International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 12, No.1 (Sep-2022)

https://dx.doi.org/10.12785/ijcds/120160

The Detection of Non-Technical Losses and Electricity Theft by Smart Meter Data and Artificial Intelligence In the Context Of Electric Distribution Utilities: A Comprehensive Review

Rakhi Yadav[1] and Yogendra Kumar[2]

[1, 2] Department,of Electrical Engineering, MANIT Bhopal, India

Received 31 Jul.2020 Revised 04 Apr. 2022, Accepted 19 Jun. 2022, Published 30 Sep. 2022

Abstract: Across the world**,** electric distribution utilities are facing two major challenges i.e., non-technical losses and electricity theft. These losses occur extremely high. Consequently, electric distribution networks' performance deteriorates drastically. Many traditional methods are in practice to detect and minimize these losses, but these are not so effective and also very time-consuming. Hence, keen researchers are looking forward to recent technology, i.e., Artificial Intelligence, because NTL detection by Artificial Intelligence is superior to the traditional techniques in terms of performance such as accuracy. By the application of AI technique on dataset generated by smart meters, electricity theft and NTL are filtered out. This paper describes the causes of NTL followed by an impact on economy. Further, we have thoroughly studied various exercises of technical surveys. Thereafter, based on different AI techniques and essential parameters, a comparison with the existing works has been analyzed. Various simulation tools and compatible environments have been explained. Moreover, multiple challenges occur during AI-based detection of NTL, and their possible solutions are also being discussed.

Keywords: Artificial Intelligence, Expert System, Electricity Theft, Non-Technical Loss, Smart Grid, Smart Meter, SVM

1. Introduction

The overall growth of any country depends on the performance of power grids. As consumers receive energy at a high cost due to limited availability of energy resources. Hence, electricity must be utilized optimally. The monitoring, controlling, and prediction of energy consumption is possible by smart grid system. In smart electrical system, the electricity consumption data is collected by the collector center through the internet and sensors. The electricity bill is calculated according to the real-time electricity consumption data collected by a smart meter. The load balancing problem can also be handled by smart electricity system. This system is very useful for minimizing the power losses and increasing revenue [1]. The device, smart meter (a computerized version of a conventional meter), makes all these operations possible smoothly.
For the calculation of electricity loss, these days, smart meters are becoming very useful. These meters send real-time information at regular intervals of time to the distribution company for better monitoring and billing. Implementation of smart meter enables bidirectional communications between the consumers and power distribution utility [2]. It is also helpful in load balancing by controlling the limit of the maximum demand of energy consumption. Moreover, these meters have the features of disconnecting and reconnecting the electricity supply automatically from any remote place [3].

The electric distribution system suffers with two losses namely as technical and non technical. Main sources of technical losses are internal resistance of electric machines. While, the major sources of non technical losses (NTL) are electricity theft, accounting, and faulty infrastructure [6]. These losses may vary up to 9-12% [4] or 2-6% [5] for the non efficient system. The grids get adversely affected by NTL. Therefore, distribution utility is worried to overcome these challenging issue. The whole world loses huge revenue due to these losses. So, the utility athourity has vital

E-mail: yadavrakhi.87@gmail.com, ykmact@yahoo.co.in

interest to minimize these losses to enhance the reliability of the grids and revenue recovery.

Many researchers have considered the above-stated issues and have given their possible solutions based on AI, smart meter data, genetic algorithms, etc,. Support vector machine (SVM) based model has been applied for detection of NTL in [7, 8]. In [9], authors have used the SVM with genetic algorithms. Expert systems with machine learning are also used for the detection of the NTL [10]. Convolution neural network (CNN) and artificial neural networks (ANN) are also applied in the references [11] and [12], respectively. Random forest (RF) and CNN-based model used in [13] for NTL detection. In 2020, the image-based supervised learning [14] and support vector data description (SVDD) [15] based model are applied to enhance the accuracy of the detection rate of NTL. Moreover, a lot of work has been done for the detection of NTL, the detailed discussion has been given in the next section. From the above existing literature, it is clear that the topic of this survey is still in trend for the research area. In our survey paper, a discussion on an AI-based model for detecting NTL along with some other important models has been precised. The details about different simulation tools have also been given, which are used for the above-mentioned models. Moreover, we have also discussed the causes, economic effects, and variations of the NTL followed by important features which are very useful to detect the NTL. Last but not the least, some most severe challenges which occur during the implementation of the machine learning-based model have been explained, and their possible solutions have also been suggested along with future research scope in this thrust area.

This paper is organized in such a way: section 2 explains the related work. Various challenges for applying the AI technique to detect and reduce the non-technical losses have been discussed in section 3 and possible solutions followed by section 4. Conclusion and possible research scope are given in section 5.

2. Literature Review

This survey has explained several reasons to occur NTL, variations of NTL at different places, economic effects of the NTL, and important features used for detection of NTL. Moreover, existing works related to AI-based detection of the NTL have been reviewed and their comparative analysis has been done. Furthermore, various available simulation tools along with its compatible environments have been discussed.

