 3495 single-phase consumers which 3097 of them are residential and the rest of them are commercial consumers. The total number of residential and commercial consumers connected to each LV feeder in this network are shown in Fig. 6.

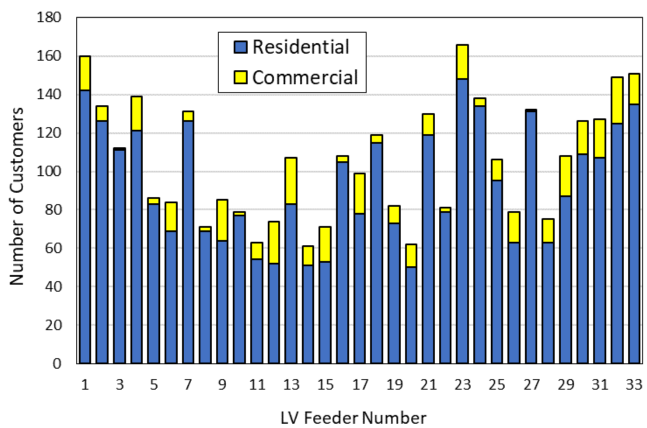

Fig. 6. The number of customers in the case study

Regarding the capacity of distribution transformers and type of the regions, each LV feeder is connected to 50~150 residential customers and up to 25 commercial customers. Although there exists an energy measurement system at the secondary side of the main substation, there are not any meters on the distribution transformers and LV feeders. In other words, the LV sectors in this network are totally unmonitored.

B. *Simulation Results*

The information about the case study of this paper can be found in Table I. As can be seen in this table, there are 33 LV feeders in this network in which a number of residential and commercial customers are connected to LV lines and are being fed by related LV (20/0.4 kV) transformers. The assumptions which are considered to simplify the simulation procedure are as follows:

1. Three-phase structure are considered symmetrical in MV sector.
2. Measuring devices' accuracy is highly acceptable.
3. Power factor is considered constant in metering period.
4. Consuming behaviors of customers in each class are highly correlated [29].

Note that, for simplicity, the three-phase structure is considered symmetrical for the MV sector in the simulation, which is not related to the LV single-phase loads. Single-phase LV loads in this study are generated randomly and they are not balanced over three phases.

Table I. LV feeders' information

Bus No.	***Residential (%)***	***Commercial (%)***	***Total Customers***	***Bills (kWh)***
1	88	12	160	39,197
2	94	6	134	30,332
3	99	1	112	23,360
4	87	13	139	34,885
5	96	4	86	18,720
6	82	18	84	22,534
7	96	4	131	28,639
8	97	3	71	15,288
9	75	25	85	24,862
10	97	3	79	16,944
11	85	15	63	16,124
12	70	30	74	22,946
13	77	23	107	30,428
14	83	17	61	16,052
15	74	26	71	20,925
16	97	3	108	23,218
17	78	22	99	27,778
18	96	4	119	25,846
19	89	11	82	20,017
20	80	20	62	16,956
21	91	9	130	30,577
22	97	3	81	17,328
23	89	11	166	40,461
24	97	3	138	29,752
25	89	11	106	25,632
26	79	21	79	21,869
27	99	1	132	27,482
28	84	16	75	19,619
29	80	20	108	29,586
30	86	14	126	31,885
31	84	16	127	33,134
32	83	17	149	39,079
33	89	11	151	36,654

In order to have a more sensible understanding of the clustering stage of the study, the positions of LV feeders in features' space are depicted in Fig. 7. As aforesaid, in this study, the authors considered four features of LV feeders to be utilized in clustering. Since drawing features' space in 3-d might not make any sense, it is shown in 2-d.

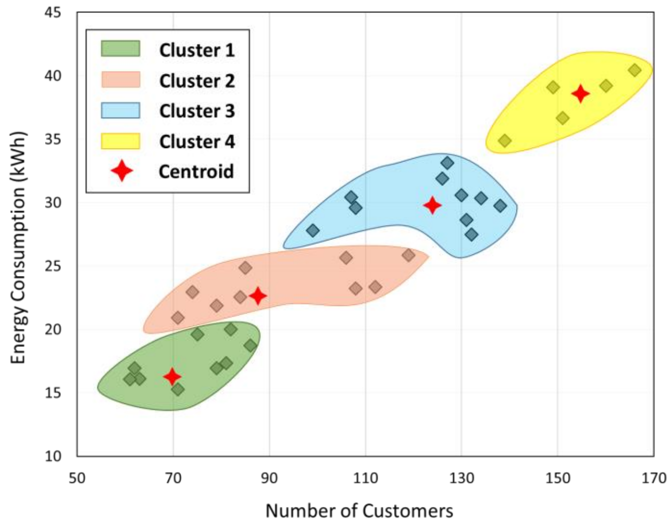

Fig. 7. The LV feeders' feature space

After collecting the primary dataset and the existed information about the network, the FCM technique is applied to the case study. With this regard, the fuzzification factor, number of iterations, and minimum improvement in the objective function are chosen 2, 100, and 10^{-5}, respectively. This computer program has been converged and finished after 10 and 50 iterations, respectively. The results of the clustering are depicted in Fig. 8.

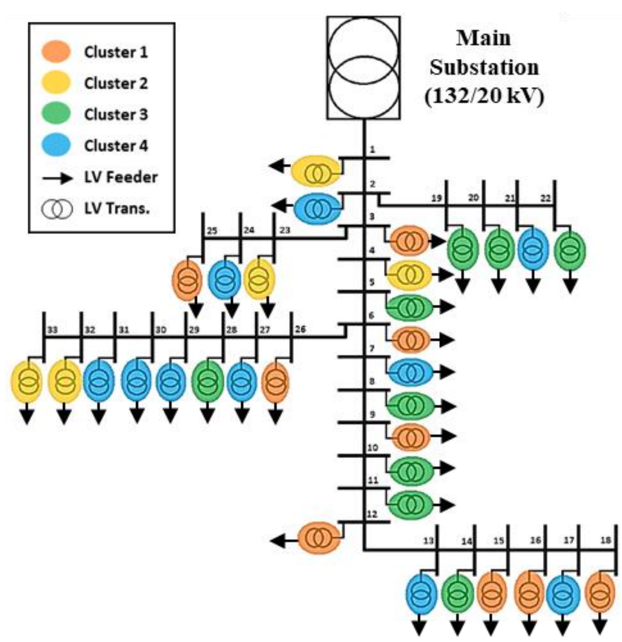

Fig. 8. Fuzzy clustering results (33-bus distribution system)

