National College of Ireland

A comparative study on optimized machine learning and deep learning models for the detection of electricity theft

MSc Research Project
MSc in Data Analytics

Oindrila Saha
Student ID: X21196061

School of Computing
National College of Ireland

Supervisor: Prof. Taimur Hafeez

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Oindrila Saha

Student ID: X21196061

Programme: MSc in Data Analytics Year: Sept 2022 - 2023

Module: Research Project

Lecturer: Prof. Taimur Hafeez

Submission Due Date: 09/11/2023

Project Title: A comparative study on optimized machine learning and deep learning models for the detection of electricity theft

Word Count: 8305 Page Count: 26

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Oindrila Saha

Date: 09/11/2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

A comparative study on optimized machine learning and deep learning models for the detection of electricity theft

Oindrila Saha
X21196061

Abstract

The act of stealing electricity is a major obstacle for utility providers on a global scale, resulting in enormous financial losses and jeopardizing the integrity of the power infrastructure. In this research, a variety of machine learning and deep learning models, combined with Particle Swarm Optimization (PSO), are utilized to conduct a comparative analysis of electricity theft detection. The main thrust of this paper is the application of PSO to improve the parameters of every model, thereby improving their predictive capabilities. This ground-breaking combination of PSO with multiple models provides significant enhancements in both precision and productivity, constituting an innovative contribution to the field. The proposed methodology includes the implementation of XGBoost with PSO, Random Forest with PSO, Decision Tree with PSO, CNN with PSO, and LSTM with PSO. This report evaluates the accuracy of each model to determine the optimal one through extensive training and testing. The classification reports offer crucial performance indicators, with a focus on accuracy, recall, precision, and F1-score. Notably, the XGBoost with PSO, Random Forest with PSO, and LSTM with PSO models stand out as the top performers, reaching an astounding accuracy of over 80%. These advanced models have exceptional ability in managing intricate, unbalanced datasets that are crucial in fraud identification. To summarize, the hybrid ML and DL techniques demonstrates great potential in improving the detection of electricity fraud. Potential future pursuits may involve delving into more optimization approaches, incorporating varied deep learning architectures, implementing real-time systems, and expanding the incorporation of broader datasets. This study establishes a foundation for novel approaches to protect power distribution systems against fraudulent actions and new threats. Utility firms can effectively utilize these models by analyzing real electricity usage data to detect instances of electricity theft and mitigate significant revenue losses in the power sector.

Keywords: Electricity Theft Detection, Hybrid Models, Particle Swarm Optimization, XGBoost with PSO, Random Forest with PSO, Decision Tree Classifier with PSO, CNN with PSO, LSTM with PSO

1. Introduction

The increasing global population and technological improvements have led to a significant surge in the need for electricity. As civilizations pursue advancement, the demand for electricity continues to rise, and this rising trend is predicted to endure, if not strengthen. Although service providers make every effort to meet the increasing demand and deliver high-