A. Causes of NTL:

It has been observed that non-technical losses or electricity theft occurs due to many reasons. The major causes of NTL are listed below-

- Un-metered supply
- Tampering with meters
- Due-bills
- Human and technical errors in meter reading
- Bypassing metering equipment
- Broken or faulty meter
- Some financially weak consumers cannot pay the electricity bill while, some consumers are unwilling to do so [16, 17].

B. Economic Effects of NTL:

The revenue increases by reducing the NTL, directly related to the inspection cost, i.e., high NTL requires more inspections. The relation between return on investment and efforts for detection of NTL is in a bell shape shown in figure 1 [10]. This figure clearly shows that the excessive money spends for detection and control of NTL after a threshold. It may be higher than the revenue recovery between return rate and efforts needed for detection and prevention from NTL which is not economic at all.

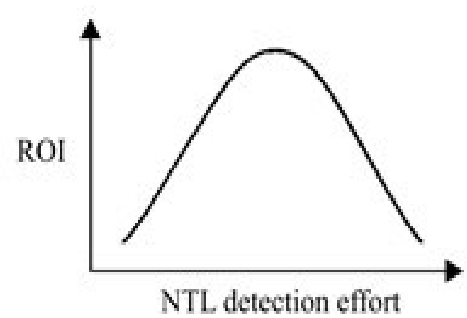

Figure 1. Relation between NTL detection efforts and ROI

C. Variation of NTL

In this survey, we have found that proportion of the NTL varies as per country. Basically, the proportion of the NTL changess on growth status of any country. In table 1, NTL proportions in the different countries have been shown.

TABLE I: List of NTL and their proportions w.r.t.,different countries

Reference	Country/ State	NTL Proportion
[16]	Turkey	4-73%
[18]	India	Up to 70%
[19]	Rwanda	18%
[20]	Brazil	3-40%

D. Features

In this survey, we have found that various types of features are being used for detecting electricity theft or NTL. Here, some important features are being discussed which have mostly been used in the literature so far.

1) Monthly consumption

Monthly consumption used as a feature in calculating the average consumption of power, which is very useful for detecting the NTL[7, 8, 9, 21, 22, 23, 24]. Based on six months' meter readings, the average consumption, standard deviation and maximum consumption have been found [25].

2) Smart meter consumption

Many researchers have used power consumption as a feature for the detection of the NTL. In different works, different time period of the consumptions has been used i.e. 15 minutes' time series used in [26, 27] whereas intervals of 30 minutes have been used in [28,29]. In references [30, 31, 32], authors have used the maximum consumption in the time series.

3) Master data

Master data has a lot of information related to the consumer's consumption and behavior. This information is used as a feature in the detection of electricity theft or NTL. The connection type is residential or industrial, location of consumers, number of phases, voltage range, and meter types are used as features in the [22]. In the reference [24], authors have used the contracted voltage, type of voltage, electricity tariff, and phases as the features.

4) Creditworthiness (CWR)

Some researchers have used this feature to detect electricity theft based on consumer behavior and financial status. It is used by the authors in [8, 9, 21]. The range of this feature varies between 1 to 5, which depends on the scenario of bill payments. This feature provides detailed information on the income, payment performance, and financial conditions of the consumer.

5) Load consumption rate (LCR):

This feature is useful for the detection of the electricity consumption nature of the consumers. If, the consumption rate of a particular consumer is very lower than the previous time consumption rate, it is the possibility that the user may be in theft categories.

6) Maximum demand indicator (MDI):

This feature is very helpful to identify that particular user who has used the overload than specified registered load. This feature is playing an important role in detecting electricity theft or NTL. The amount of excessive energy demand is calculated as:

$$Overload = \mathrm{MDI} - \mathrm{Load}$$

7) Arear/ late payment surcharge (LPSC):

This feature shows that customer ignorance or delaying payments. The amount of LPSC is increases as delaying the electricity bill payment from the due time of payment. This feature is used in the model for the detection of NTL.

E. Expert systems and fuzzy systems

Several AI-based models detect NTL or electricity theft. In this subsection, we have mainly discussed those works which are based on either expert system or fuzzy system, or both. The fuzzy logic expert system and SVM are used to detect NTL in the reference [7]. This approach has been used 100k consumer's dataset for the experiment. In this approach, recorded recall is 0.72. Another approach for detecting NTL is based on fuzzy c-means clustering [25], in which customers are categorized based on the different features. This approach has 0.745 precision value.

In reference [10], an expert system is used for detection of NTL; the abstract view of this model is displayed by figure 2. In this approach, initially, the consumer's consumption and inspection dataset are collected. In the next step, the data which are collected in the previous steps are loaded in the database. Thereafter, useful feature extraction is performed according to the dataset. In order to detect NTL, a classifier is trained according to the selected features and previously carried out inspection datasets. The prediction process of NTL is performed in the next step. Finally, the suspicious consumers are filtered out for physical inspection.