The results of meter placement are illustrated in Table II in which selected candidate feeders, as well as their features, are shown. The candidate feeders are supposed to be as follows:

Table II. Candidate LV feeders

Cluster No.	*Candidate Feeder*	*Number of Residentials*	*Number of Commercials*	*Consumption (kWh)*
1	16	105	3	23,218
2	32	125	24	39,079
3	20	50	12	16,956
4	24	134	4	29,752

The daily average of clusters' PLPs are shown in Fig. 9 to Fig. 12. These load patterns are obtained based on the likelihood of customers of each consumption class in all clusters. For more understandability of patterns, these profiles are divided by their peak values to become per unit values.

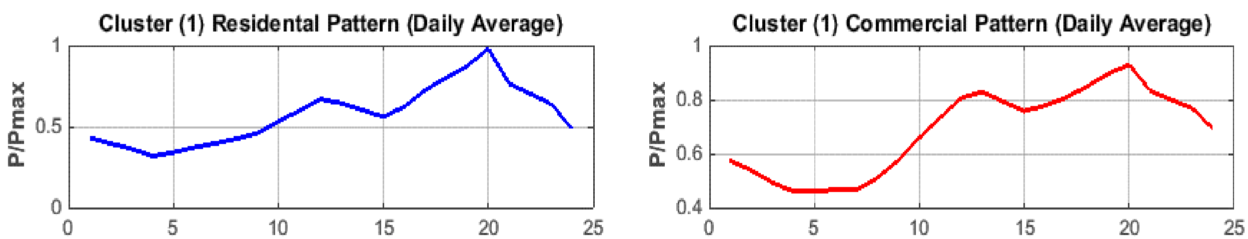

Fig. 9. Extracted PLPs of Cluster 1 (blue: residential, red: commercial)

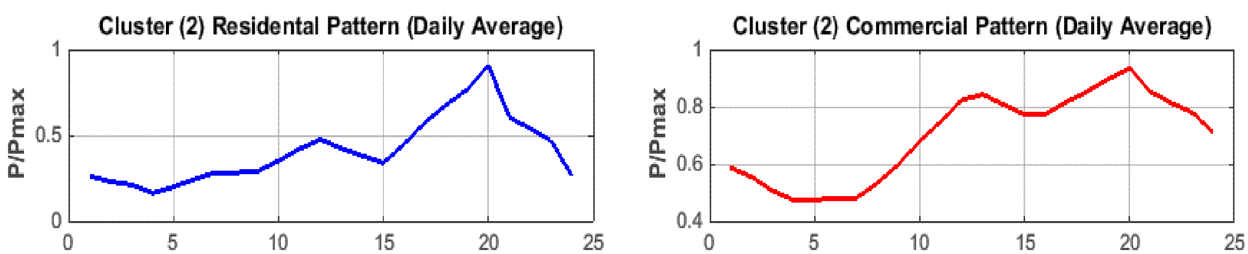

Fig. 10. Extracted PLPs of Cluster 2 (blue: residential, red: commercial)

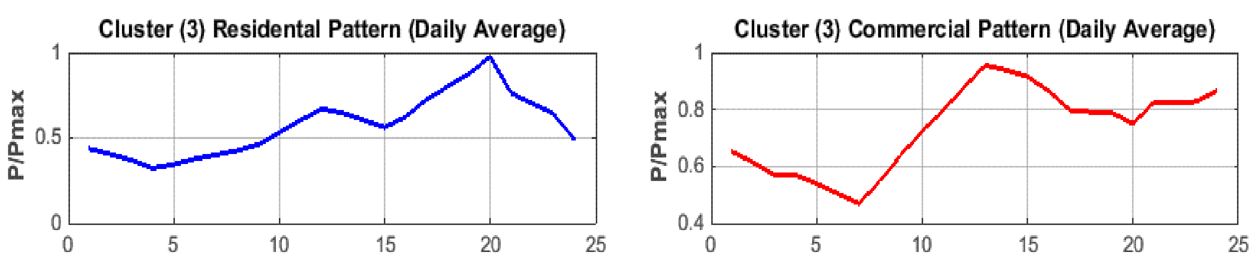

Fig. 11. Extracted PLPs of Cluster 3 (blue: residential, red: commercial)

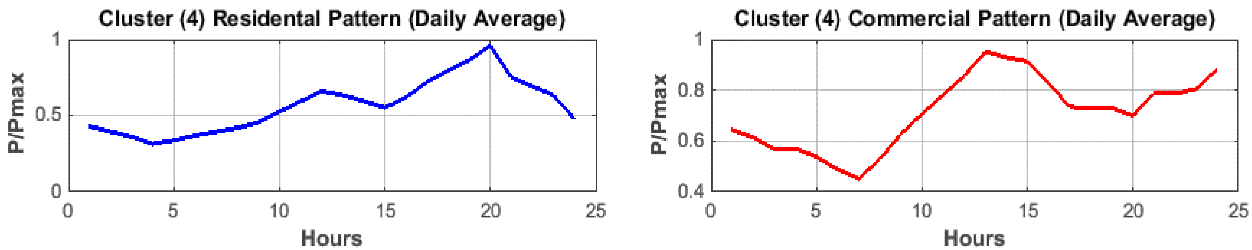

Fig. 12. Extracted PLPs of Cluster 4 (blue: residential, red: commercial)

A comparison between this paper's load estimation method and well-known load allocation technique proposed in [30] is depicted in Fig. 13. This figure provides a 24-hour mean load profile comparison of LV transformers located in candidates LV feeders. As can be seen clearly in Fig. 13, even in feeders 16, 20, 24, and 32 which measurement devices are installed, the obtained LV transformers' load by load allocation method are not equal to actual values. This means that the existing number of meters are not adequate for achieving a suitable amount of accuracy. In contrast, the estimated load of LV transformers is precisely the same as recorded data obtained from meters. Thereby, the presented load estimation method in this paper is quite accurate and could be utilized for loss estimation purposes.

