International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

2023 IJIEMR. Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 31ST Jan 2022. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 01

DOI: 10.48047/IJIEMR/V12/ISSUE 01/85

Title **ANALYSIS OF WOMEN SAFETY IN INDIAN CITIES USING MACHINE LEARNING ON TWEET**

Volume 12, ISSUE 01, Pages: 889-896

Paper Authors
Mr.Ganti manikanta,Mr.Akshay Venishetty,
Mr.Voma Arun Kumar

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

ELECTRICITY THEFT DETECTION IN POWER GRIDS WITH DEEP LEARNING AND RANDOM FORESTS

1) Mr.Ganti manikanta,Student,Department of Information Technology,Gurunanak Institutions Technical Campus,Telangana,India. gantimanikanta7@gmail.com
2) Mr.Akshay Venishetty,Student,Department of Information Technology,Gurunanak Institutions Technical Campus,Telangana,India. venishettyakshay82@gmail.com
3) Mr.Voma Arun Kumar,Student,Department of Information Technology,Gurunanak institutions Technical Campus,Telangana,India arunkumarvoma@gmail.com

ABSTRACT: Electricity theft has a significant negative impact on power grids and lowers operational profitability, making it one of the primary causes of non-technical losses (NTLs) in distribution networks. An original mixture convolutional neural network-random forest(CNN-RF) model for programmed identification of power robbery is proposed in this review. The's model will probably assist with overhauling associations in settling the issues of deficient electrical audit and irregular power use. A convolutional neural network (CNN) is used in this model to get to know the characteristics that different between different hours of the day and different days from enormous and fluctuating splendid meter data at first. The back expansion strategy is used to invigorate network limits during the planning stage, and a dropout layer is familiar with reduce the gamble of overfitting. The random forest (RF) is then prepared utilizing the attributes that were procured to decide whether the customer has been taking power. The matrix search strategy is utilized to choose the best boundaries for making the RF in the cross breed model. To wrap things up, tests considering authentic energy use data are finished, and the outcomes show that the proposed area model beats various approaches concerning precision and adequacy.

Keywords *– convolutional neural network random forest (CNN-RF)*

1. INTRODUCTION

Energy loss during electricity transmission and distribution is a significant issue that power providers worldwide must address. Technical losses (TLs) and nontechnical losses (NTLs) are the two primary types of -e energy losses [1]. -e TL is a necessary component of the movement of electricity and is brought on by internal processes in the transformers and transmission lines of the power system [2]; The hole between complete misfortunes and TLs, which is for the most part welcomed on by energy burglary, is known as the NTL. In reality, actual attacks such line tapping, meter crushing, or meter perusing control represent most of force burglary [3]. - These kind of electrical fraud practises might cost power providers money. Consider the projected $4.5 billion in yearly misfortunes welcomed on by energy burglary in the United States (US) [4]. Furthermore, it's assessed that energy robbery costs service organizations worldwide more than $20 billion yearly [5]. Also, examples of power burglary could affect the security of the power framework. As an illustration, the high demand on electrical systems brought on by power theft may result in fires that endanger public safety. Consequently, for the wellbeing and soundness of the power framework, exact power burglary location is fundamental.