F. Neural networks

Neural networks (NN) are motivated by the working of the human brain and its ability to learn complex hypotheses from data. A detailed description of NN with an example is given in [33]. Shaun Li *et al.* [13] have introduced a hybrid model for the detection of electricity theft. This model detects electricity theft in an automatic manner. In this model, CNN is applied for features learning from massive smart meter data, and the backpropagation is used for updating network parameters during the training phase. Moreover, the risk of overfitting is reduced by the dropout layer. This work has also focused on the data imbalance problem, resolved by SMOTE algorithm. In the second phase, the model is trained by

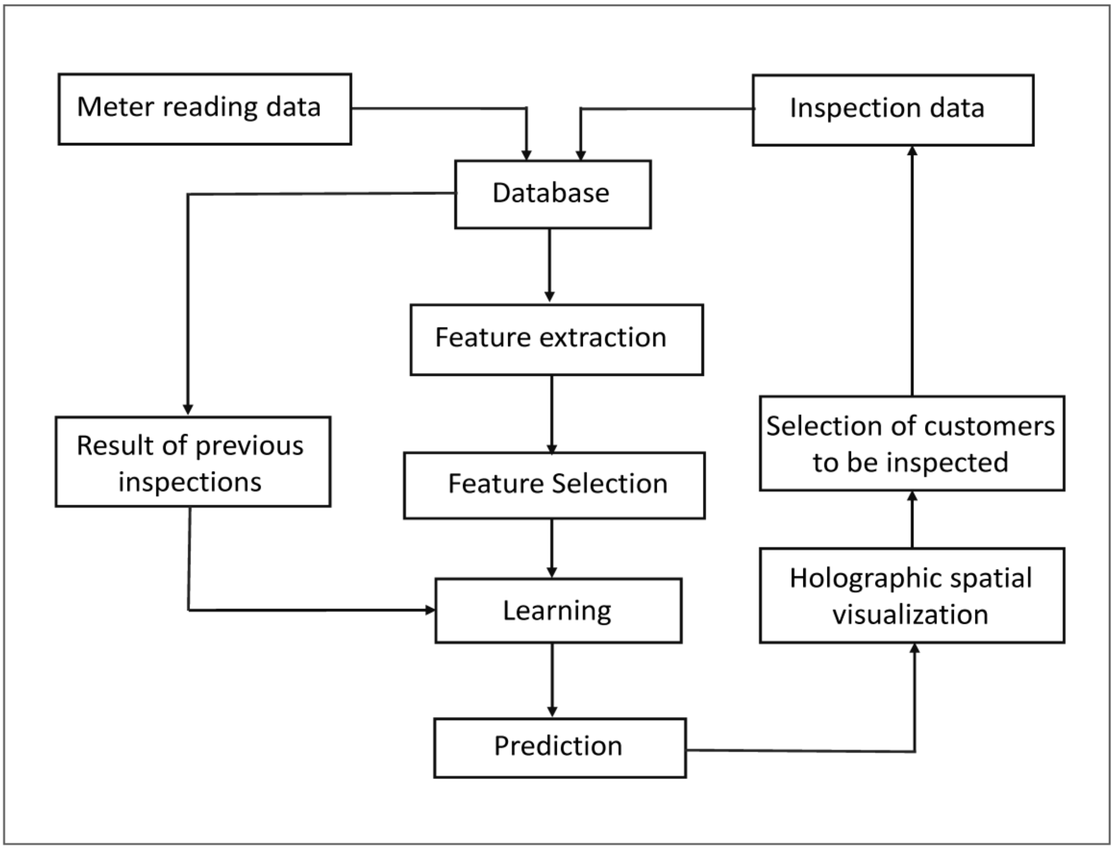

Figure 2. Flow chart for the detection of NTL by applying Machine Learning techniques and expert knowledge

using random forest (RF) and obtained features. Here, we have explained the AI-based model with the help of a flowchart which is shown in the figure 3. To test the accuracy and efficacy of the approach on the real data set from Ireland and London, some AI techniques such as SVM, RF and LR are applied on the same problem as a benchmark. The average test value of Precision and Recall are 0.97 for both and the test value of AUC on SEAI and LCL are 0.99 and 0.97, respectively. Neural networks based models have also been used in [26, 34, 35]. In Extreme learning machines (ELM), the learning process is done from hidden to output layer. In [34], ELM has been used to detect NTL. The accuracy of this approach is 0.546. In reference [35], the authors have used dataset size is 20k having the precision and accuracy 0.626 and 0.686, respectively.

G. Support vector machines

Support Vector Machines (SVM) are the supervised learning model used to analyze data for classification and regression. Several approaches based on SVM have been introduced in [8, 21, 30]. The creditworthiness ranking (CWR) has been used as a feature in [8], and test recall has been recorded as 0.53. The test accuracy and recall reference [21] are 0.86 and 0.77, respectively. In [30], authors have run the 5k customers dataset on SVM, kNN and neural network classifiers. This approach has achieved the test accuracies 0.9628, 0.9620, and 0.9448 correspondings to SVM, kNN, and neural network.