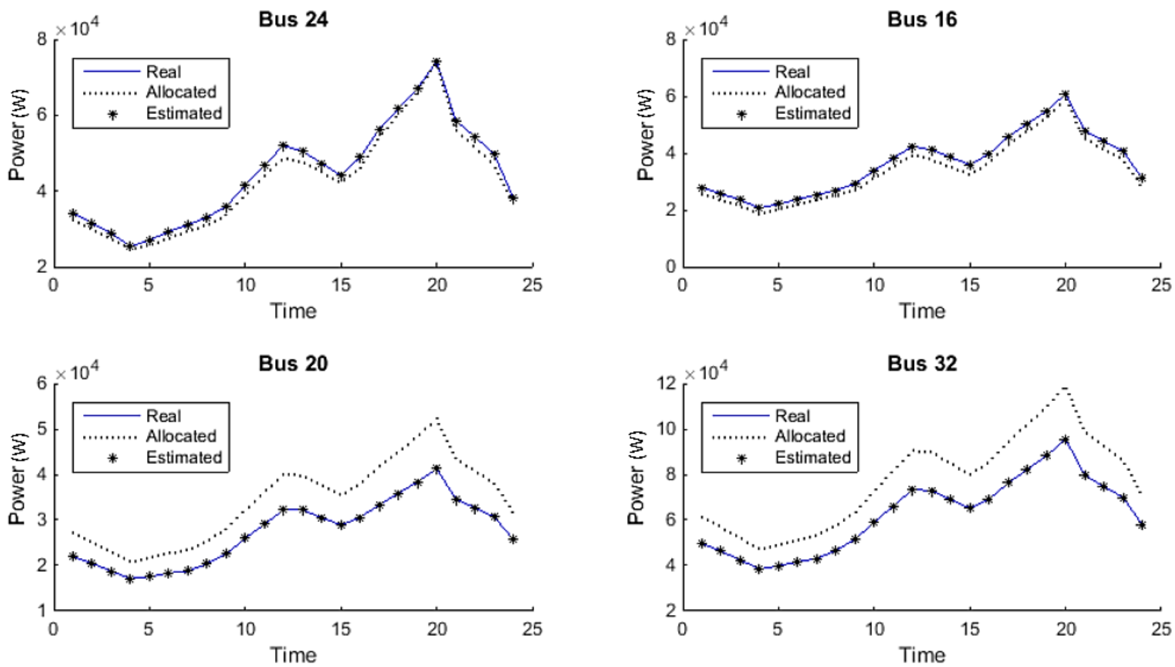

Fig. 13. Transformers' load estimation (daily mean) results for candidate feeders (Allocated: [30])

The results of the proposed load estimation technique demonstrated that the estimated load in buses that have meters are following their actual values. It must be mentioned that an FCM step is used in the load allocation method to make the results more accurate. Despite using an extra clustering step in the load allocation process used in [30], the results show that the proposed load estimation method proposed in this paper in selected buses is by far more accurate than that of in [30]. The related results of load estimation in other buses are also more acceptable than those obtained in the load allocation method. To clarify more, the results of average load estimation errors, as well as average load allocation errors, can be found in Fig. 14. This figure vividly proves the fact that the proposed load estimation approach has higher accuracy in comparison to previous works.

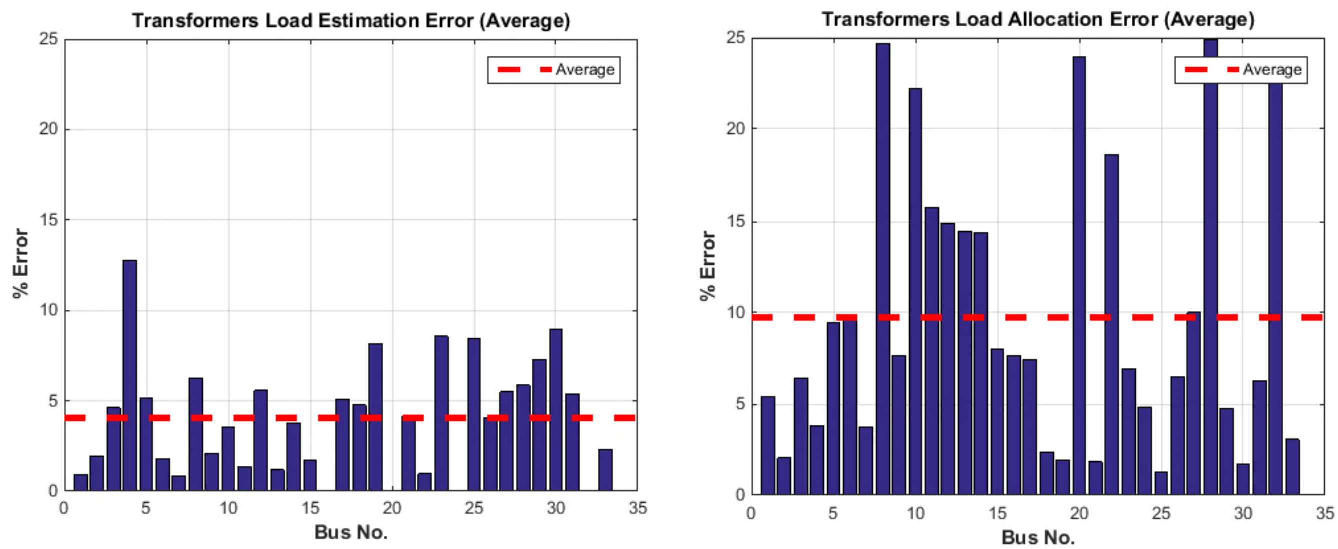

Fig. 14. Load estimation accuracy comparison (left: this paper - right: [30])

In order to obtain the values of R for each LV feeder, it is considered that $\alpha = 0.01$. Afterward, the calculated values for LV feeders will be sorted ascending. Again, index R will be calculated and compared for other values of α (0.05, 0.015, and 0.02) in Fig. 15, to show the robustness of the defined index.