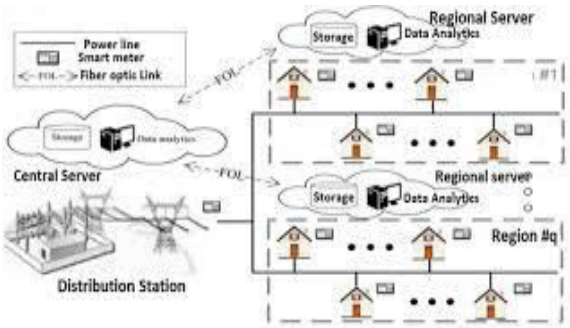

Fig.1: Example figure

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

While advanced metering infrastructure (AMI) was carried out in brilliant lattices, power utilities got tremendous volumes of energy utilization information from shrewd meters on a successive premise. This data is useful for identifying electricity theft [6, 7]. But there are two sides to every coin, and the AMI network provides a new avenue for attempts to steal power. These AMI assaults may be undertaken via a variety of tactics, including digital tools and cyberattacks. The three main methods of detecting power theft include evaluating problematic gear or equipment, reviewing unlawful line diversions, and contrasting malicious metre records with the good ones. Notwithstanding, these techniques are very tedious and costly while doing a framework's entire finished meter confirmation. Also, these manual techniques can't forestall cyberattacks. Various methodologies have been proposed lately to address the previously mentioned issues. State-based, gametheory-based, and computerized reasoning based models are the three fundamental classifications used to arrange these methods [8].

2. LITERATURE REVIEW

Electricity theft: overview, issues, prevention and a smart meter based approach to control thef:

In developing nations, non-technical loss (NTL) during the transmission of electrical energy is a significant issue, and utility companies have found it extremely challenging to identify and combat the thieves. The majority of NTL is made up of electricity theft. These losses have an impact on supply quality, add to the strain on the producing station, and change the rate charged to loyal consumers. The causes that cause customers to steal power are covered in this essay. Various strategies for distinguishing and assessing the robbery are tended to considering these adverse consequences. This article recommends a brilliant meter, outer control station, symphonious generator, and channel circuit structural plan. The goal of this effort is to discourage unlawful use while also successfully conserving and using energy. Additionally, smart metres are made to offer information on a number of aspects relating to current power use. The communicating end data of the dispersion feeder is utilized by an outer control station to ascertain NTL in the feeder. If a significant volume of NTL is found, the feeder's harmonic generator is activated to add an extra harmonic component and damage the appliances of the unlawful customers. A cost-benefit analysis for implementing the suggested method in India is provided as an example.

Technical and nontechnical losses in power system and its economic consequence in Indian economy:

India has chronic energy shortages that are becoming worse. The country's overall economic growth has been severely hampered by these shortages. Since the amount of specialized and non-specialized misfortunes in the conveyance framework, These huge misfortunes are credited to various variables, including lacking T and D limit, an unnecessary number of change steps, improper burden appropriation, far reaching rustic jolt, and so forth. The distinction between the metered units of power entering the dispersion organization and those leaving the organization, paid for through power accounts, whether assessed or metered, over a particular timeframe, could be alluded to as just misfortunes. Specialized misfortunes are the electrical framework misfortunes welcomed on by assistant supplies, flow streams, and organization impedance. Specialized misfortunes might have direct associations with network activities or speculation. Theft, unbilled accounts, approximated client accounts, botches coming about because of the estimation of utilization by unmetered supplies, and metering deficiencies are a couple of instances of non-specialized misfortunes. This' paper will probably examine specialized and non-specialized misfortunes in influence frameworks utilizing a contextual analysis and MATLAB reproduction.

A multi-sensor energy theft detection framework for advanced metering infrastructures:

The advanced metering infrastructure (AMI), which replaces conventional analog meters with digital smart meters, is an essential component of the smart

grid. Smart meters have made it conceivable to productively deal with countless end buyers, yet they have likewise made AMI an engaging objective for neighborhood and far off actual control fully intent on taking energy. Despite the fact that energy theft can be detected by smart meters thanks to a variety of sensors and data sources, various methods frequently report inaccurate results. We discuss AMIDS, a framework for AMI interruption identification that combines data from smart meter use and sensor data to more effectively distinguish energy theft. To all the more exactly portray and distinguish theft related conduct, AMIDS incorporates utilization information with meter review logs of physical and digital exercises. Our test discoveries on common and strange burden designs show the way that AMIDS can precisely recognize endeavors at energy robbery. Moreover, AMIDS appropriately recognized lawful burden profile changes and hurtful ones that were marked vindictive by less complex examinations.