T. Hu et al. [15] have focused on two major issues (class imbalance and high dimensionality of the data) in NTL detection. Authors have proposed a novel technique for the detection of NTL, which is based on deep learning. Deep neural networks perform well on high-dimensional data. A bidirectional Wasserstein generative adversarial network (BiWGAN) has been proposed for feature extraction in this proposed method. The class imbalance problem is handled by using support vector data description (SVDD). In which, 5000 consumer's data are used. The performance metrics used in this approach are detection rate (DR) and false-positive rate (FPR). The test DR and FPR values are 87.78 % and 3.7 %, respectively.

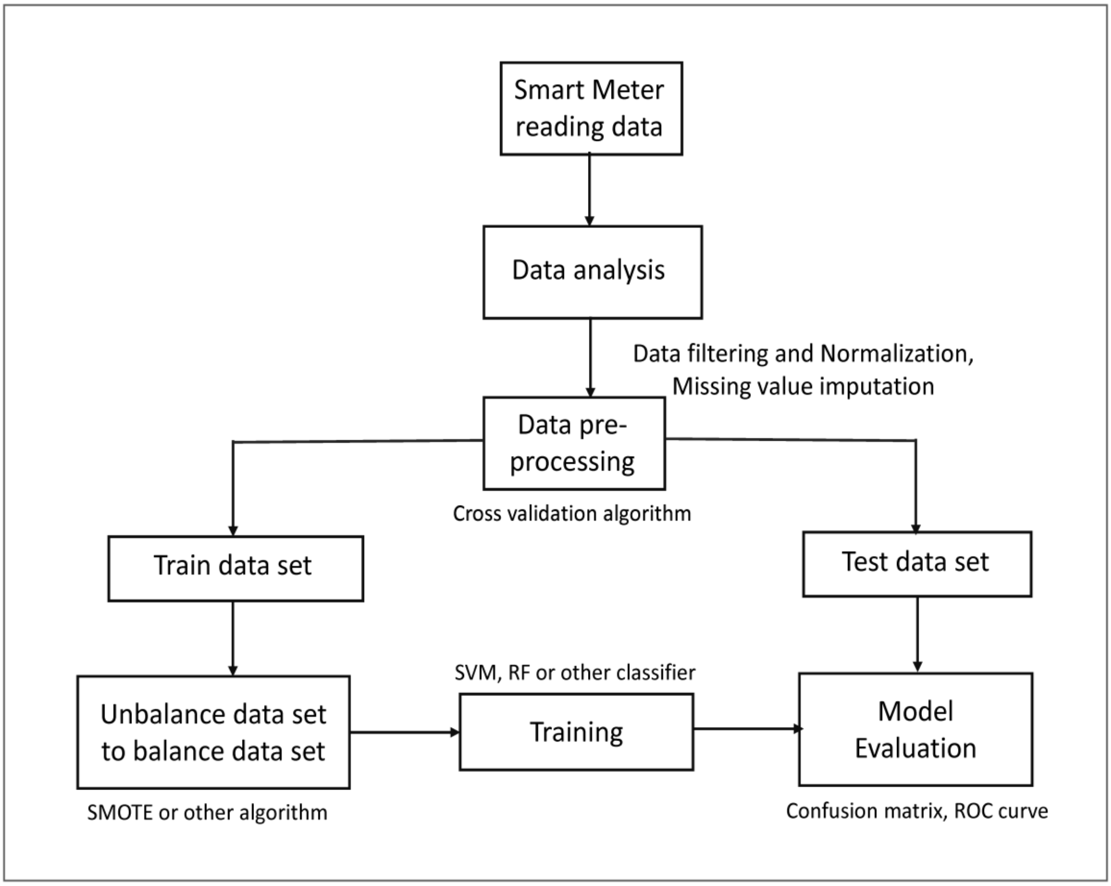

Figure 3. flow chart for the detection of electricity based on AI model

The inspection dataset and the monthly consumer consumption dataset have been used in the reference [23]. An expert system is used which optimizes the fuzzy system parameters using stochastic gradient descent [37] by data set. The fuzzy system performs better than the Boolean system. The AUC scores of this approach are 0.55, 0.55, and 0.465 on an NTL proportion of 5% for the optimized fuzzy system, SVM, and Boolean system.

H. Genetic algorithms

Authors have proposed a genetic SVM-based model [9] for 1,171 customers, which is an extension of works [8, 21]. The genetic algorithm achieves the main motive to optimize the hyperparameters of the SVM. The recall value of this model is 0.62.E. coasta et al. [38] have introduced a non-linear restricted optimization problem named as stratified sampling. This sampling technique helps to reduce the total energy loss and gets increment in accuracy of the detection of NTL by minimizing statistical variance.

I. Rough sets

A rough set-based model has been used in reference [39] to detect NTL. In this technique, dataset of 40K customers is used for training and testing but this model has poor accuracy of 0.2. This paper also lacks detail on attributes used by customers. Another technique [24] has also used the rough set analysis applied for NTL detection, which has used the features related to [22]. This is a supervised technique that describes the fraud and regular users. This model achieves a good test accuracy of 0.9322.