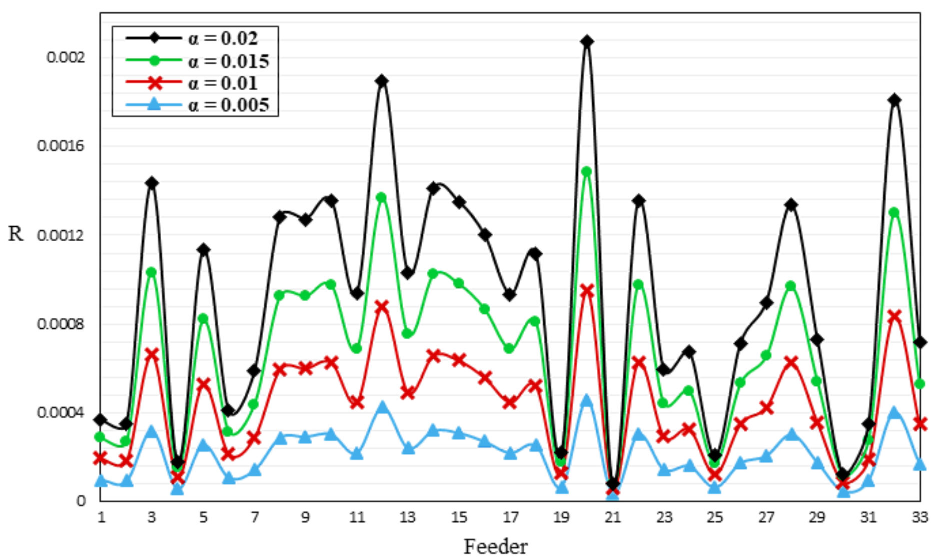

Fig. 15. NTL possibility per capita for LV feeders

This comparison illustrates the independency of LV feeders ranking in terms of NTL to eligible values of α. This studies then follows by an intentional 100kWh/30days electricity theft is applied to all feeders so as to find the sensitivity of R to theft in every single feeder separately. The increase in R for LV feeders after occurring new thefts proves the robustness of this index. These results are shown in Fig. 16.

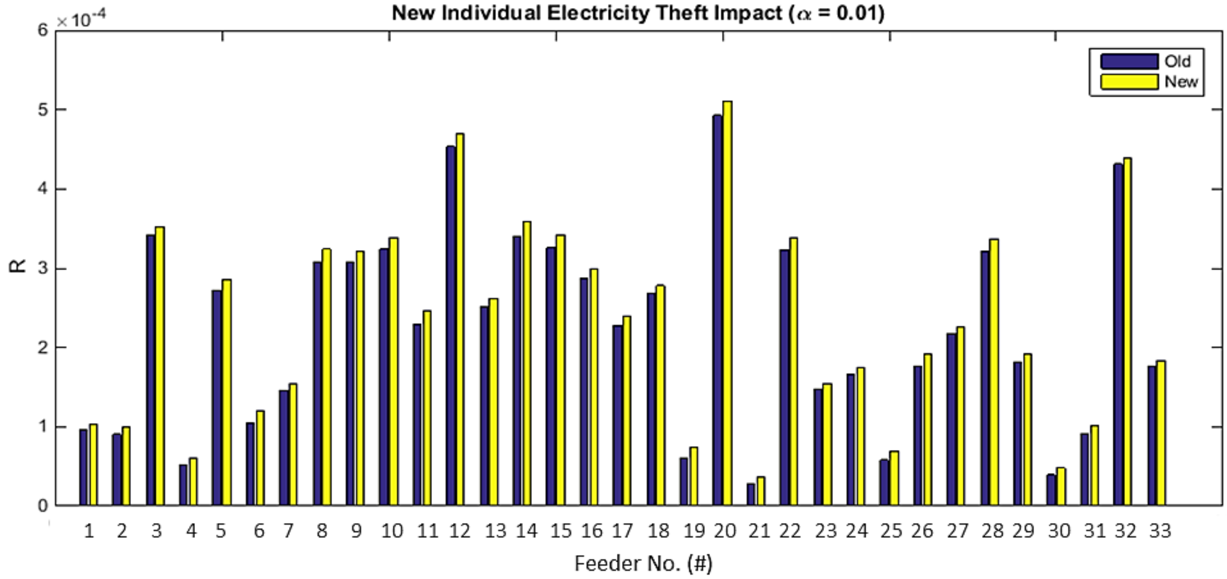

Fig. 16. Impact of intentional theft on *R* (in each feeder separately)

Finally, a comparison between the value of total energy loss, percentage of NTL, and *R* for all LV feeders are presented in Fig. 17. Regarding this result, as mentioned before, those LV feeders with higher values of *R* are supposed to be on the top of the investigation list.

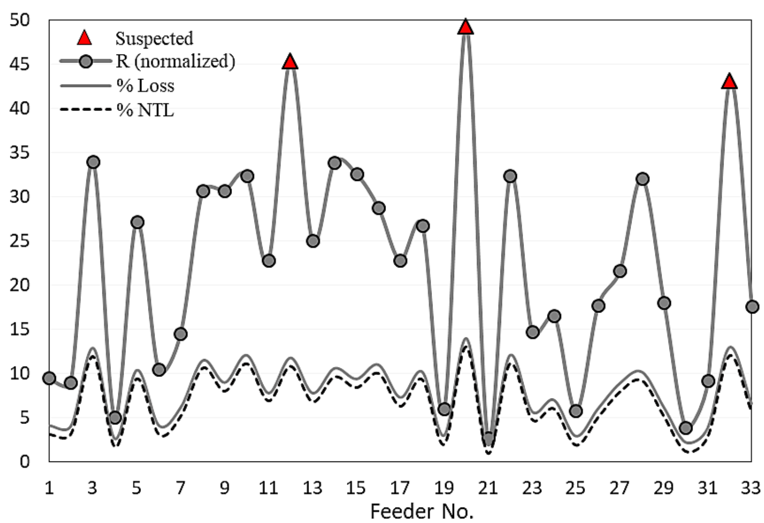

Fig. 17. Comparison of %Loss, %NTL, and *R* in LV feeders

This paper truly shows that the proposed method can be useful to find out the LV feeders with a high possibility of NTL for further investigations. Due to the complexity of radial distribution networks and also the huge number of customers, the idea of investigating every single consumer for finding fraud is not only time-consuming but highly costly. Therefore, prioritizing the most probable areas of fraud for investigation can be noticeably time-efficient and cost-efficient.

