International Journal of Engineering Works
https://www.ijew.io/
ISSN-p: 2521-2419
ISSN-e: 2409-2770
Vol. 9, Issue 02, PP. 22-27, February 2022
https://doi.org/10.34259/ijew.22.9022227

A Systematic Review on Supervised Learning Techniques in Electricity Theft Detection

Sohrab Alam[1], Majid Ashraf [2], Salman Alam[3]
[1,2] Electrical Engineering Department, University of Eneginnering and Technology, Peshawar
[3]Computer Science Department, COMSATS University, Abbottabad
[1]sohrab.alamuet@gmail.com
Received: 15 January, Revised: 28 January, Accepted: 04 February

***Abstract*—** In smart grids, the term non-technical losses impose challenges to perform classification, optimization, data analytics, and regression analysis in almost all areas of real-world research. The primary raw data suffers from an un-uniform distribution of one class over the other class in the case of machine learning(ML) and Deep learning(DL). The data imbalanced, Overfitting, high False positive rate, handling of high dimensional data, and generalization error impose challenges for the industries and academia to detect the thieves of electricity efficiently. The aim of this article also to present a comparative analysis of the approaches from the reference of data preprocessing, algorithms, models, and hybrid paradigms for the coeval imbalance data analysis, overfitting, generalization error, and high False Positive rate and the comparative study of different supervised techniques and its application area.

***Keywords*—** Data imbalance, overfitting, Non Technical losses, overfitting.

I. INTRODUCTION

The power losses have mainly categorized into two categories. i.e. Technical Losses(TL) and Non-Technical Losses(NTL)[1]. The TL is primarily due to internal parts of electric devices, which are reducible but not completely removable. To reduce the former one many researchers work on modifying the semiconductors, transistors, transformers, etc. to reduce the Technical losses. The main and high level of loss is Non-Technical losses, which are mainly occurred due to unlawful connection to main wire, hacking, and reading tempering of meters, the infrastructure is shown in Figure 1. In smart grids, the term non-technical losses are also called the amount of energy that is consumed but not billed [2]. The NTL has badly damaged the economy of a country like Pakistan has faced 0.89 billion US dollar losses yearly, India 16.5 billion US dollars, and Brazil 3 billion US dollars losses [3]. In the time of traditional grid physical inspections have been performed to detect electricity theft(ET), In which extra manpower was needed to detect the abnormal users of electricity [4]. Later, the traditional grid was modified into a smart grid. A smart grid is the current shape of a traditional grid that provides manner conversation among the powerhouse and consumption side and additionally it much improve the security and safeness.Smart grid is a complex,high-speed along side cyber security

TABLE I. NTL LOSSES

Country	Losses/year
USA	$6 billion
Pakistan	$0.89 billion
UK	$0.23 billion
Brazil	$ 10.50 billion
India	$16.20 billion

that's why smart grid is much safer and stable than traditional grid. [5]. Nowadays smart meters are used to record the data of consumed electricity, which allows the exchange of data between companies and consumers[6]. Some advancement in the smart grids shows that data analysis on smart grid data helps to detect abnormal users of electricity. In [7] the author shows the ETD has been categorized into Hardware and Data-driven based solutions, the former one is expensive, time-consuming, and very hard to maintain and the latter one is promising to defined electricity theft. Many approaches such as Support Vector Machine(SVM) and Multilayer Perceptron(MLP), but it does not accurately detect electricity theft [8]. SVM is only promising for low dimensional data on the other side MLP is not applicable due to overfitting and also not accurately measured local minima. So a very promising approach is needed to detect electricity theft. While much Deep Learning(DL) techniques have been analyzed such as Long Short Term Memory(LSTM) [9]. However, LSTM needs high memory and time for computation. In [10] author has proposed an approach Elephant herd optimization(EHO) based Convolutional neural network(CNN, which solves many issues like Data imbalanced, overfitting, Generalization error, Weak Classification, and no reliable evaluation matrices[11]. Electricity theft may be in the shape of trickery which consists of meter tampering, unlawful connections (direct hooking), destruction of electricity meters, stopping the rotating disk of electricity meters, and pretending to bill. Energy robbery instances happen with-inside the maximum areas of the world.