J. Other methods

Apart from the above-discussed methods, some more techniques have also been used to detect the NTL. One of these has been described in [40], in which some searching techniques are used for customer master data like genetic, greedy, and best-first search (BFS). In which, the shape factor is calculated by using consumption time series, and it is also used as a feature. The used classifier is the decision tree, and the tested accuracy is 0.9997. A long short-term memory (LSTM) and CNN based approach have been introduced in reference [11], which is also considered the class imbalance problem. The used dataset for training and testing is 10k, and the test accuracy of the approach is 0.89. In [12], Multilayer Perceptron ANN was used for electricity theft detection. In which SOM was trained for clustering customers based on the load curve. In this technique, customers have been categorized into two groups i.e. one is a fraud, and another is non-fraud. The test accuracy of this model is 93% on the 5000 customer's dataset, while the false positive rate is less than 2%.

J. Li and F. Wang [14] have introduced a new and different method for detecting NTL by using image-based semi-supervised learning. In which, real data of the smart meter is transformed into super images, which helps to predict the consumption behavior of customers. This new data format has also been used along with some new features for the analysis of anomalies. The authors have used data set from the power grid of China for the experiments. The test value of AUC is found 0.94. A graph-based classifier have been used in reference [31]. In this approach, plots the graph between feature space and uses the training samples or prototypes. Optimum path forest (OPF) does not learn parameters. Thus, the learning process becomes fast, but predicting capability becomes slower when it is compared with parametric methods. These algorithms are used in [41] and found the test accuracy 0.9. Feature-engineering-based model also used for NTL detection [42]. This approach used the advantages of Finite Mixture Model clustering (FMMC) and Genetic Programming (GP). GP algorithm is used for the identification of new features and FMMC is used for customer segmentation. The data set size has applied 4000 households of 18 months. It gives better accuracy other than the SVM and NN by applying Gradient Boosting Machine algorithm. A fuzzy-based clustering model has been used in reference [43] for the detection of NTL. The objective of this model is to predict the users' consumption behavior. This technique is being tested on real-world dataset. The size of the used dataset is more than 2k users. The AUC score of this approach is 0.741.

Some other methods are also popular to detect NTL, such as decision tree, logistic regression, Gradient-Boosted Tree, random forest, and k-Nearest Neighbors. All the above-discussed literature have been compared to each other based on different performance measures, the number of customers, and data sources which are shown in table 2.

4. Suggested Methodology

After reviewing the related research thoroughly, the following methodology can be implemented to overcome these issues.

A. Handling class imbalance and evaluation Metric:

In the existing literature, authors have used various evaluation parameters such as detection rate, accuracy, area under curve (AUC), precision, and recall. The class imbalance problem affects the accuracy of various models. Hence, to resolve this problem, some other evaluation metrics are also used, such as inspection cost, the rank of the customer, and possible increment in revenue. Mathew's correlation coefficient (MCC) is another important evaluation metric that has been used in reference [44]. It is defined as-

$$MCC = \frac{TN*TP-FN*FP}{\sqrt{(FP+TP)(TP+FN)(TN+FN)(TN+FP)}} \quad (1)$$

Where *TP, TN, FN* and *FP* are the true positive, true negative, false negative, and false-positive rate values. These are the different confusion matrices.

B. Feature description:

In the artificial intelligence-based NTL detection model, the feature description and selection process are two important factors because these have the impact on accuracy of the model. During this survey, we have found that AI-based model has used both traditional [8, 21, 23, 35] and smart meter data [27, 28, 29, 32, 40, 41, 45]. Most of the existing works have used manual feature extraction, which can be done by automatic feature extraction and self-learning.

C. Construction of a publicly available data set:

We have found that different works of literature have used different data sets. Then, the question arises, how to compare different models and how to decide which one is the best model for detection of NTL? The quality of the data set is also not good, due to which data preprocessing is required before the application of any new classifier model. So, there is an urgent need of the publically available standard data set which can be used as the benchmark and should satisfy all the following properties-

- At least one year's real data set must be available because consumers' consumption varies according to the climate.
- The size dataset and inspection data should be sufficient.
- Categorization of the data sets based on residential and industrial consumers.

TABLE II. Summary of models, performance measures and data sets.