V. DISCUSSION

A. Optimal Number of Clusters

In order to achieve an acceptable result in the clustering procedure, it is necessary to know the best number for clusters. There is not any proven way to find an exact number for the clustering program because the results are strongly dependent on the dimension of the questioned dataset. However, several criteria have been introduced that can calculate a suitable value for the number of clusters. The nearest integer number to the result of these criteria would be the optimum number of clusters based on our dataset. One of the most popular criteria in this regard is Silhouette which demonstrates a clustering validity test is as follows [31]:

$$s(x) = \frac{b(x) - a(x)}{\{b(x), a(x)\}} \quad ; \quad -1 \leq s(x) \leq 1 \tag{30}$$

Where,
$s(x)$: Silhouette value for x
$a(x)$: Average dissimilarity of x to all members in its cluster
$b(x)$: Minimum of {average dissimilarity of x to members in neighboring clusters}

Through using (30), a specific value will be obtained for each object. Then, the final value for s can be achieved by averaging of all calculated $\{s(x)\}$. The more resulted criterion s is near 1, the more clusters' members are near their centroids. This means the running clustering program is working properly. For calculating silhouette values based on the different number of clusters, it is needed to select a range of numbers manually as {1,2,3,4,…}, then the resulted s could define the most suitable number of clusters as is shown in Fig. 18.

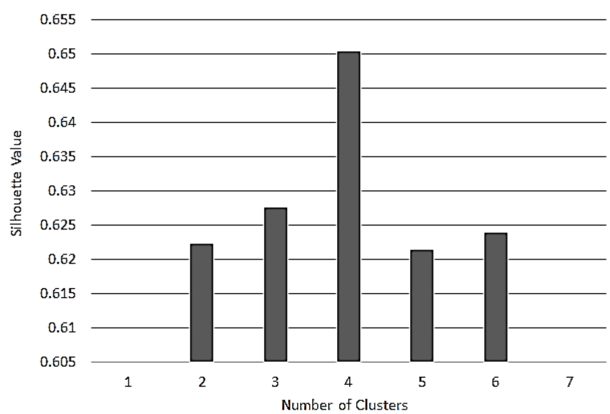

Fig. 18. Silhouette index analysis (case study)

As can be seen in Fig. 18, the optimum number of clusters that yields the best result in this case study is 4. Therefore, at least 4 meters should be installed on the secondary side of the LV transformers to yield better results. Another validity assessment for the aforesaid criterion is carried out as follows.

B. Alteration in Number of Clusters / Meters

As it was explained in previous sections, the optimal number of clusters in this paper was obtained four which means LV feeders should be classified into four groups. Since PLPs extraction, as well as load estimation, must be run in each group of feeders separately, there must be at least one installed meter in each group.

At this stage, the paper is supposed to assess the accuracy of the employed method for finding the number of clusters (meters) through increasing the number of clusters manually. For this purpose, the PLPs extraction, as well as load estimation, has been repeated for 2, 3… 7 (or more) clusters. The results can be found out in Fig. 19.

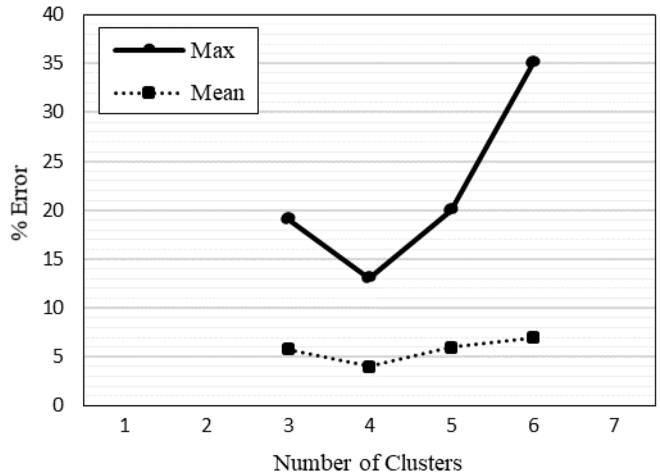

Fig. 19. Impact of alternation in number of clusters (case study)

As can be seen in Fig. 19, the resulted mean/max error of estimations base on 4 clusters is the lowest amongst the others. It should be noted that the value of errors for 2 clusters as well as 7 (or more) clusters has not been depicted due to their high amounts. The results of this step show that in this case study there are four groups of dissimilar customers in LV feeders in terms of consumption manners.

C. Geographical Division of the Network

The energy loss in such traditional networks is mostly originated from NTL, and the NTL is mostly happening in the LV sector. Furthermore, the proposed method of NTL detection in this paper studies the periodic energy of LV feeders. In order to analyze the impact of grouping the network base on the geographical location of LV feeders on load estimation, the case study is divided into four groups of LV feeders according to their MV feeder from which they are being fed as in Fig. 20.

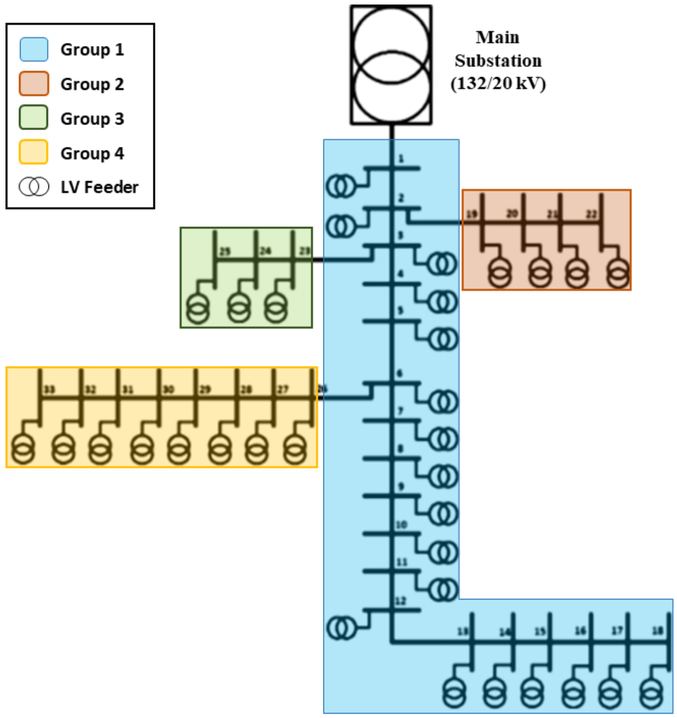

Fig. 20. Geographical division of the case study

It should be noted that in this shape of grouping, the proposed clustering stage, as well as meter placement stage, are bypassed. Moreover, the proposed meter placement method could not be implemented in this approach since it was based on the clustering concepts and the clusters' centroid. However, the existing meters are installed randomly on the LV feeders of each group. The result of the load estimation with this approach is depicted in Fig. 21.

