JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Real-Time Smart Grid Identification:AI-Enabled Electricity Theft Detection

[1]Sripavan B, [2]Numair Shaikh, [3]Spandan M N, [4]Ananya Richu

[5]Elaiyaraja P

[1234]B.E Student, Department of CSE, Sir M Visvesvaraya Institute of Technology, VTU, Bengaluru, India
[5]Assistant Professor, Department of CSE, Sir M Visvesvaraya Institute of Technology ,VTU, Bengaluru, India

***Abstract* :**
Electricity theft is a global problem that negatively affects electricity companies and consumers. It destabilizes the business devel opment of energy companies, causes electricity risks and affects high energy prices for consumers. The development of smart grid s plays an important role in the exploration of electricity because they generate a lot of information, including consumer informati on, that can be used to identify electricity Theft through machine learning and deep learning. This work presents a theft detection method that uses the combined time and frequency features of a deep neural network as a classification system. We address data weaknesses such as missing data and inconsistent classes through data correlation and synthetic data generation techniques.

I. Introduction

Electricity theft is a problem affecting power companies worldwide. Nonprofit losses (NPLs) cost utilities worldwide more than $9 6 billion annually, with electricity theft being the primary cause. According to the World Bank report, 50 percent of the electricity p roduced in Saharan Africa is stolen. Usage is less than utilization. As a result, energy companies suffer huge financial losses due to electricity theft. The report stated that in 2015, India lost 16.2 billion dollars, Brazil lost 10.5 billion dollars and Russia lost 5.1 billi on dollars. It is estimated that electricity theft costs South Africa approximately US$1.31 billion in lost revenue each year. It may c ause excessive electricity, generator overload, and public safety such as electric shock. The increase in electricity prices also directl y affects all consumers. The use of smart grids takes a lot of time to solve the problem of electricity theft. Smart grids generally con sist of traditional power lines, smart meters and meters, and computing centers that monitor and control the grid, all connected by c ommunications.

Smart meters and meters collect information about energy use, grid conditions, energy costs and more. The system is expensive, ine ffective and cannot detect cyber attacks. Recently, researchers have focused on the use of machine learning classification technique s and intelligent electronic data analysis. These theft investigations have proven to be very cheap. However, existing classification methods consider time domain features and not active areas, thus limiting their performance. The main reason for the delay in solvi ng this problem is that the spread of smart projects has been completed in developing countries, while developing countries are lagg ing behind.

Challenges in the delivery of smart projects include lack of communication and users' concerns about the confidentiality of informa tion provided by smart meters. However, many developed and developing countries are reportedly considering smart meters with g oals such as solving NTL. The most efficient electronic search method is based on the distributed network (DNN) method. We sho w that using frequency domain features improves classification performance compared to using time domain features alone. We use actual electricity consumption data published by the State Grid Corporation of China (SGCC). This file contains electronic data fro m January 2014 to October 2016.

II. Research objective

The project aims to create an instant grid identification system that integrates smart intelligence energy theft function. The main go al is to create and use a powerful framework that can describe various aspects of the smart grid, such as energy consumption, distrib ution patterns, and imbalances that indicate the possibility of theft. The system will detect unauthorized access to the grid through a dvanced machine learning algorithms and data analytics and provide instant alerts to energy service providers, focusing on timely i ntervention. This study also aims to evaluate the effectiveness of the system in reducing revenue from electricity theft, increasing gr id security and improving resource allocation in power.