quality services, obstacles continue to arise, resulting in financial setbacks on both technical and non-technical aspects. Within this particular framework, the notion of Non-Technical Loss (NTL), driven by factors such as malfeasance and energy pilferage, has surfaced as a remarkable concern. These losses can constitute a significant proportion, anywhere from 10% to 40%, of the total energy distribution. For example, due to theft in electrical utilities, India incurs a loss of approximately $4.5 billion annually where the utilities sector in the United States of America incurs a financial loss of around $1.6 billion (Ahmad et al., 2017). In accordance with British Columbia Hydro, Canada experiences an annual financial loss of around $100. The issue of electrical fraud and illegal activities, which significantly contribute to NTL, has emerged as a critical concern that requires immediate action. Irrespective of its scope, this phenomenon is widespread on a global scale. For example, empirical investigations carried out in India have unveiled that larceny and fraudulent activities account for the squandering of over 20% of the overall electricity produced (Razavi et al., 2019). Service providers have acknowledged the necessity of addressing these losses and have engaged in various studies and projects to understand the financial consequences of non-technical losses and devise solutions to minimize them. Service providers face the issue of properly identifying both specific and general losses in this endeavour. The principal objective is to reduce the loss of revenue through the prevention of energy fraud and unauthorized usage. The process of identifying energy losses can be intricate, given that service providers might suffer unjustified losses due to false accusations. Achieving an optimal equilibrium between customer satisfaction and loss reduction is of the utmost importance. Contemporary technologies, with AI in particular, have surfaced as potent instruments in confronting these obstacles (Ismail et al., 2020). AI technologies provide advanced analytics and anomaly detection abilities that enable service providers to successfully battle NTL. An important challenge encountered by service providers is the identification of anomalies, which can be addressed by the continuous monitoring of user behaviour and consumption trends. Real-time analysis has the potential to offer significant insights into atypical patterns of energy consumption, thereby notifying service providers of possible occurrences of fraud and misuse. For the purpose of achieving reliable and precise detection of electricity theft, a total of five separate machine learning and deep learning approaches have been employed, with each model being combined using Particle Swarm Optimization (PSO). A comparative analysis was conducted as well to determine which fraud detection method is more accurate. The suggested models exploit the advantages of hybrid approaches mixed with the optimization skills of PSO. By utilizing this cutting-edge methodology, service providers can optimize their capacity to promptly identify occurrences of electrical pilferage in real-time, empowering them to promptly intervene to mitigate losses. In addition, this model can easily combine with other sophisticated technologies, such as intelligent meters, systems for managing energy usage, and solutions for intelligent billing, thereby strengthening the efforts to reduce energy waste and prevent fraudulent activities. The use of these technologies not only protects the revenue of service providers but also enables customers by providing them with information about their electricity consumption and the capacity to contribute to ecological sustainability. Fig.1 depicted the underlying notion of this research.

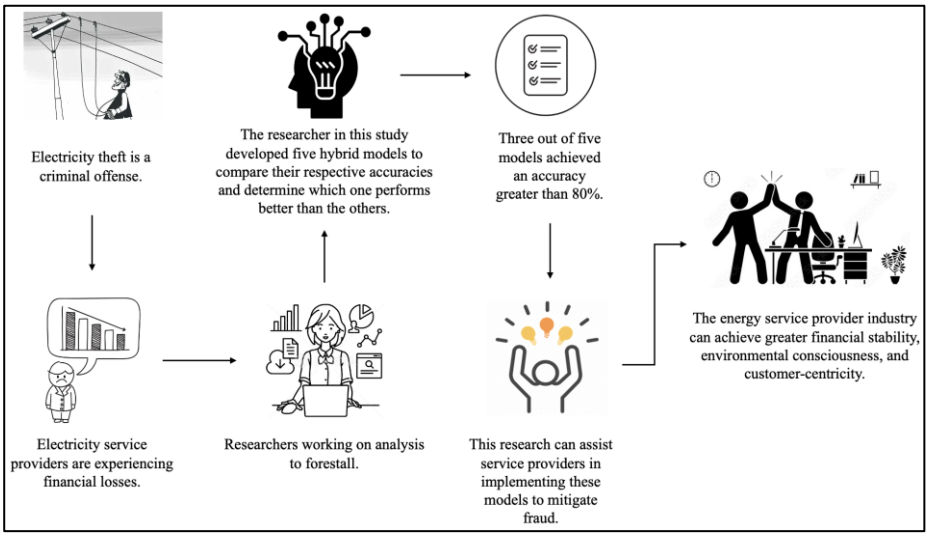


Fig.1: The underlying notion of this study

The topic-driven research questions (RQ) and research objectives (RO) are as follows:

RQ1: What is the efficacy of the proposed hybrid techniques that are integrated with PSO in detecting electricity theft?
RQ2: What is the comparative performance of the hybrid model in relation to the traditional machine learning and deep learning methods?
RQ3: What are the prospects and possibility of working with hybrid machine learning or deep learning methods?
RO1: The objective is to create hybrid ML and DL approaches integrated with particle swarm optimization to achieve accurate and efficient detection of electricity theft.
RO2: To assess and compare the effectiveness of the proposed hybrid models.
RO3: To train and evaluate the performance of the following models using PSO: XGBoost, Random Forest Classifier, Decision Tree Classifier, CNN, and LSTM.
RO4: To construct classification reports for each model, outlining crucial performance parameters.
The paper is organized in the following manner. Section 2 comprises a compilation of Related Work from several papers that have conducted research on the detection of electricity theft. The Methodology part is elaborated upon in Section 3. Design Specifications are detailed in Section 4 of the paper, whereas Implementation Steps are described in Section 5. Section 6 encompasses the Results and Evaluation of all five experiments conducted in relation to the research. Section 7 provides a comprehensive Discussion of the study's findings. Finally, the Conclusion and Future studies are addressed in Section 8.

2. Literature Review

The literature review is broken down into four sections: the first section encompasses techniques related to deep learning, the second section demonstrates approaches to machine learning, the next section introduces approaches to hybrid models and finally, the concluding section contains the approach using blockchain technology.

2.1 Approaches based on Deep Learning

(Zheng et al., 2017) presented a comprehensive and sophisticated CNN model to address the limitations of prior approaches that just relied on one-dimensional (1-D) data, thereby neglecting the cyclic patterns inherent in electricity usage. Extensive trials were conducted employing authentic electricity usage data that was made available to the public by the state government of China. Their empirical results revealed that the proposed wide and deep CNN outperformed conventional methodologies, such as linear regression, CNN, Random Forest and SVM. To mitigate the problem of overfitting, the researchers incorporated a dropout layer into the deep CNN module they developed.

Another study (Eddin et al., 2022) presented effective detectors with superior detection rates, utilizing a practical energy profile dataset encompassing the recorded injected energy over a span of three years. A significant constraint of this work was the absence of a benchmarking dataset that included both benign and attack instances for ET attacks within the generating domain. However, they had overcome this constraint by creating artificial attack samples. They chose two characteristics to improve the system's detection and handle the fluctuations in data over the different seasons. The GRU-RNN and BLSTM-RNN models were employed due to their aptitude for pattern recognition in datasets with temporal dependencies. Their proposed methodologies attained a 96.67% detection rate, surpassing the performance of the most advanced single-source model available. The results indicated that the model exhibits exceptional performance, achieving the maximum detection rate by utilizing the power generation profile of a single data source. In addition, they have developed a distinctive model, the BLSTM model, which effectively addresses the problem of electricity theft in wind and solar distributed generation units. The findings from both inquiries demonstrated that the models which are smaller perturbation-trained, exhibit more resilience in performance compared to models trained with deliberate variations in ETA.

(Lin et al., 2021) unveiled the development of an adaptive Time-Series Recurrent Neural Network (TSRNN) architecture for the purpose of detecting instances of electricity theft based on analysing time-series data of power usage. The data was continually observed from January, 2017 to March, 2019 for a total of 820 days. Through the monitoring of ARP, PEA, and SHO data, anomalous samples were assigned to users suspected of stealing electricity, while normal samples were assigned to the remaining common users. To rectify the data imbalance caused by the minority of abnormal users in the acquired data, the SMOTE technique was employed. This technique generated 222 simulated abnormal examples, resulting in a normal to abnormal ratio of approximately 3:1. The experimental results showed that the suggested adaptive TSRNN design, when paired with SMOTE, is capable of identifying aberrant electricity stealing behaviour.