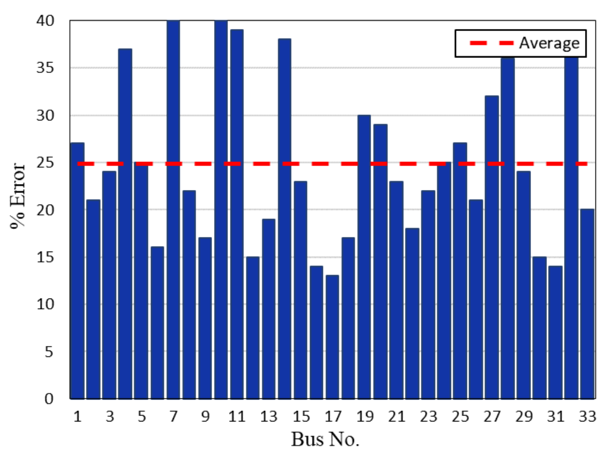

Fig. 21. Load estimation accuracy for geographical division

As was expected, the geographical grouping of LV feeders yielded more inaccurate results in the load estimation stage compared to the proposed method in this paper. Therefore, this method is not suitable for NTL detection targets since the accuracy of load estimation directly affects the rest of the NTL detection procedure. This is due to the fact that the introduced method in this paper is based on the clustering concept, and the extracted load patterns, as well as the estimated loads, were attained utilizing homogeneity of LV feeders in terms of consumption patterns.

D. Impact of Imbalance in the Network

In order to analyze the effect of the network's imbalance in the proposed NTL detection procedure, the studied network is considered once more with different degrees of imbalance between phases. In this matter, the total consumption of the LV feeders remained unchanged while the single-phase loads became more uneven. Accordingly, a 2% and 10% increase in phase imbalance imposed on the network and results of the energy loss as well as the R index were depicted in Fig. 22 and Fig. 23, respectively. The results in these figures show that the increase in imbalance will escalate the energy loss in the network. It must be mentioned that the electricity theft also in this experiment remained unchanged so as the see how imbalance affects the results of NTL detection. As could be seen in Fig. 22 and Fig. 23, the resulted energy loss in this condition has been increased in LV feeders while the values of the R index did not see considerable change compared to the results in Fig. 17. The results illustrate that the increase in imbalance will result in energy loss escalation which is mostly due to increasing TL in the network's branches. However, the most crucial result from this test is the possibility of NTL in LV feeders. As can be observed in Fig. 22 and Fig. 23, the LV feeders which their R index is higher than the others, are quite the same in different imbalance experiments as well as the result of the paper in Fig. 17. It could be concluded that the suspicious feeders which cause a huge portion of NTL in the network could still be recognized for further investigation.