All most 102 nations such as Pakistan confronted the energy theft issue because of bad infrastructure, political uncertainty, extraordinarily degree corruption, low degree authorities efficiency, selections of non-technical staff, lack of accountability, regulation, and order of situations. Pakistan is one of the countries in which all-natural sources are to be had for the era of energy, however woefully there may be a loss of a long time making plans for the era of energy and detection of electricity theft. The economic system loss is about Rs 90 billion in line with 12 months because of robbery of electricity at lucrative, home and business level. As reported by the document of Northeast Group it said $89.3 billion losses took place all through the yr 2015 around the world. Because of electricity theft, there are about $16.20 billion losses of the income of India, Brazil 10.5bil US dollars losses, Russia 5.1bil US dollars losses, and Pakistan 0.89 billion US dollar losses. Table 1 encapsulate the country's financial system losses because of robbery. There are different forms of losses such as technical and non-technical losses throughout the distribution of electricity [12].

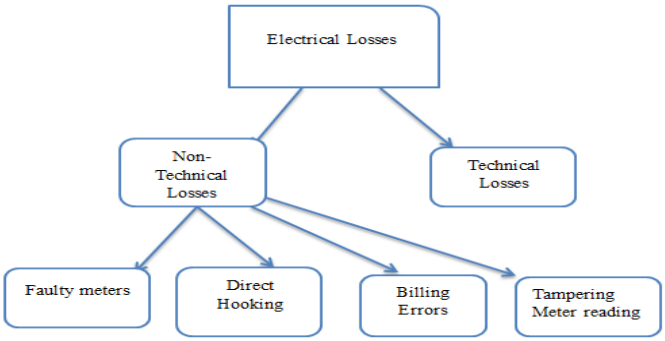

Figure 1: Types of Electricity losses

II. ELECTRICITY THEFT DETECTION

The current work done in electricity theft detection(ETD) classify further into two types of solutions i.e. hardware solution and data-driven-based solution. The hardware-based solution needs high manpower, more hardware tools, expensive and more time needed to detect electricity theft.

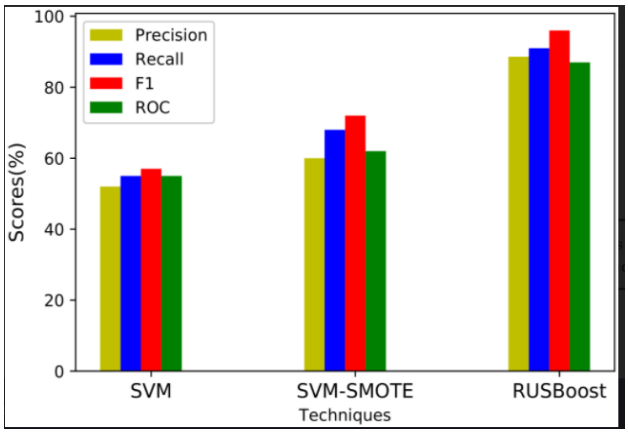

Figure 2:Performance comparison [13]

The latter one is based on Machine learning and deep learning, which draw the usage pattern based on the smart meter data to detect ET. The existing work done for electricity theft detection is mentioned in Table II. The state-of-the-art methods compared in [9],[13].