Electricity theft: a comparative analysis:

Extortion (altering meters), burglary (making associations without authorization), odd charging designs, and neglected bills are instances of power robbery. Out of 102 nations, estimates of the volume of force robbery are made between 1980 and 2000. The research suggests that stealing is on the rise across the majority of the globe. Reduced energy sales revenue and the need to increase consumer prices are the financial effects of theft. Theft of electricity is highly correlated with measures of good governance; it is more prevalent in nations with weak accountability systems, unstable governments, and high levels of corruption. By using management techniques like inspection and monitoring, technical solutions like tamper-proof metres, and in certain cases reorganising the ownership and control of power infrastructure, electricity theft may be decreased.

Improving knowledge-based systems with statistical techniques, text mining, and neural networks for nontechnical loss detection:

Power distribution businesses now face a number of issues connected to energy losses. For instance, a malfunction in the customer's measuring equipment or illicit manipulation might prevent the billing of the used energy. These losses are known as non-technical losses (NTLs), and their magnitude typically surpasses that of distribution infrastructure-related losses (technical losses). There have not yet been any tests that make use of a Knowledge-Based System (KBS) that is designed with the controllers' knowledge and expertise in mind. Generally, countless investigations have utilized information mining to distinguish NTLs. A KBS was developed as part of the ongoing review that uses text mining, brain organization, and measurable techniques to identify NTLs and is based on the knowledge and expertise of the overseers. To extricate data from tests, text mining, brain organizations, and factual methodologies were used. This data was then changed over into rules, which were then joined with the principles that were delivered by the monitors' skill. Genuine examples taken from Endesa information bases were utilized to assess this framework. With in excess of 73 million clients, Endesa is a major player in both the European and South American business sectors, and it is one of the largest circulation companies in Spain.

3. METHODOLOGY

As is shown, classification and clustering models are additional categories into which existing machine learning systems may be divided. Despite the fact that the AI recognition methods introduced above are one of a kind and great, their outcomes are as yet deficient for commonsense use. For example, most of these frameworks need to physically remove highlights, which is somewhat because of their powerlessness to deal with high-layered information. To be sure, the mean, standard deviation, most extreme, and least utilization information are exemplary hand-planned components. The 2D qualities in the savvy meter information can't be separated physically since it is a difficult and tedious activity.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Disadvantages:

1. The performances are still insufficient for practise.
2. Time-consuming procedure; data from smart metres cannot be used to extract 2D characteristics.

PROPOSED SYSTEM:

The methodology of this paper aims to prioritize a list of customers who are most likely to have an abnormality in their electricity meter and to provide utilities with this list. The three main steps of the system for detecting electricity theft are depicted in Figure 2.

(i) **Data analysis and preprocess:** We look the factors that impact power clients' ways of behaving first to legitimize the utilization of a CNN for highlight extraction. We consider various assignments for the information preprocess, including information cleaning (resolving outliers), missing worth ascription, and information change.

(ii) **Generation of train and test datasets:** To determine whether the methods described in this paper are adequate, the cross-approval strategy divides the preprocessed dataset into the train dataset and the test dataset. While the train dataset is used to prepare the model's boundaries, the test dataset is used to evaluate how well the model holds up to brand-new, untested client tests. The unbalanced concept of the dataset has the potential to have a negative impact on the presentation of administered AI strategies because power theft customers are strikingly larger than nonfraudulent ones. The synthetic minority oversampling technique (SMOT) calculation is used to equalize the number of power burglaries and nonfraudulent customers in the train dataset in order to lessen this predisposition.

(iii) **Classification using the CNN-RF model:** In the proposed CNN-RF model, convolution and down examining at various times of the day and on various days are used to first help the CNN extract the highlights from massive and dynamic bright meter data. After that, RF characterization is prepared to determine whether the customer has taken power by utilizing the gathered attributes. Finally, the matrix and receiver operating characteristic (ROC) curves are used to assess the CNN-RF model's accuracy on the test dataset.