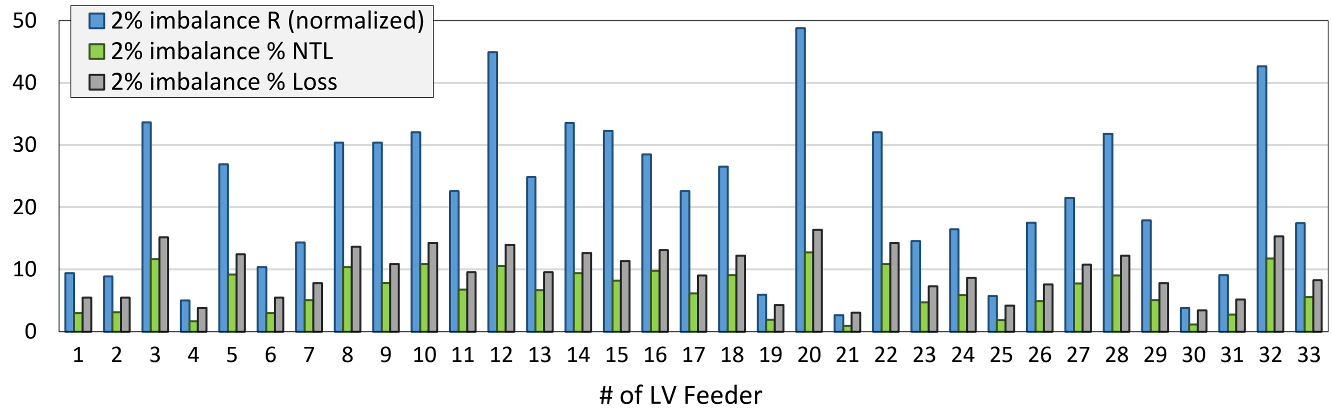

Fig. 22. Impact of 2% increase load imbalance in LV feeders

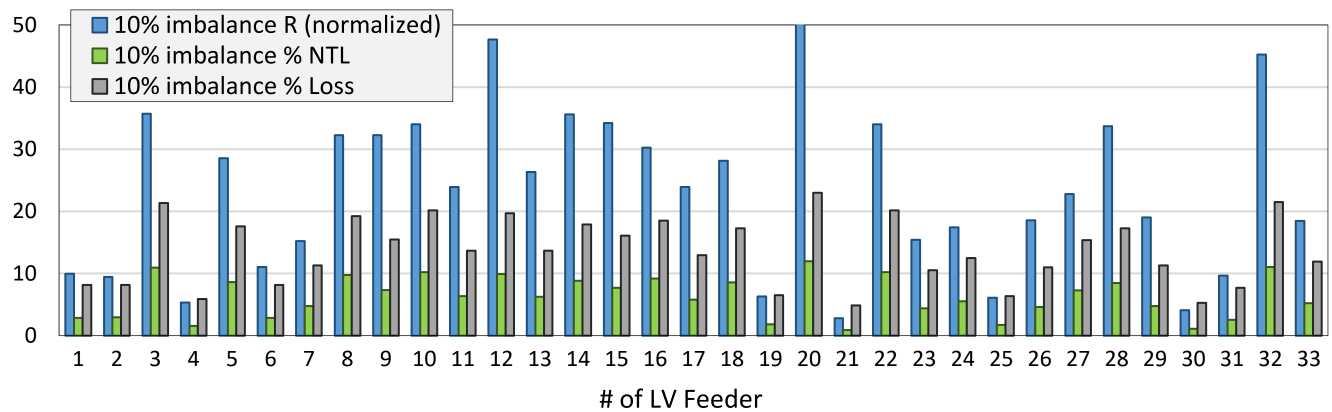

Fig. 23. Impact of 10% increase in load imbalance in LV feeders

VI. Conclusion and Future Works

In this paper, a new NTL evaluation method based on load estimation was proposed in order to find the high-priority feeders in terms of NTL. Regarding the results of this study, it could be seen that the average of errors in the load estimation method experienced about 5.5% decrease compared to the well-known load allocation method. According to the results, the error of load estimation in 80% of feeders was obtained less than 6%. Moreover, in LV feeder number 4 which transformer was far from the related cluster centroid, the error of load estimation was calculated 12.5%. This means that the aforesaid meter placement procedure was by far more suitable than random meter placement. Since the resulted scheme for meter placement is highly vital in the loss estimation procedure, this accuracy can truly affect the results of technical and non-technical losses approximation. Hence, the proposed method seems appropriate to be utilized in the NTL evaluation process. Regarding the results of the NTL evaluation stage in LV feeders, it can be seen that in feeders' number {12, 20, 32} the value of index R was by far higher than others which can be considered as the top-priority group of feeders to be investigated for fraud detection.

This study could be expanded in future works by, for instance, considering the presence of DGs in the network. In this regard, the issue could be seen from two viewpoints. One is the presence of small-scale DGs like PV panels in customer premises. The other is the presence of DGs in grid-level scale within the distribution network. Primarily, it will not be far-fetched to assume that either on customer-level or grid-level scale, DGs must have their own measuring device to measure their generation in practice. In customer-level DGs, the power measured by their separate meter is mostly in hand and the generation and consumption data could be treated either aggregated or separated accordingly since both consumption data and generation data are available. However, in grid-level DGs, they might (or might not) have their own transformer. In case they had their own transformer, the production of them is available and could be considered as an independent node in the network. In case they do not have their own transformer and are directly connected to one of the nodes because they at least have their own energy meter, their generation could be deducted from the total metered energy by the transformers' energy meter. Therefore, in this method, they can be treated as separate generation data and could be considered as a new feature in the clustering stage. Moreover, another future direction of the current work could be studying the impact of metering noises on the result of the NTL detection, which could be further investigated as a sensitivity analysis problem.

APPENDIX

The idea of utilizing α, the relating linear coefficient between energy loss and energy consumption in (18), originated from the results of previous numerical analysis on the historical data from the main substation. In this light, the historical data on the monthly injected energy that the measurement system of the main substation recorded is required. Besides, the historical data on the consumption bills of customers is also available. The difference between these two parameters yields the total energy loss in the network. By drawing the monthly samples of energy loss and total consumption bills as a scatter, the relation between these two parameters was achieved. Afterward, the linear regression tool was utilized in order to obtain the relation between energy loss and total energy consumption in the network.

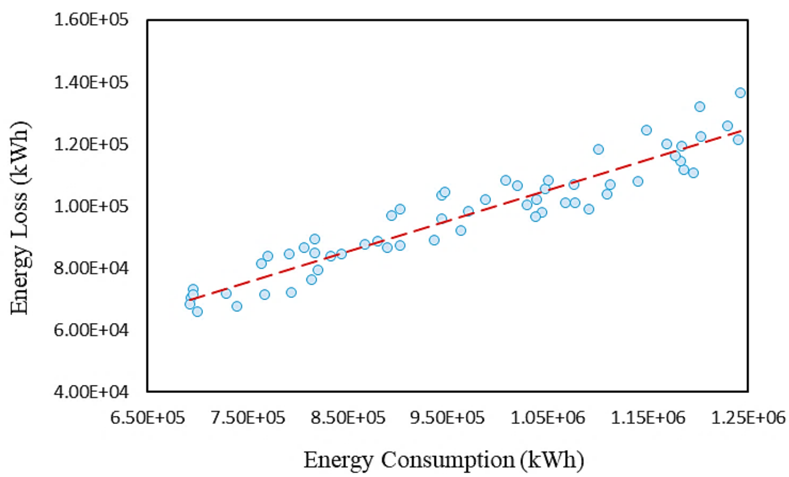

Fig. 24. Correlation between loss and energy consumption

The actual relation between these two parameters would be quadratic due to the loss formula. However, the utilized monthly sampling data over five years shows that the relation between energy consumption and energy loss could be approximated linearly. This conclusion then could be approximately extended to the whole network due to the structure of the network as well as the consumption pattern of residential/commercial customers in distribution networks. The result of this numeral analysis is depicted in Fig. 24. The results of the load estimation as well as NTL detection prove that the deployed coefficients yield acceptable results.

REFERENCES

[1] Manito ARA, Bezerra UH, Soares TM, Vieira JPA, Nunes MVA, Tostes MEL, et al. Technical and non-technical losses calculation in distribution grids using a defined equivalent operational impedance. IET Gener Transm Distrib 2019;13:1315–23. https://doi.org/10.1049/iet-gtd.2018.5334.

[2] Yip S-C, Tan W-N, Tan C, Gan M-T, Wong K. An anomaly detection framework for identifying energy theft and defective meters in smart grids. Int J Electr Power Energy Syst 2018;101:189–203.