2.2 Approaches based on Machine Learning

Electricity theft detection poses significant challenges due to the poor classification of imbalanced electricity consumption data, overfitting problems, especially the high false positive rate associated with present algorithms. In order to overcome the aforementioned

constraints, (Khan et al., 2020) introduced a novel framework that relies on supervised machine learning methods and actual electricity use data. A class imbalance problem is tackled using the Adasyn algorithm. It is utilized to accomplish two goals. Firstly, it strategically augments the number of samples belonging to the minority class in the dataset. Also, it serves to mitigate the model's inclination towards favouring samples from the majority class. Subsequently, the equilibrated data is inputted into a VGG-16 module to identify irregular patterns in electricity usage. Ultimately, they utilized a classification technique called Firefly Algorithm based Extreme Gradient Boosting. The researchers utilized advanced techniques, namely SVM, CNN, and Logistic Regression, for conducting comparative analysis. The validation process included the assessment of precision, recall, F1-score, MCC, ROC-AUC, and PR-AUC measures. The simulation results demonstrated that the Adasyn method, when applied to the FA-XGBoost classifier, significantly enhanced its performance, resulting in an accuracy rate exceeding 92%. Moreover, the VGG-16 module exhibited superior overall performance, achieving an accuracy rate of over 82% on both training and testing data.

By contrast, the alternative study (Hussain et al., 2021) attained a 93% accuracy rate using their proposed framework. The authors have introduced a new machine-learning framework called NGBoost, which utilizes feature engineering and natural gradient descent ensemble boosting, to detect fraud in power consumption data. Initial impute of missing data records in the smart metre dataset was performed utilizing an imputation technique based on the random forest algorithm. The second phase employed the MWMOTE algorithm, which applies a majority weighted minority oversampling strategy to address the issue of imbalanced data distribution across several classes. Ultimately, they employed the tree SHAP additive-explanations method to identify the impact (either positive or negative) of each input characteristic on predicting the target variable.

2.3 Approaches based on Hybrid Models

(Li et al., 2019) introduced a novel hybrid model, combining a convolutional neural network (CNN) with a random forest (RF), for the purpose of automatically detecting electricity theft. The suggested approach utilizes CNN as an automated feature extractor for analysing smart meter data. The Random Forest algorithm is then employed as the output classifier. In order to mitigate the risk of overfitting, they fine-tuned the parameters and implemented a fully integrated layer during the training phase with 0.4 dropout rate. Furthermore, the SMOT technique is utilized to address the issue of data imbalance. The results demonstrated that the CNN-RF model, as presented, is a highly promising classification technique in the field of electricity theft detection due to two distinct characteristics: Firstly, the hybrid model has the advantage of automatically extracting features, unlike other traditional classifiers that heavily depend on manually designing features, which is a tedious and time-consuming process. The second aspect is to the hybrid model's ability to amalgamate the merits of the RF and CNN, which are widely recognized and effective classifiers in the domain of electricity theft detection.

(Bian et al., 2021) proposed a model called PSO-Attention-LSTM, which is designed to detect and assess anomalous electricity consumption behaviour exhibited by consumers. The model combined the Particle Swarm Optimization (PSO) algorithm with Long-Short Term Memory

(LSTM) and incorporates an attention mechanism. By employing the attention mechanism to assign varying weights to the hidden state of LSTM, the model effectively mitigated the loss of historical data, fortified critical information, and omitted superfluous data. The PSO optimizer is employed to address the challenging task of model parameter choice, enhancing the model's performance by optimizing the hyperparameters. Subsequently, they constructed a model leveraging the TensorFlow framework, utilizing historical power usage data and relevant weather attributes.

(Bitirgen & Filik, 2022) also presented the technique for enhancing the performance of CNN-LSTM using Particle Swarm Optimization to identify False Data Injection Attacks in the Smart Grid system. The implemented model utilized Phasor Measurement Unit (PMU) (Wallace & -T. Lu, 2016) data to identify an atypical measurement value for the purpose of classifying the anomaly kind. The researchers conducted a comprehensive quantitative analysis by employing advanced DL model architectures such as CNN-LSTM, PSO-LSTM, and LSTM methods to validate the accuracy and efficacy of the presented model. The findings demonstrated that the model surpassed other deep learning models in terms of performance. Furthermore, the model exhibited a commendable level of precision, offering valuable guidance for ensuring the reliable and secure functioning of smart grid infrastructure.