Advantages:

1. To help utility suppliers in tracking down answers for the issues of wasteful power review and capricious power utilization

2. To decide if the buyer is taking power.

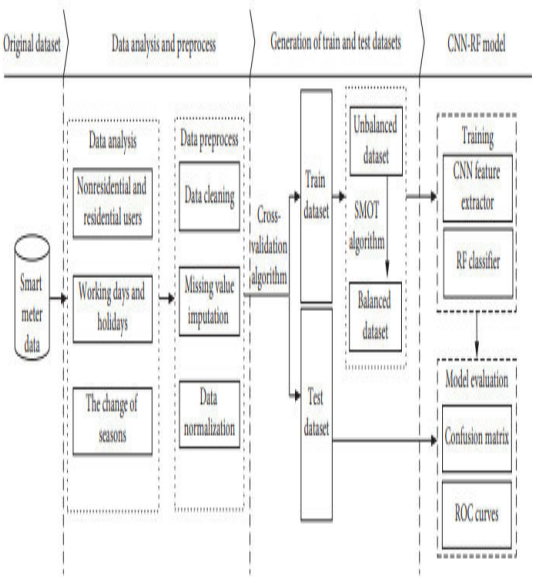

Fig.2: Flow diagram

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

MODULES:

Author of proposed paper is doing the following:

1) **Reading dataset:** utilizing this module and perusing a dataset of force utilization
2) **Preprocess dataset:** By removing missing data, this module will normalise and clean the dataset.
3) **Train CNN Model:** With the assistance of this module, we will prepare CNN on a dataset, extricate its prepared elements, and afterward input those learned highlights into the irregular timberland procedure to make a model for anticipating burglary. We constructed a DROPOUT layer to dispose of pointless highlights.
4) **Train CNN with Random Forest:** A random forest was constructed using a standard dataset without the use of CNN elements in this module, and accuracy, review, FSCORE, and exactness were then determined., and then records exactness, review, accuracy, and FSCORE.
5) **Train CNN with SVM:** This module will prepare SVM with CNN highlights, but accuracy, review, FSCORE, and precision are not completely established.
6) **Train Random Forest without CNN:** The accuracy, review, FSCORE, and exactness of the random forest were then determined using a standard dataset and without the use of CNN elements.
7) **Train SVM without CNN:** On a typical dataset, we prepared SVM without using CNN highlights in this case, and then we evaluated accuracy, review, FSCORE, and exactness.
8) **Comparison Graph:** We will show an examination chart between all calculations utilizing this.
9) **Predict Electricity Theft:** We will transfer test data with the assistance of this module, and CNN-RF will determine whether test records contain information about energy theft.

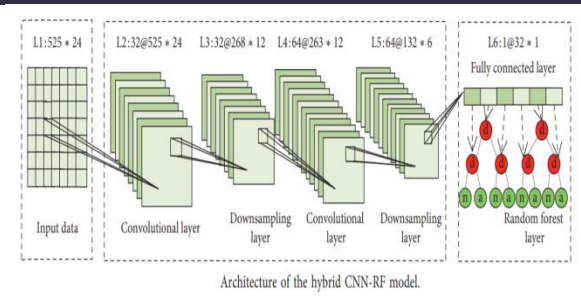

Fig.3: Architecture

4. IMPLEMENTATION

CNN:

Convolutional neural networks, also known as convnets or CNNs, are a component of machine learning. It is a subset of the various models of artificial neural networks that are utilized with various data sets and purposes. A CNN is a specific kind of organization plan for deep learning calculations that is used for errands like picture acknowledgment and pixel information handling. In spite of the fact that there are various types of brain networks in profound learning, CNNs are the favored organization plan for distinguishing and perceiving objects. They are in this way undeniably appropriate for computer vision (CV) exercises and for applications where precise item distinguishing proof is vital, such face and self-driving auto frameworks.