[3] Kim JY, Hwang YM, Sun YG, Sim I, Kim DI, Wang X. Detection for Non-Technical Loss by Smart Energy Theft With Intermediate Monitor Meter in Smart Grid. IEEE Access 2019;7:129043–53. https://doi.org/10.1109/access.2019.2940443.

[4] Huang S-C, Lo Y-L, Lu C-N. Non-technical loss detection using state estimation and analysis of variance. IEEE Trans Power Syst 2013;28:2959–66.

[5] Hock D, Kappes M, Ghita B. Using multiple data sources to detect manipulated electricity meter by an entropy-inspired metric. Sustain Energy, Grids Networks 2020;21:100290. https://doi.org/10.1016/j.segan.2019.100290.

[6] Villar-Rodriguez E, Del Ser J, Oregi I, Bilbao MN, Gil-Lopez S. Detection of non-technical losses in smart meter data based on load curve profiling and time series analysis. Energy 2017;137:118–28.

[7] Spirić J V, Stanković SS, Dočić MB. Identification of suspicious electricity customers. Int J Electr Power Energy Syst 2018;95:635–43.

[8] FG SR, Ugarte LF, Sarmiento DA, de Almeida MC. Assessing the Normalized Residuals Technique with AMB-SE for Non-Technical Loss Detection. 2019 IEEE Milan PowerTech, IEEE; 2019, p. 1–6.

[9] Madrigal M, Rico JJ, Uzcategui L. Estimation of Non-Technical Energy Losses in Electrical Distribution Systems. IEEE Lat Am Trans 2017;15:1447–52. https://doi.org/10.1109/TLA.2017.7994791.

[10] Yip S-C, Wong K, Hew W-P, Gan M-T, Phan RC-W, Tan S-W. Detection of energy theft and defective smart meters in smart grids using linear regression. Int J Electr Power Energy Syst 2017;91:230–40.

[11] Simões PFM, Souza RC, Calili RF, Pessanha JFM. Analysis and short-term predictions of non-technical loss of electric power based on mixed effects models. Socioecon Plann Sci 2020:100804.

[12] Ahmad T. Non-technical loss analysis and prevention using smart meters. Renew Sustain Energy Rev 2017;72:573–89.

[13] de Souza MA, Pereira JLR, Alves G de O, de Oliveira BC, Melo ID, Garcia PAN. Detection and identification of energy theft in advanced metering infrastructures. Electr Power Syst Res 2020;182:106258.

[14] Nazmul Hasan M, Toma RN, Nahid A Al, Manjurul Islam MM, Kim JM. Electricity theft detection in smart grid systems: A CNN-LSTM based approach. Energies 2019;12:1–18. https://doi.org/10.3390/en12173310.

[15] Lu X, Zhou Y, Wang Z, Yi Y, Feng L, Wang F. Knowledge embedded semi-supervised deep learning for detecting non-technical losses in the smart grid. Energies 2019;12:1–18. https://doi.org/10.3390/en12183452.

[16] Viegas JL, Esteves PR, Vieira SM. Clustering-based novelty detection for identification of non-technical losses. Int J Electr Power Energy Syst 2018;101:301–10.

[17] Figueroa G, Chen YS, Avila N, Chu CC. Improved practices in machine learning algorithms for NTL detection with imbalanced data. IEEE Power Energy Soc Gen Meet 2018;2018-Janua:1–5. https://doi.org/10.1109/PESGM.2017.8273852.

[18] Ghasemi AA, Gitizadeh M. Detection of illegal consumers using pattern classification approach combined with Levenberg-Marquardt method in smart grid. Int J Electr Power Energy Syst 2018;99:363–75.

[19] Lydia M, Kumar GEP, Levron Y. Detection of Electricity Theft based on Compressed Sensing. 2019 5th Int Conf Adv Comput Commun Syst ICACCS 2019 2019:995–1000. https://doi.org/10.1109/ICACCS.2019.8728396.

[20] Henriques HO, Correa MRLS. Use of smart grids to monitor technical losses to improve non-technical losses estimation. SBSE 2018 - 7th Brazilian Electr Syst Symp 2018:1–6. https://doi.org/10.1109/SBSE.2018.8395924.

[21] Aryanezhad M. A novel approach to detection and prevention of electricity pilferage over power distribution network. Int J Electr Power Energy Syst 2019;111:191–200.

[22] Ahmad T, Chen H, Wang J, Guo Y. Review of various modeling techniques for the detection of electricity theft in smart grid environment. Renew Sustain Energy Rev 2018;82:2916–33.

[23] Messinis GM, Hatziargyriou ND. Review of non-technical loss detection methods. Electr Power Syst Res 2018;158:250–66.

[24] Viegas JL, Esteves PR, Melício R, Mendes VMF, Vieira SM. Solutions for detection of non-technical losses in the electricity grid: A review. Renew Sustain Energy Rev 2017;80:1256–68. https://doi.org/https://doi.org/10.1016/j.rser.2017.05.193.
[25] Firoozi H, Mashhadi HR, Armin MA. Distribution Substations Load Estimation with Minimal Meter Placement through a Fuzzy C-Means Clustering Approach. Electr. Distrib. (CIRED), 2016 4th Conf., 2016, p. 1–5.
[26] Ministry of Energy n.d. https://www.moe.gov.ir.
[27] Li C, Wang J, Li X. An optimized FCM method for electric load clustering. 2008 Third Int. Conf. Electr. Util. Deregul. Restruct. Power Technol., IEEE; 2008, p. 882–6.
[28] Liberti L, Lavor C, Maculan N, Mucherino A. Euclidean distance geometry and applications. SIAM Rev 2014;56:3–69.
[29] Labeeuw W, Deconinck G. Residential electrical load model based on mixture model clustering and Markov models. IEEE Trans Ind Informatics 2013;9:1561–9.
[30] Carmona C, Romero-Ramos E, Riquelme J, Gomez-Exposito A. Distribution transformer load allocation from substation measurements and load patterns. 2010 IEEE PES Innov. Smart Grid Technol. Conf. Eur. (ISGT Eur., IEEE; 2010, p. 1–8.
[31] Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987;20:53–65. https://doi.org/https://doi.org/10.1016/0377-0427(87)90125-7.