III. Literature survey

1. IEEE(2022) - Electricity Theft Detection in Smart Grids Based on Deep Neural Network
2. MDPI(2021) - Electricity Theft Detection in Smart Grid Systems: A CNN-LSTM Based Approach
3. Science Direct(2021) - A practical feature engineering framework for electricity theft detection in smart grids
4. Hindawi(2020) – Electricity Theft Detection in Power Grids with Deep Learning and Random Forests
5. Research Gate(2020) - Adapting Big Data Standards, Maturity Models to Smart Grid Distributed Generation: Critical Review
6. Science Direct(2022)- Pattern-based and context aware electricity theft detection in smart grid.
7. Science Direct (2020) - Detection and identification of energy theft in advanced metering infrastructures
8. The Institution of Engineering and Technology (2021) - Electricity theft detection in smart grid using random matrix theory
9. Bulletin of Electrical Engineering and Informatics (2021) - Electricity Theft Detection in Smart Grid Based on Deep Learning
10. Malaviya National Institute of Technology (2021) - Intelligent energy cyber physical systems (iECPS)

IV. Proposed system

The planning system is a combination of hardware and software designed to update the intelligent control plan and make it more eff icient against electricity. It includes advanced metering equipment (AMI) equipment equipped with IoT sensors to capture realtime data on energy consumption, grid performance, and the environment. This information is securely sent to a central cloudbased platf orm, where it is processed and analyzed by intelligent algorithms to identify patterns that indicate illegal activity, such as meters or meters. The dashboard system provides utility operators with insight and alerts, allowing them to make quick adjustments and prev ent losses. Additionally, the system uses machine learning models that constantly learn and adapt to changing theft techniques, ensu ring long-term efficiency and reliability in network management and good operation..

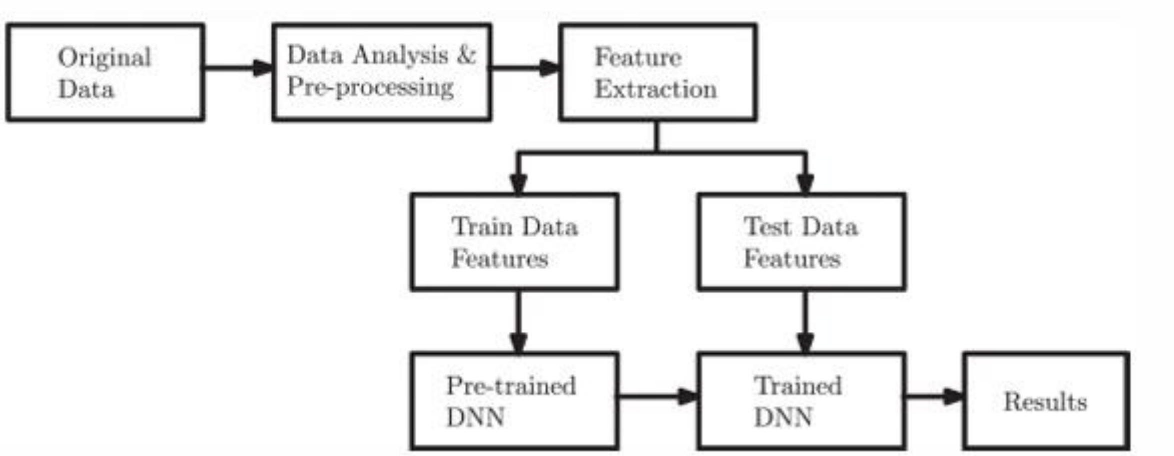


Figure 1: Methodology

1) **Data set analysis and prioritization**

We use actual power consumption data published by the State Grid Enterprise of China. This data includes daily electricity consum ption data from January 2014 to October 2016. It is equal to the total daily electricity consumption. The data used includes 42372 o bservations, of which 3615 are electronic data of unethical consumers and the remaining observations are electronic data of custom er loyalty.

2) Feature extraction The auxiliary data used in this project is non-uniform time data.

A univariate measurement is a single measurement over time. To solve the classification problem, data can be represented by its fea tures (attributes) that can be included in the classification. Classify data based on the similarity of features of datasets in different m odels. In this study, time and frequency domain features were extracted and used as input to a deep neural network for classification

3)Classification performance was compared in terms of recording time, recording frequency, and combination of two recor dings.