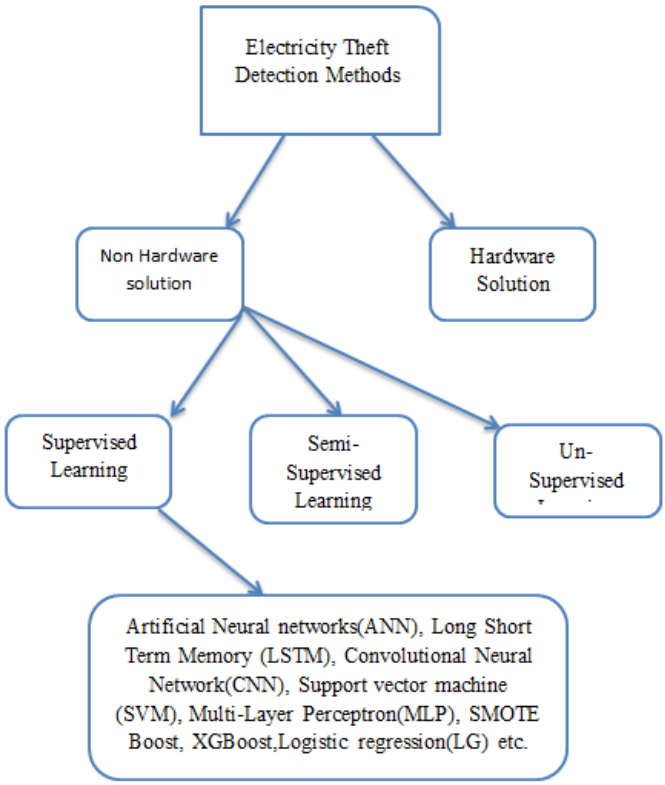

Figure 3: Electricity Theft detection Methods

Machine learning is further classified into supervised learning [14], unsupervised learning[15], and semi-supervised learning [16]. The types of electricity theft detection are shown in Figure 3. Figure 2 shows the comparative analysis of Support Vector Machine(SVM), SVM-SMOTE, and RUSBoost. The RUSBoost has better evaluation matrices.

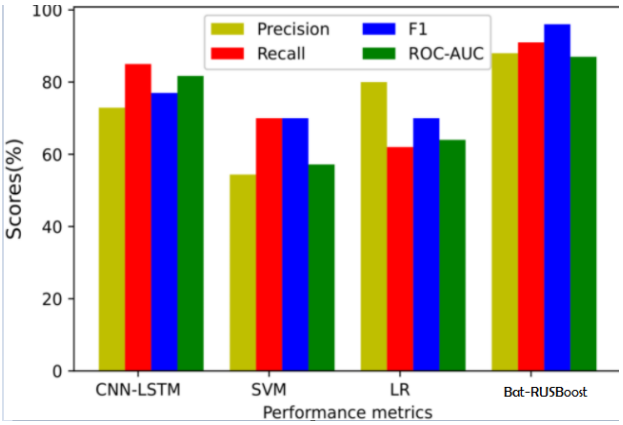

Figure 4: Performance comparison [13]

TABLE II. PERFORMANCE OF SUPERVISED MACHINE LEARNING TECHNIQUES IN THE LITERATURE. MODWPT: MAXIMUM OVERLAP PACKET TRANSFORM; ANN: ARTIFICIAL NEURAL NETWORK; LSTM: LONG SHORT TERM MEMORY; CNN: CONVOLUTIONAL NEURAL NETWORK; XGBOOST: EXTREME GRADIENT BOOSTING TECHNIQUE; SGCC: STATE GRID CORPORATION OF CHINA; SMOTE: SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE; RF: RANDOM FOREST; BBHA: BINARY BLACK HOLE

Problem Adressed	**Techniques**	**Data set**	**Contributions**	**Limitations**
To overcome the NTL in Spain[19].	LSTM and ANN	Endesa	NTL detection by Combining the ancillary and energy consumption data	Before classification the data is imbalanced
To perform the accurate prediction by selecting the refined input[20].	CNN	SGCC	Extract the local and global features from the data that represent the real theft cases	No parameter tuning
The revenue losses in Brazil due to electricity theft [21]	BBHA	National grid of Brazil	Use of binary black hole optimization technique to identify the NTL	No reliable evaluation is performed to validate the performance of the system model
The biasses of classifiers due to unbalance distribution of labels in data [22]	LSTM, Bat based RUSBoost	SGCC	LSTM extracts the relevant data and classification accuracy is enhanced by ensemble learning	The computational time of the system model is high
To accurately detect the real cases of electricity theft in Spain using the supervised learning algorithms [23]	XGBoost	Endesa Spain	The XGBoost is utilzed that operates as an ensemble and boosted the classification performance	The data pre-processing is not considered to refined the input data
The over fitting problem is addressed in neural networks [24]	CNN, SMOTE	Ireland and London grid data	The generalized performance is achieved by using the decision trees along with CNN	The SMOTE generate synthetic data, which increase the execution time of the model
The unbalanced data problem and overfitting issues are considered [25]	SOSTLink, Bidirectional GRU	SGCC	The SOSTLink is used to balance the labels in data the generalized performance is achieved by enhanced GRU	The model is complex the execution time of the model is high
To secure smart grid from the electricity theft [26]	LSTM	Smart meter data	The LSTM captures the anomalous consumption pattern and differentiate between electricity thieves and honest consumers	The unbalanced data problem is not considered
To reduce the problem of electricity theft [27]	Adaboost, ELM	China electric grid	he SMOTE is used to balance the labels in data and ensemble learning model enhance the classification accuracy	The SMOTE creates the synthetic data that cuase overfitting problem and also increases the computational complexity
The unbalanced data problem and overfitting issues are considered [28]	RUSBoost, MODWPT	SGCC	The MODWPT produce the refined input and RUSBoost method gives good results in ETD	The random under sampling technique reduce the data size and result in under fitting the mode
The unbalance data issue and overfitting issues are condsiderd[29]	LSTM, RUSBoost	SGCC	The LSTM extract features and RUSBoost method gives good results in ETD	Need large memory and computational time
Improve ROC,PR, AUC [30]	LSTM,MLP	Endesa	Integrate primary data effectively to detect non technical losses	Imbalance data