SVM:

Support Vector Machine, or SVM, is one of the most common managed learning calculations used to deal with Grouping and Relapse issues. Nevertheless, it is frequently used in Machine Learning Characterization problems. The SVM's computation will probably spread out the ideal line or decision limit that can segment n-layered space into classes, allowing us to quickly portray new information of interest from this point forward. This ideal choice limit is known as a hyperplane. SVM selects the wild vectors and points that guide the construction of the hyperplane. The SVM method is based on support

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

vectors, which are used to deal with these outrageous situations.

RANDOM FOREST:

The well-known machine learning algorithm Random Forest is included in the supervised learning strategy. It can be used to solve ML problems related to classification and regression. The concept of ensemble learning, in which a number of classifiers are combined to solve challenging problems and boost model performance, serves as its foundation. Random Forest is a classifier that, as the name suggests, averages a number of decision trees on various subsets of the dataset to increase the accuracy of predictions from that dataset. Instead of relying solely on a single decision tree, the random forest makes predictions based on the votes of the majority of predictions. Overfitting is prevented and accuracy is improved thanks to the forest's greater number of trees.

5. EXPERIMENTAL RESULTS

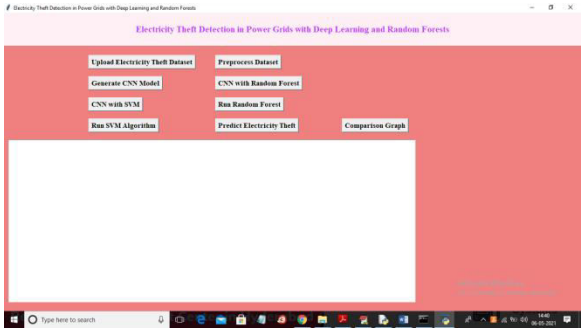

Fig.4: Home screen

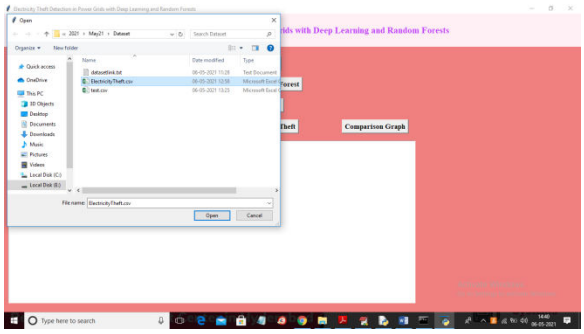

Fig.5: Upload electricity theft dataset

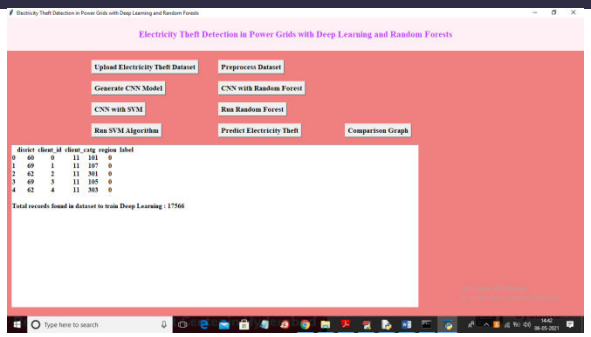

Fig.6: Preprocess dataset

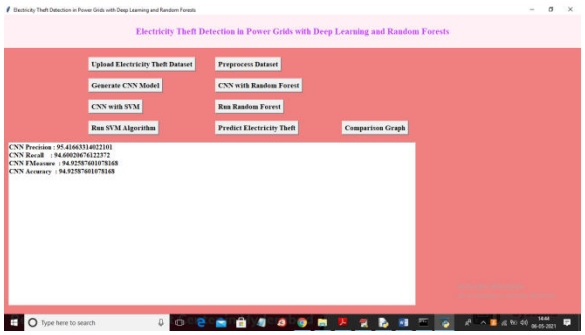

Fig.7: Generate CNN model

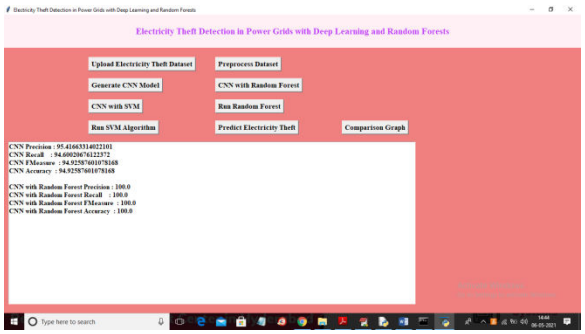

Fig.8: CNN with RF

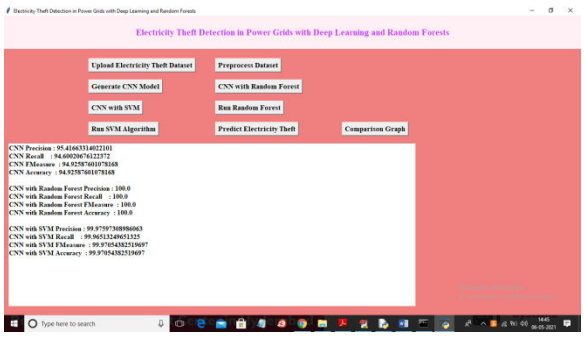

Fig.9: CNN with SVM

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