In order to achieve the best classification in a reasonable time, we use the Bayesian optimization method to tune the following hype rparameters: number of hidden layers, size of each layer, activation strength and function. Bayesian optimization is derived from Ba yes' theorem stated for conditions A and B.

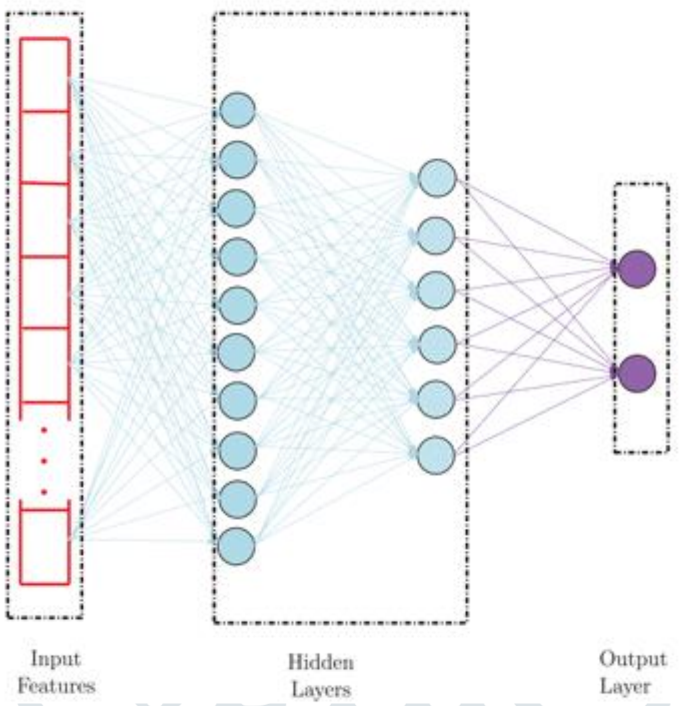
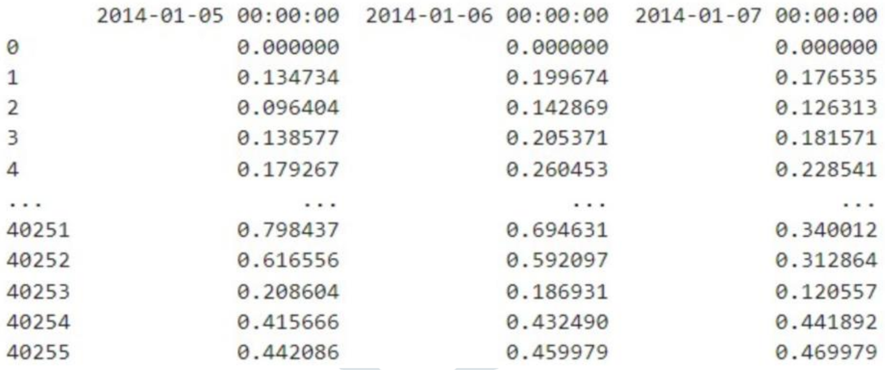


Figure 2: A typical representation of DNN

1) RESULTS AND CONCLUSION



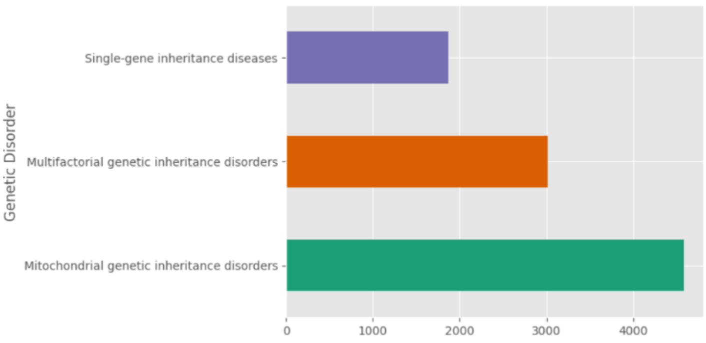
```
         2014-01-05 00:00:00  2014-01-06 00:00:00  2014-01-07 00:00:00
0                   0.000000             0.000000             0.000000
1                   0.134734             0.199674             0.176535
2                   0.096404             0.142869             0.126313
3                   0.138577             0.205371             0.181571
4                   0.179267             0.260453             0.228541
...                      ...                  ...                  ...
40251               0.798437             0.694631             0.340012
40252               0.616556             0.592097             0.312864
40253               0.208604             0.186931             0.120557
40254               0.415666             0.432490             0.441892
40255               0.442086             0.459979             0.469979
```