The authors (Almazroi & Ayub, 2021) have utilized the CNN-LSTM methodology to detect instances of electric fraud. Black Widow Optimization (BWO) and Blue Monkey Optimization (BMO) are two meta-heuristic approaches that are employed to determine the most optimal values for the hyperparameters of CNN-LSTM. Classifier training was improved as a result of the optimization of hyperparameters. Following rigorous simulation, they proposed the CNN-LSTM-BWO and the CNN-LSTM-BMO methodologies which attained an accuracy exceeding 90%. Comparatively, the proposed approaches outperformed every existing scheme. The model performance had achieved a notable degree of precision and a minimal rate of errors. In addition, the statistical analysis demonstrated that the proposed methodologies were superior.

Likewise, (Ullah et al., 2020) introduced a Hybrid Deep Neural Network, referred to as CNN-GRU-PSO HDNN, which combines the CNN, Gated Recurrent Unit (GRU), and PSO. The researchers employed CNN to conduct feature selection and extraction, resulting in a reduction of both dataset dimensionality and redundancy. Moreover, the categorization of the given data into genuine and counterfeit consumers is accomplished by the utilization of the GRU-PSO technique. Subsequently, the model's performance is evaluated by comparing it to many benchmark techniques, including Logistic Regression, SVM, LSTM, and GRU. The accuracy of the suggested model is verified by assessing its performance using multiple performance metrics such as AUC, precision, accuracy, recall, and F1-Score. The simulation results demonstrated that the suggested model surpassed the existing strategies in addressing ETD and class unbalanced concerns. Additionally, the suggested model exhibits greater resilience and precision compared to the current approaches.

(Xia et al., 2022) proposed a method for detecting electricity theft using a diverse ensemble learning approach. The system incorporates a stacking integrated structure that combines several powerful individual learners. In order to select the heterogeneous powerful classifier combination of LSTM, Light gradient boosting machine, and KNN for the foundational model layer of the stacking structure, the researchers initially utilized grey relation analysis (GRA). This decision was made on the criteria of the highest comprehensive evaluation index value.

In addition, the SVM model that yielded satisfactory results, chosen as the model for the meta-model layer. A hybrid multifaceted learning model is thus developed to detect electrical theft in the stacking structure. The findings indicated that the proposed method exhibits superior detection performance compared to the standard state-of-the-art detection approach.

On the contrary, an alternative scholarly article (Pereira et al., 2016) presented Social-Spider Optimization (SSO), an evolutionary algorithm aimed at addressing the issue of tuning SVM parameters. The researchers evaluated the resilience and effectiveness of SSO in contrast to PSO and NGHS (Novel Global Harmony Search) (Zou et al., 2010), two widely employed optimization techniques. Additionally, this SSO optimizer has the potential to be applied to the feature selection process. The authors asserted that the SSO-SVM technique, which they have proposed, is being assessed for the first time on publicly available datasets in order to detect larceny in power distribution systems.

However, (Miao et al., 2022) developed an electricity theft detection model incorporating an SVM-based approach and an LSTM-Stacked Auto Encoder (LSTM-SAE). With its effective feature extraction capabilities for highly dimensional non-linear time series, LSTM-SAE is a viable approach for extracting the latent characteristics of electricity usage data. The researchers employed SVM as a classifier, with its parameters adjusted using the PSO algorithm. The LSTM-SAE extracted features and used an SVM classifier to get the results. The efficacy and precision of the proposed method are validated by calculating the false positive rate (FPR) and detection rate (DR) via experimental simulation of real power grid data.

Motivated by the robust feature extraction and data reconstruction capabilities of autoencoders, (Huang & Xu, 2021) discussed the development of a stacked sparse denoising autoencoder designed for the purpose of electricity fraud identification. In order to enhance the capability of feature extraction and ensure robustness, the autoencoder incorporated sparsity and noise. To optimize these hyper-parameters, the particle swarm optimization algorithm is utilized.