Reference	Model	Accuracy	Precision	AUC	Recall	NTL/theft proportion	#Customers	Dataset source
7	SVM+ Fuzzy	-	-	-	0.72	-	100k	TNBD
7	SVM-FIS	0.72	-	-	-	-	36176	TNBD
8	SVM	-	-	-	0.53	-	<400	TNBD
8	SVM	0.60	-	-	-	-	36176	TNBD
9	Genetic SVM	-	-	-	0.62	-	1171	TNBD
9	Genetic- SVM	0.62	-	-	-	-	186,968	TNBD
11	CNN, LSTM	0.89	0.90	-	0.87	-	17120	-
12	SOM, MP-ANN	0.934	-	-	-	-	5000	-
13	RF-CNN	-	0.97	0.99	0.97	-	-	Ireland
14	Image-based semi-supervised	-	-	0.94	-	-	-	-
15	SVDD	-	-	-	-	-	5k	ISET
21	SVM(Gauss)	0.86	-	-	0.77	-	<400	-
23	Bool rules	-	-	0.47	-	5%	700k	-
23	Fuzzy rules	-	-	0.55	-	5%	700k	-
23	SVM (linear)	-	-	0.55	-	5%	700k	-
23	Bool rules	-	-	0.48	-	20%	700k	-
23	Fuzzy rules	-	-	0.55	-	20%	700k	-
23	SVM (linear)	-	-	0.55	-	20%	700k	-
24	Rough sets	0.93	-	-	-	-	N/A	-
25	Fuzzy classification	0.745	-	-	-	-	-	Brazil
26	SOM	0.93	0.85	-	0.98	-	2k	-
27	SVM(Gauss)	0.98	-	-	-	-	1350	-
28	Fuzzy logic	0.55	-	-	-	-	-	TNBD
30	KNN	0.96	-	-	-	-	5k	-
30	SVM	0.96	-	-	-	-	5k	-
30	NN	0.94	-	-	-	-	5k	-
35	NN	0.835	0.249	-	-	-		Brazil
41	OPF	0.90	-	-	-	-	736	-
41	NN	0.53	-	-	-	-	736	-
41	SVM(Gauss)	0.89	-	-	-	-	736	-
41	SVM(Linear)	0.45	-	-	-	-	736	-
41	Decision Tree	0.99	-	-	-	-	N/A	-
43	Fuzzy clustering	-	-	0.741	-	-	$\geq$ 2k	
46	ARMA models	-	-	-	0.62	-	108	-
47	Wide and Deep CNN	0.9404	-	-	-	-	42372	State Grid Corporation of China (SGCC)
48	DT coupled SVM	0.925	-	-	-	-	N/A	N/A
49	(SVM,OPF, C4,5 tree)	0.862	0.544	-	0.64	-	-	Uruguayan Electric Company (UTE)
50	CNN, LSTM	0.966	-	-	-	-	12180	-
51	K-mean-SVM	-		-	0.94	-	5000	CER in Ireland.

5. CONCLUSIONS AND FUTURE RESEARCH SCOPE:

This paper has reviewed various severe impacts of NTLs and electricity theft on economies and major revenue losses. Further, it gives a detailed review of AI-based detection of non-technical losses and electricity theft techniques. Our study has found that the SVM and neural network are the most commonly used methods so far. These methods are applied to features like connection and customer type in addition to consumer consumption profiles i.e, maximum demand, change in demand, and average load. The size of data sets has been used in a wide range i.e. 100 to more than one million in the literature of this survey. This paper has also discussed various challenges of AI-based detection of NTL and electricity detection techniques. It has also discussed some possible solutions of the above stated challenges. In addition, the different simulation tools and compatible environments have also been discussed, which are used during the detection of the NTL.

The following thrust areas of research are still needed to work for any researcher-

- To make the standard data set which can be made publically available.
- To design a model for class imbalance handling.
- To make an automated and robust expert system for the detection of NTL.

REFERENCES:

[1] Mavridou, A.; Papa, M. A situational awareness architecture for the smart grid. In Global Security, Safety and Sustainability & e-Democracy; Springer: Germany, pp. 229–236, 2011.

[2] Zheng, J.; Gao, D.W.; Lin, L. Smart meters in smart grid: An overview. In Proceedings of the IEEE Green Technologies Conference, USA, pp. 57–64, 2013.

[3] Depuru, S.S.S.R.; Wang, L.; Devabhaktuni, V.; Gudi, N. Smart meters for power grid—Challenges, issues, advantages and status. In Proceedings of the IEEE/PES Power Systems Conference and Exposition, USA, pp. 1–7, 2011.

[4] M. Golden and B. Min, Theft and loss of electricity in an Indian state, International Growth Centre, Tech. Rep., 2012.

[5] T. B. Smith, Electricity theft: A comparative analysis,Energy Policy, vol. 32, no. 18, 2004, pp. 2067–2076.

[6] P. Antmann, Reducing technical and non-technical losses in the power sector. World Bank, Washington, DC, 2009.

[7] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and F. Nagi, improving svm-based nontechnical loss detection in power utility using the fuzzy inference system. *IEEE Transactions on power delivery*, 2011, pp. 1284–1285.

[8] J Nagi, A.M. Mohammad, K.S. Yap, S.K.Tiong, and S.K. Ahmed. Non-technical loss analysis for detection of electricity theft using support vector machines. In *PECon, IEEE, 2008, pp.* 907–912.

[9] J. Nagi, K.S. Yap, S.K. Tiong, S.K. Ahmed, and A.M. Mohammad, Detection of abnormalities and electricity theft using genetic support vector machines. In *TENCON,* IEEE, 2008, pp. 1–6.