In [13] the author proposed LSTM and Bat-based RUSBoost technique compared with the state of the art technique, in which the author has shown the RUSBoost has better evaluation matrices but has high computational time and need long memory for LSTM. Recently author proposed a hybrid synthetic minority oversampling technique edited nearest neighbor (SMOTE-ENN) and Elephant herd optimization technique based convolutional neural network(EHO-CNN)[11].

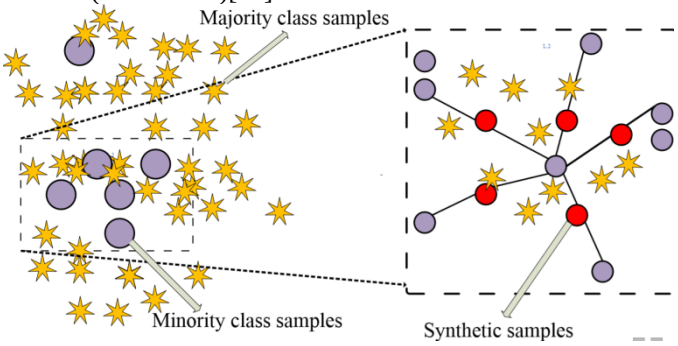

Figure 5: SMOTE-ENN analysis

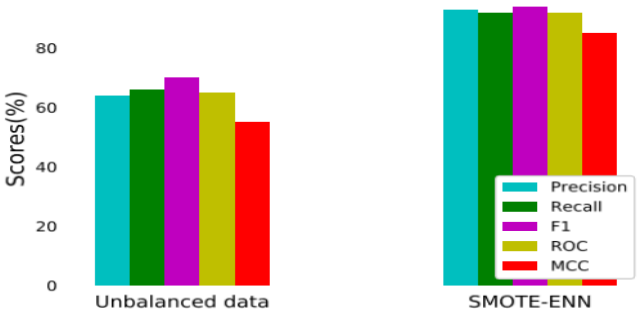

Figure 6:Performance Comparison [9]

In [9] the author proposed a technique for imbalanced data to balance, the mechanism of SMOTE-ENN has shown in Figure 5 and his best performance based on evaluation matrices has shown in Figure 6. Later for better classification the author also proposed the EHO-CNN technique which is promising compared to the state of the art.

A. Elephant Herding Optimization

Elephant Herding Optimization (EHO) is a smart, intelligent, and swarm-based metaheuristic search technique, which is developed by Wang at the end of the year 2015 [31], for clarifying the optimization issues. The design comes from the designing of the real elephant's grazing behavior in the wild. The behavior of herding is summarized as follows: Schools of elephants consist of several subgroups, so-called clans are made up of the number of Calves and females [32]. Each clan is moved under the guidance of a matriarch (adult female elephant) [31]. EHO model based on the elephant herding behaviors, which is used in operations:

- The Clan update, By which the elephant's cutting-edge positions in every clan are revising. In every extended family, the lady elephants live beneath the management of the Adult Female called matriarch, and the placement of different extended family elephants are stimulated via way of means of the matriarch place positions, in a manner that during ci extended family the placement of k elephant is revising the use of Eq. 1 [33].