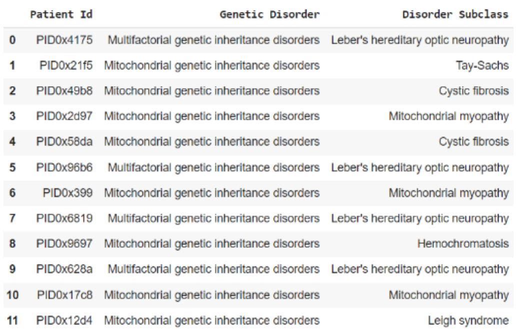
Figure 5.1: Snippet of processed dataset]

In Table 5.1,Each column provides valuable information about the patients, their medical history, and relevant clinical parameters. Analyzing these data can lead to insights into disease patterns, risk factors, and treatment outcomes. Analyzing the rich array of patient data encapsulated within each column of this dataset offers a gateway to uncovering intricate insights into disease patterns, risk factors, and treatment outcomes.

From patient demographics like age and gender to genetic predispositions inherited from both maternal and paternal sides, these data points illuminate the complex interplay between genetic and environmental factors in disease manifestation. The presence or absence of specific genes provides a molecular blueprint, shedding light on the underlying mechanisms driving inherited disorders and offering potential avenues for targeted therapies. Physiological markers such as blood cell count, respiratory rate, and heart rate offer real-time snapshots of a patient's health status, enabling clinicians to diagnose and manage conditions ranging from hematological disorders to respiratory and cardiovascular diseases.

Ultimately, this comprehensive dataset serves as a treasure trove of information for healthcare professionals and researchers alike, offering a holistic view of patient health and disease trajectories. Through meticulous analysis and interpretation, we can unlock valuable insights that drive advancements in personalized medicine, optimize healthcare delivery, and enhance patient outcomes.

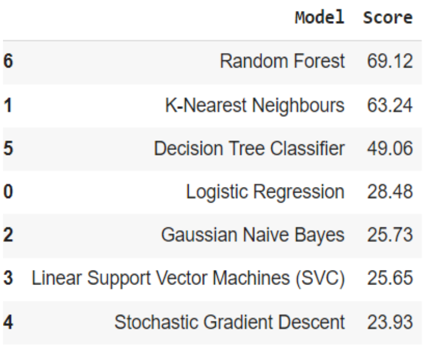




	Patient Id	Genetic Disorder	Disorder Subclass
0	PID0x4175	Multifactorial genetic inheritance disorders	Leber's hereditary optic neuropathy
1	PID0x21f5	Mitochondrial genetic inheritance disorders	Tay-Sachs
2	PID0x49b8	Mitochondrial genetic inheritance disorders	Cystic fibrosis
3	PID0x2d97	Mitochondrial genetic inheritance disorders	Mitochondrial myopathy
4	PID0x58da	Mitochondrial genetic inheritance disorders	Cystic fibrosis
5	PID0x96b6	Multifactorial genetic inheritance disorders	Leber's hereditary optic neuropathy
6	PID0x399	Mitochondrial genetic inheritance disorders	Mitochondrial myopathy
7	PID0x6819	Multifactorial genetic inheritance disorders	Leber's hereditary optic neuropathy
8	PID0x9697	Mitochondrial genetic inheritance disorders	Hemochromatosis
9	PID0x628a	Multifactorial genetic inheritance disorders	Leber's hereditary optic neuropathy
10	PID0x17c8	Mitochondrial genetic inheritance disorders	Mitochondrial myopathy
11	PID0x12d4	Mitochondrial genetic inheritance disorders	Leigh syndrome