[10] P. Glauner, Artificial Intelligence for the detection of electricity theft and irregular power usage in emerging markets, PhD-FSTC-2019-07 The Faculty of Sciences, Technology and Communication, *Université du Luxembourg.*

[11] Md. Nazmul Hasan, R. N.Toma, Abdullah-Al Nahid, M MManjurul Islam, Jong-Myon Kim, Electricity Theft Detection in Smart Grid Systems: A CNN-LSTM Based Approach, Energies 2019.

[12] M. A. de Souza, J. L.R. Pereira, Guilherme de O. Alves, B. C. de Oliveira, Igor D. Melo, Paulo A.N. Garcia, Detection and identification of energy theft in advanced metering Infrastructures, Electric Power Systems Research 182, 2020.

[13] S. Li, Y. Han, Xu Yao, S. Yingchen, J. Wang, and Q. Zhao, Electricity Theft Detection in Power Grids with Deep Learning and Random Forests, Hindawi Journal of Electrical and Computer Engineering Volume 2019, Article ID 4136874.

[14] J. Li and F. Wang· Non-Technical Loss Detection in Power Grids with Statistical Profile Images Based on Semi-Supervised Learning, Sensors (Basel); 20(1), 2020

[15] T. Hu, Q. Guo, H. Sun, T. Huang and J. Lan, "Nontechnical Losses Detection Through Coordinated BiWGAN and SVDD," in *IEEE Transactions on Neural Networks and Learning Systems*, doi: 10.1109/TNNLS.2020.2994116.

[16] C. Yurtseven, The causes of electricity theft: An econometric analysis of the case of Turkey, Utilities Policy, vol. 37, 2015, pp. 70–78.

[17] F. B. Lewis, Costly 'throw-ups': Electricity theft and power disruptions, The Electricity Journal, 2015, pp. 118–135.

[18] S. K. Katiyar, "Political economy of electricity theft in rural areas: A case study from Rajasthan", Economic and Political Weekly, 2005, pp. 644–648.

[19] F. M. Mwaura, "Adopting electricity prepayment billing system to reduce non-technical energy losses in Uganda: Lesson from Rwanda", Utilities Policy, vol. 23, 2012, pp. 72–79.

[20] C. C. Ramos, D. Rodrigues, A. N. de Souza and J. P. Papa, On the study of commercial losses in Brazil: A binary black hole algorithm for theft characterization, IEEE Transactions on Smart Grid, 2018, pp. 676–683.

[21] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and M. Mohamad. Non technical loss detection for metered customers in power utility using support vector machines. *IEEE transactions on Power Delivery*, 2010, pp. 1162–1171.

[22] B. C Costa, Bruno LA Alberto, Andr´e M Portela, W Maduro, and E. O Eler, Fraud detection in electric power distribution networks using an ann based knowledge-discovery process, *International Journal of Artificial Intelligence & Applications*, 2013.

[23] P. Glauner, A. Boechat, L. Dolberg, R. State, F. Bettinger, Y. Rangoni, and D. Duarte. Large-scale detection of non-technical losses in imbalanced data sets. In *Innovative Smart Grid Technologies Conference* 2016.

[24] Josif V Spiri´c, Slobodan S Stankovi´c, Miroslav B Doˇci´c, and Tatjana D Popovi´c. Using the rough set theory to detect fraud committed by electricity customers. *International Journal of Electrical Power &Energy Systems*, 62, 2014, pp. 727–734.

[25] E. Werley S Angelos, O. R Saavedra, O. A Carmona Cort´es, and Andr´e Nunes de Souza, Detection and identification of abnormalities in customer consumptions in power distribution systems. *IEEE Transactions on Power Delivery*, 2011, pp. 2436–2442.

[26] Jose E Cabral, Joao OP Pinto, and Alexandra MAC Pinto. Fraud detection system for high and low voltage electricity consumers based on data mining. In *IEEE Power & Energy Society General Meeting*, 2009, pp. 1–5.

[27] S. Shekara S. Reddy Depuru, L. Wang, and V. Devabhaktuni, Support vector ma- chine based data classification for detection of electricity theft. In *Power Systems Conference and Exposition (PSCE), ,2011,* pp. 1–8.

[28] J Nagi, KS Yap, F Nagi, SK Tiong, SP Koh, and SK Ahmed, NTL detection of electricity theft and abnormalities for large power consumers in Malaysia. *Student Conference*, 2010, pp. 202–206.

[29] S. Sahoo, D. Nikovski, T. Muso, and K.Tsuru. "Electricity theft detection using smart meter data". In *Innovative Smart Grid Technologies Conference (ISGT), IEEE Power & Energy Society*, pp. 1–5, 2015.

[30] C. C. Oba Ramos, Andre Nunes De Souza, D. S. Gastaldello, and J. P. Papa. Identification and feature selection of non-technical losses for industrial consumers using the software weka, In *Industry Applications (INDUSCON), IEEE, 2012,* pp. 1–6.