$$Xnew, ci, k = Xci, k + \alpha(Xbest, ci - Xci, k)r \quad (1)$$

Where Xci, k represents the old position while the Xnew, ci, k is used for a brand new up to date position for k elephant in each ci clan, α represents a scale operator ϵ [0, 1] for figuring out the outcomes of matriarch ci on Xci, k. Xbest, ci represents the matriarch ϵ [0, 1] which improves the elephant population's diversity withinside the next seek phase [31]. in eq: 1 the number of matriarch elephants Xbest, ci in clan ci is never changed while Xbest, ci may be revised by Eq. 2.

$$A = Xnew, ci, k = \beta Xcenter, ci \quad (2)$$

- Separation, which complements the populace range with inside the subsequent seek phase. In every clan of elephants, a male elephant(adult) leaves the colony to stay alone after it becomes mature. In optimization issues, this keeping apart procedure is known as the setting apart operator. In the EHO method, the grownup male elephant with a very bad performance separates at every generation using Eq. 3 [31, 32].

$$Xworest, ci = Xmin + (Xmax - Xmin + 1) * r \quad (3)$$

Where Xworest, ci represents the worst male elephant withinside the ci clan [31] Xmin is used for lower bounds and Xman for the Upper limits of elephant's place positions. r is a form of stochastic i.e. random probability distribution and uniform distribution [0, 1] [31, 32].

B. Convolutional neural network

Y.LeCun proposed CNN, which is a kind of deep neural network [33]. It consists of multiple stacks of hidden layers, which are convolutional, pooling, flattens, dropout, and dense layers which are shown in Figure 7. These layers operate as a sequential model. CNN is not like the conventional ML model in the circumstances of feature extrication. Along with classification, it also extracts relevant features by the convolutional layer and pooling layer [9].

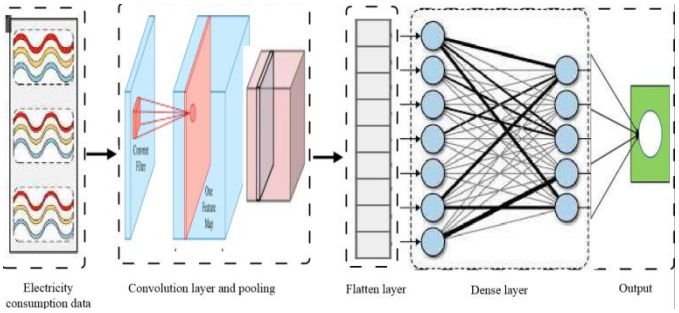

Figure 7: Convolutional neural network

III. SUMMARY

The non-technical losses have occurred due to the electricity theft. The power companies and governments of all the countries try to minimize the non-technical losses. Many researchers proposed numerous approaches to detect electricity theft more accurately but the main problem for the researcher is Primary raw data which are highly imbalanced, due to the nature of data the classifier is biased toward majority instances because the minority instances are very less in numbers i.e CNN [21]. In the existing techniques, some have limitations of poor reliable evaluation.

In [13] proposed BBHA, consider only mean accuracy which is not promising for electricity theft detection. Some approaches are only applicable for low dimensional data like XGBoost [23], Adaboost [27], and SVM [29] not for high. Besides this, some authors proposed promising approaches like LSTM and Bat-RUSBoost [13] and SMOTE-ENN and EHO-CNN [9]. In [13] the main issue of using LSTM which needs large memory and high computational time, but its performance is high based on comparative analysis. In [9] the author proposed approach has a better performance based on comparative analysis which is shown in summary table III.

TABLE III. COMPARATIVE SUMMAY

Model	Accuracy	ROC	Recall	F1-Score
SVM	0.577	0.53	0.842	0.817
LR	0.732	0.63	0.629	0.66
RUSBoost	0.82	0.875	0.831	0.83
Bat-RUSBoot	0.879	0.879	0.9109	0.961
ANN	0.75	0.768	0.77	0.82
XGBoost	0.78	0.739	0.787	0.825
CNN–LSTM	0.742	0.517	0.851	0779
EHO-CNN	0.901	0.92	0.915	0.94

CONCLUSION

This article proposed a comprehensive analysis of supervised learning challenges because of imbalanced data, overfitting, poor classification, Generalization error, Gradient vanishing, missing values, and outliers. From the comparative analysis, it has been concluded that the EHO-CNN approach has better performance compare with other existing models.