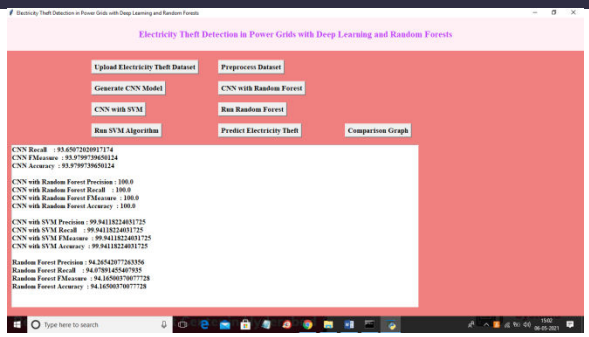

Fig.10: Random forest algorithm

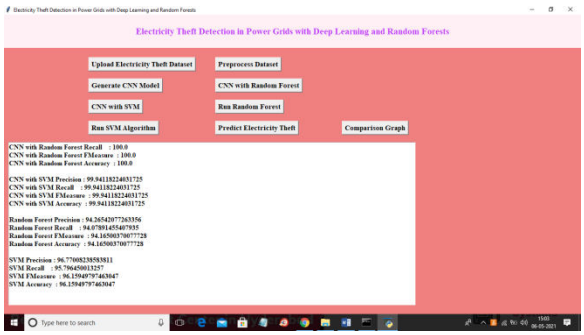

Fig.11: SVM algorithm

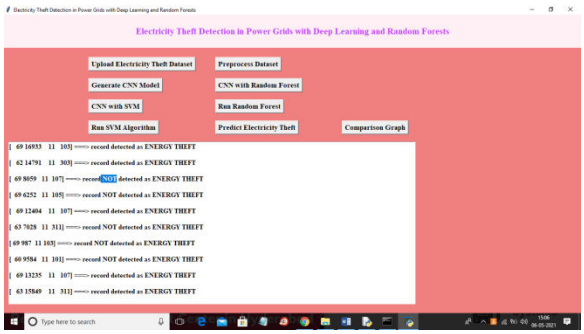

Fig.12: Test data prediction

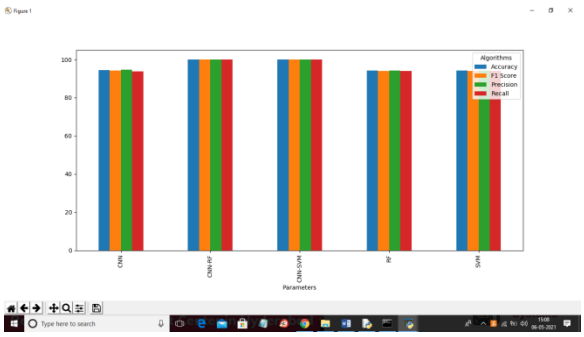

Fig.13: Comparison graph

6. CONCLUSION

This study presents a novel CNN-RF model for identifying electricity theft. When analyzing data from smart meters, this method uses the RF as the output classifier and the CNN as an automated feature extractor. A completely associated layer with a dropout pace of 0.4 is made during the preparation stage on the grounds that numerous boundaries should be upgraded, expanding the gamble of overfitting. The SMOT calculation is likewise used to resolve the issue of information lopsidedness. SVM, RF, GBDT, and LR are just a few of the deep learning and machine learning techniques whose application is compared to a similar problem. SEAI and LCL datasets were used to test these approaches. The proposed CNN-RF model is a seriously encouraging order approach in the space of force robbery location, as per the outcomes, which highlight two attributes: The recovery of great hand-planned highlights is a difficult and laborious process, but it is necessary for the majority of other traditional classifiers to advance. The mixture model, in any case, can consequently remove highlights. The second is that since CNN and RF are the two most broadly involved and viable classifiers in the field of power burglary location, the mixture model joins their benefits.

7. FUTURE WORK

Future research will focus on how the granularity and duration of smart meter data might affect consumers' right to privacy because the detection of electricity theft affects that right. Consider employing the proposed hybrid CNN-RF model, for instance, for load forecasting.

REFERENCES

[1] S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni, “Electricity theft: overview, issues, prevention and a smart meter based approach to control theft,” Energy Policy, vol. 39, no. 2, pp. 1007–1015, 2011.