[31] C. C.O. Ramos, A. N. Sousa, J. P. Papa, and Alexandre X. Falcao. A new approach for nontechnical losses detection based on optimum-path forest. *IEEE Transactions on Power Systems*, 2011, pp. 181–189.

[32] C. C.O. Ramos, A. N Souza, Joao P. Papa, and Alexandre X Falcao. Learning to identify nontechnical losses with optimum-path forest. In *Proceedings of the IWSSIP,* 2010, pp. 154–157.

[33] C. M Bishop, Neural networks, a pattern recognition perspective, 1996.

[34] AH Nizar, ZY Dong, and Y Wang. Power utility nontechnical loss analysis with extreme learning machine method. *IEEE Transactions on Power Systems*, 2008, pp. 946–955.

[35] C. Muniz, K. Figueiredo, M. Vellasco, G. Chavez, and M. Pacheco, Irregularity detection on low tension electric installations by neural network ensembles, In *International Joint Conference on Neural Networks*, 2009, pp. 2176–2182.

[36] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, Greg S Corrado, A. Davis, J. Dean, M. Devin Tensorflow: Large-scale machine learning on heterogeneous distributed systems., 2016.

[37] L Bottou. Stochastic learning. In *Advanced lectures on machine learning*, 2004, pp. 146–168.

[38] E. Costa, F. Fabris, A. R. Loureiros, H. Ahonen, and F. Miguel Varej˜ao. Optimization metaheuristics for minimizing variance in a real-world statistical application. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, 2013, pp. 206–207.

[39] JE Cabral, JOP Pinto, EM Gontijo, and J Reis. Rough sets based fraud detection in electrical energy consumers. In *WSEAS International Conference on Mathematics and Computer in Physics, Cancun, Mexico*, 2004, pp. 413–416.

[40] A. H. Nizar, J. H. Zhao, and Z. Y. Dong. Customer information system data preprocessing with feature selection techniques for nontechnical losses prediction in an electricity market, In *International Conference on Power System Technology*, IEEE, 2006, pp. 1–7.

[41] C. C.O. Ramos, A. N. Sousa, J. P. Papa, and Alexandre X. Falcao, Fast non-technical losses identification through optimum-path forest. In *Intelligent System Applications to Power Systems, IEEE, pp. 2009*

[42] R. Razavi, A. Gharipour, M. Fleury, Ikpe Justice Akpan, a practical feature-engineering framework for electricity theft detection in smart grids, Applied Energy 238, 2019, pp. 481–494.

[43] J. L. Viegas, P. R. Esteves and S. M. Vieira, Clustering-based novelty detection for identification of non-technical losses, Electrical Power and Energy Systems 101, 2018, pp. 301–310.

[44] B. W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme", Biochimica et Biophysica Acta (BBA) - Protein Structure 405 (2): pp. 442-451, 1975.

[45] S. Shekara, S.R. Depuru, L. Wang, V. Devabhaktuni, and R. C Green. High performance computing for detection of electricity theft. *International Journal of Electrical Power & Energy Systems*, 2013, pp. 21–30.

[46] Mashima D. and Cárdenas, A. A., Evaluating electricity theft detectors in smart grid networks. *InProceedings of* Research in Attacks, Intrusions, and Defenses. *Springer, 2012, pp.* 210-229.

[47] Zheng, Z.; Yang, Y.; Niu, X.; Dai, H.-N.; Zhou, Y. Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids. IEEE Trans. Ind. Inform. 2018, 14, pp. 1606–1615.

[48] Jindal, A.; Dua, A.; Kaur, K.; Singh, M.; Kumar, N.; Mishra, S. Decision tree and SVM-based data analytics for theft detection in smart grid. IEEE Trans. Ind. Inform. 2016, 12, pp. 1005–1016.

[49] Di Martino, M.; Decia, F.; Molinelli, J.; Fernández, A. Improving Electric Fraud Detection using Class Imbalance Strategies. In Proceedings of the International Conference on Pattern Recognition Applications and Methods (ICPRAM), Vilamoura, Portugal, 2012; pp. 135–141.

[50] Bhat, R.R.; Trevizan, R.D.; Sengupta, R.; Li, X.; Bretas, A. Identifying nontechnical power loss via spatial and temporal deep learning. In Proceedings of the 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE: Piscataway, NJ, USA, 2016; pp. 272–279.

[51] Jokar, P., Arianpoo, N. and Leung, V. C. M., Electricity Theft Detection in AMI Using Customers' Consumption Patterns, *IEEE Transactions on Smart Grid.*7, 2016, pp. 216 – 226.

Rakhi Yadav has completed M. Tech. and Ph.D. from MANIT Bhopal. Her area of interest is smart grid, electric distribution system, and advanced metering infrastructure.

Yogendra Kumar has completed M. Tech. from MANIT, Bhopal, and Ph.D. from IIT Roorkee. He is a professor in Electrical Engineering Department, MANIT, Bhopal (M.P.). He also has been worked as HOD and dean (S&W). His area of interest is smart grid, electrical distribution system, neural network, fuzzy logic. He has published many research papers in reputed journals like IEEE Transactions etc.