REFERENCES

[1] Buzau, M.M., Tejedor-Aguilera, J., Cruz-Romero, P. and Gomez Exposito, A., 2018. Detection of non-technical losses using smart meter data and supervised learning. IEEE Transactions on Smart Grid, 10(3), pp. 2661-2670.

[2] Jamil, A., Alghamdi, T. A., Khan, Z. A., Javaid, S., Haseeb, A., Wadud, Z., and Javaid, N. (2019). An Innovative Home Energy Management Model with Coordination among Appliances using Game Theory. Sustainability, pp. 1-23.

[3] Hussain, Z., Memon, S., Shah, R., Bhutto, Z.A. and Aljawarneh, M., 2016. Methods and Techniques of Electricity Thieving in Pakistan. Journal of Power and Energy Engineering, pp. 1-10.

[4] Jamil, A., Alghamdi, T. A., Khan, Z. A., Javaid, S., Haseeb, A., Wadud, Z., and Javaid, N. (2019). An Innovative Home Energy Management Model with Coordination among Appliances using Game Theory. Sustainability, pp. 1-23.

[5] Kaur, H., Pannu, H. S., and Malhi, A. K. (2019). A systematic review on imbalanced data challenges in machine learning: Applications and solutions. ACM Computing Surveys (CSUR), pp. 1-36.

[6] Buzau, M.M, Tejedor-Aguilera, J, Cruz-Romero, P, Gomez Exposito, 2019. Hybrid deep neural networks for detection of non-technical losses in electricity smart meters. IEEE Trans. Power System, pp. 1254–1263

[7] Sana Mujeeb and Nadeem Javaid, ESAENARX and DE-RELM: Novel Schemes for Big Data Predictive Analytics of Electricity Load and Price, Sustainable Cities and Society, pp. 2210-6707

[8] Jokar, P., Arianpoo, N., and Leung, V. C. (2015). Electricity theft detection in AMI using customers' consumption patterns. IEEE Transactions on Smart Grid, pp. 216- 226.

[9] Zheng, Z., Yang, Y., Niu, X., Dai, H. N., and Zhou, Y. (2017). Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids. IEEE Transactions, pp. 1606–1615.

[10] Alam S, Ashraf M, Alam S,Khan M. A hybrid SMOTE-ENN and EHO based CNN for electricity theft detection. Applied Sciences.

[11] Leite, J. B., and Mantovani, J. R. S. (2016). Detecting and locating non-technical losses in modern distribution networks. IEEE Transactions on Smart Grid, pp. 1023- 1032.

[12] Glauner, P., Meira, J. A., Valtchev, P., State, R and Bettinger, F. (2016). The challenge of non-technical loss detection using artificial intelligence: A survey. pp. 1-16.

[13] Adil, M., Javaid, N., Qasim, U., Ullah, I., Shafiq, M., and Choi, J. G. (2020). LSTM and Bat-Based RUSBoost Approach for Electricity Theft Detection. Applied Sciences. pp. 1-21.

[14] Ramos, C. C., Rodrigues, D., de Souza, A. N., Papa, J. P. (2016). On the study of commercial losses in Brazil: a binary black hole algorithm for theft characterization. IEEE Transactions on Smart Grid, pp. 676-683

[15] Buzau, M.M., Tejedor-Aguilera, J., Cruz-Romero, P. and Gomez Exposito, A., 2018. Detection of non-technical losses using smart meter data and supervised learning. IEEE Transactions on Smart Grid, 10(3), pp. 2661-2670.

[16] Li, S.; Han, Y.; Yao, X.; Yingchen, S.; Wang, J.; Zhao, Q, 2019. Electricity Theft Detection in Power Grids with Deep Learning and Random Forests. J. Electr. Comput. Eng, pp. 1-12

[17] Gul, H., Javaid, N., Ullah, I., Qamar, A. M., Afzal, M. K., and Joshi, G. P. (2020). Detection of Non-Technical Losses using SOSTLink and Bidirectional Gated Recurrent Unit to Secure Smart Meters. Applied Sciences, pp. 1-21.

[18] Jamil, A., Alghamdi, T. A., Khan, Z. A., Javaid, S., Haseeb, A., Wadud, Z., and Javaid, N. (2019). An Innovative Home Energy Management Model with Coordination among Appliances using Game Theory. Sustainability, pp. 1-23.