[2] J. P. Navani, N. K. Sharma, and S. Sapra, "Technical and nontechnical losses in power system and its economic consequence in Indian economy," International Journal of Electronics and Computer Science Engineering, vol. 1, no. 2, pp. 757–761, 2012.

[3] S. McLaughlin, B. Holbert, A. Fawaz, R. Berthier, and S. Zonouz, "A multi-sensor energy theft detection framework for advanced metering infrastructures," IEEE Journal on Selected Areas in Communications, vol. 31, no. 7, pp. 1319–1330, 2013.

[4] P. McDaniel and S. McLaughlin, "Security and privacy challenges in the smart grid," IEEE Security & Privacy Magazine, vol. 7, no. 3, pp. 75–77, 2009.

[5] T. B. Smith, "Electricity theft: a comparative analysis," Energy Policy, vol. 32, no. 1, pp. 2067–2076, 2004.

[6] J. I. Guerrero, C. Leon, I. Monedero, F. Biscarri, and ´ J. Biscarri, "Improving knowledge-based systems with statistical techniques, text mining, and neural networks for nontechnical loss detection," Knowledge-Based Systems, vol. 71, no. 4, pp. 376–388, 2014.

[7] C. C. O. Ramos, A. N. Souza, G. Chiachia, A. X. Falcão, and J. P. Papa, "A novel algorithm for feature selection using harmony search and its application for non-technical losses detection," Computers & Electrical Engineering, vol. 37, no. 6, pp. 886–894, 2011.

[8] P. Glauner, J. A. Meira, P. Valtchev, R. State, and F. Bettinger, "-e challenge of non-technical loss detection using artificial intelligence: a surveyficial intelligence: a survey," International Journal of Computational Intelligence Systems, vol. 10, no. 1, pp. 760–775, 2017.

[9] S.-C. Huang, Y.-L. Lo, and C.-N. Lu, "Non-technical loss detection using state estimation and analysis of variance," IEEE Transactions on Power Systems, vol. 28, no. 3, pp. 2959–2966, 2013.

[10] O. Rahmati, H. R. Pourghasemi, and A. M. Melesse, "Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran region, Iran," CATENA, vol. 137, pp. 360–372, 2016.