Table 5.2: Patient Genetic Disorder Information

Table 5.2, provides valuable information about the genetic disorders present in the patient population, enabling further analysis of disease prevalence, subtype distribution, and potential associations with other patient characteristics or medical outcomes.



	Model	Score
6	Random Forest	69.12
1	K-Nearest Neighbours	63.24
5	Decision Tree Classifier	49.06
0	Logistic Regression	28.48
2	Gaussian Naive Bayes	25.73
3	Linear Support Vector Machines (SVC)	25.65
4	Stochastic Gradient Descent	23.93

Table 5.3: Performance Comparison of Machine Learning Models

The Table 5.3, presents the performance scores of different machine learning models utilized in a classification or prediction task. Each row corresponds to a specific model, identified by its name or identifier, while the accompanying score represents the model's effectiveness in making predictions. These scores, often derived from evaluation metrics such as accuracy, precision, recall, F1-score, or area under the ROC curve (AUC), offer valuable insights into the relative performance of each model. By examining these scores, stakeholders can assess the strengths and weaknesses of various machine learning approaches and identify the most suitable model for the given dataset and prediction objective. This information guides decision-making processes in selecting the optimal model for deployment in real-world applications, ultimately enhancing the efficiency and accuracy of predictive analytics systems.

In conclusion, this project has demonstrated the potential of machine learning approaches to significantly enhance the prediction, diagnosis, and management of genetic disorders. By leveraging the multi-label multi-class genomes and genetics dataset and employing advanced data analysis techniques such as Genetic Exploratory Data Analysis (GEDA) and feature engineering, we have developed a robust predictive model for genetic disorders.

The results of our study show that our model can accurately predict the likelihood of an individual having a specific genetic disorder based on their genetic markers and clinical data. The model also provides interpretable insights into the genetic factors contributing to the prediction, which can be valuable for healthcare professionals in understanding the underlying mechanisms of genetic disorders.

Furthermore, the application of the novel ETRF feature extraction technique has enriched the feature set and improved the performance of the models. By combining the strengths of Extra Trees (ET) and Random Forest (RF) algorithms, the ETRF technique has provided a more comprehensive input for the learning models, leading to better predictive capabilities. Overall, this project has demonstrated the potential of machine learning approaches to revolutionize the field of genetic disorder prediction and management. The insights gained from this study can have far-reaching implications for personalized medicine, early detection, and preventive healthcare. By continuing to refine and improve our models, we can further enhance their accuracy and effectiveness, ultimately improving patient outcomes and quality of life for individuals with genetic disorders.

REFERENCES

1) L. J. Lepolesa, S. Achari and L. Cheng, "Electricity Theft Detection in Smart Grids Based on Deep Neural Network," in IEEE Access, vol. 10, pp. 39638-39655, 2022, doi: 10.1109/ACCESS.2022.3166146.

2) https://www.linkedin.com/pulse/shocking-truth-how-electricity-theft-costing-millions-amith vijayan

3) https://www.indiacode.nic.in/showdata?actid=AC_CEN_19_22_00001_200336_1517807317930&orderno=136

4) K. Phil, MATLAB Deep Learning: With Machine Learning Neural Networks and Artificial Intelligence, Seoul, South Korea:Apress, 2017.

5) Z. Zheng, Y. Yang, X. Niu, H.-N. Dai and Y. Zhou, Electricity Theft Detection, Sep. 2021, [online] Available: https://github.com/henryRDlab/ElectricityTheftDetection.