[19] Glauner, P., Meira, J. A., Valtchev, P., State, R and Bettinger, F. (2016). The challenge of non-technical loss detection using artificial intelligence: A survey. pp. 1-16.

[20] Buzau, M.M, Tejedor-Aguilera, J, Cruz-Romero, P, Gomez Exposito, 2019. Hybrid deep neural networks for detection of non-technical losses in electricity smart meters. IEEE Trans. Power System, pp. 1254–1263

[21] Ramos, C. C., Rodrigues, D., de Souza, A. N., Papa, J. P. (2016). On the study of commercial losses in Brazil: a binary black hole algorithm for theft characterization. IEEE Transactions on Smart Grid, pp. 676-683.

[22] Adil, M., Javaid, N., Qasim, U., Ullah, I., Shafiq, M., and Choi, J. G. (2020). LSTM and Bat-Based RUSBoost Approach for Electricity Theft Detection. Applied Sciences. pp. 1-21.

[23] Punmiya, R., and Choe, S. (2019). Energy theft detection using gradient boosting theft detector with feature engineering, pp. 2326–2329..

[24] Li, S.; Han, Y.; Yao, X.; Yingchen, S.; Wang, J.; Zhao, Q, 2019. Electricity Theft Detection in Power Grids with Deep Learning and Random Forests. J. Electr. Comput. Eng, pp. 1-12.

[25] Gul, H., Javaid, N., Ullah, I., Qamar, A. M., Afzal, M. K., and Joshi, G. P. (2020). Detection of Non-Technical Losses using SOSTLink and Bidirectional Gated Recurrent Unit to Secure Smart Meters. Applied Sciences, pp. 1-21.

[26] Fenza, G.; Gallo, M.; Loia, V. Drift-aware methodology for anomaly detection in smart grid. IEEE Access 2019, pp. 9645–9657

[27] Qin, H., Zhou, H., and Cao, J. (2020). Imbalanced Learning Algorithm based Intelligent Abnormal Electricity Consumption Detection. Neuro computing, pp. 112-132.

[28] Avila, N.F., Figueroa, Chu, (2018). NTL detection in electric distribution systems using the maximal overlap discrete wavelet-packet transform and random under sampling boosting. IEEE Transactions on Power Systems, pp. 7171-7180.

[29] Buzau, M. M., Tejedor-Aguilera, J., Cruz-Romero, P., and Gomez-Exp ´ osito, A. ´ (2018). Detection of non-technical losses using smart meter data and supervised learning. IEEE Transactions on Smart Grid, 10(3), pp: 2661-2670.

[30] Wang, S., and Chen, H. (2019). A novel deep learning method for the classification of power quality disturbances using deep convolutional neural network. Applied energy, pp. 1126-1140.

[31] G. Wang, S. Deb and L. Coelho, "Elephant Herding Optimization", 2015 3rd International Symposium on Computational and Business Intelligence (ISCBI), 2015.

[32] Fenza, G.; Gallo, M.; Loia, V. Drift-aware methodology for anomaly detection in smart grid. IEEE Access 2019, pp. 9645–9657.

[33] Hasan, M., Toma, R. N., Nahid, A. A., Islam, M. M., Kim, J. M. (2019). Electricity Theft Detection in Smart Grid Systems: A CNN-LSTM Based Approach. Energies. pp. 1-18.

[34] Ding, L., Li, H., Hu, C., Zhang, W., and Wang, S. (2018). Alexnet Feature Extraction And Multi-Kernel Learning for Object-Oriented Classification. Int. Arch. Photogramm. Remote Sens.

Sohrab Alam (DOB 3 Apr 1993) Belongs to beautiful District of Khyber Pakhtunkhwa Pakistan named Malakand. Received his BSc Electronics Engineering 2018 from University of Engineering and Technology Peshawar Khyber Pakhtunkhwa and MS in 2021 in Electrical Engineering from Department of Electrical & Electronics University of Engineering and Technology Peshawar.

How to cite this article:

Sohrab Alam, Majid Ashraf, Salman Alam "A Systematic Review on Supervised Learning Techniques in Electricity Theft Detection", International Journal of Engineering Works, Vol. 9, Issue 02, PP. 22-27, February 2022, https://doi.org/10.34259/ijew.22.9022227.