(Gu et al., 2022) introduced a novel approach called Low False Positive Rate -Deep Neural Network (LFPR-DNN), which use deep learning techniques to identify instances of electricity theft. According to the authors, this is the inaugural method for detecting electricity theft that is trained using False Positive Rate as the optimisation objective. The offered two-phase training technique efficiently decreased the false positive rate. They have demonstrated that by modifying the parameters of the target function during the second phase of training, it is possible to achieve adaptable performance that can meet various specific requirements. The empirical findings on the Irish dataset demonstrated that their approach may proficiently address data imbalance. In addition, when compared to other advanced classifiers, the LFPR-DNN model demonstrated superior performance, surpassing even powerful ensemble classifiers like extreme gradient boosting. The sensitivity analysis findings demonstrate that their method has a strong resistance to changes in the sample rate.

(Banga et al., 2021) have experimented with the ensemble techniques along with the machine learning models. The data imbalance issue was effectively resolved through the implementation of six data balancing techniques: Adaptive Synthetic Sampling (ADASYN), Random over Sampler, SMOTEENN (Edited Nearest Neighbour), Synthetic Minority over Sampling (SMOTE), Support Vector Machine-Synthetic Minority over Sampling, and SMOTE Tomek Links. Their model was composed of two steps. At first, they applied twelve

classification algorithms (Support Vector Machine, Multi-Layer Perceptron, Logistic Regression, Adaboost, Light GBM, K-Nearest Neighbour, Decision Tree, Extra Tree, Bagging, Random Forest, XGBoost, and Naïve Bayes) to the balanced data. During the subsequent phase, the two ensemble techniques, maximum stacking and voting, are implemented on the top five algorithms that exhibit the highest performance. Upon contrasting the evaluation outcomes of each model, it was determined that the SMOTEENN with stacking ensemble technique yields the most optimal performance, as evidenced by its accuracy value exceeding 96%. Furthermore, the efficacy of the model is verified through the implementation of the one-way statistical test ANOVA (Analysis of variance).

(Asif et al., 2021) introduced a novel hybrid deep learning model called Alexnet-Adaboost-ABC for the purpose of detecting instances of electricity theft. The model specifically addresses the issues of overfitting, class imbalance, and generalization problems. Their model consistently surpassed existing methodologies, offering a resilient and potent mechanism for detecting electricity theft in power systems.

(Javaid et al., 2023) proposed a novel Electricity Theft Detection model called SSA-GCAE-CSLSTM, which effectively addressed the issues of class imbalance and curse of dimensionality. The model combines the salp swarm algorithm (SSA), gate convolutional autoencoder (GCAE), and cost-sensitive learning and long short-term memory (CSLSTM). In addition, they developed a hybrid GCAE model by combining a gated recurrent unit with a convolutional autoencoder. The implied model consists of five submodules: electricity theft categorization, data balancing, data preparation, hyperparameters' optimization, and dimensionality reduction. The proposed model is evaluated against two fundamental models, GCAE–CSLSTM and CSLSTM, as well as seven benchmark models: Extreme Gradient Boosting, SVM, Decision Tree, Adaptive Boosting, Extra Trees, Random Forest, and CNN. The findings demonstrated that the SSA–GCAE–CSLSTM model achieved a precision rate above 99% with an accuracy of 92.25%, outperforming the other methods in terms of ETD.

(Akram et al., 2021) devised a novel and inventive model called CNN with RUS Boost Bird Swarm Algorithm (Rus-BSA) as well as RUS Boost Manta Ray Foraging Optimization (Rus-MRFO) model to detect the electricity fraud. The models RUS-MRFO and RUS-BSA, exhibited accuracy rates of 91.5% and 93.5%, respectively. The proposed methodologies have demonstrated encouraging outcomes, and the authors asserted that it possesses significant potential for future application.

2.4 Approaches based on Other Technology

Muzumdar et al., 2022) proposed a system for detecting energy theft in smart grid NAN using blockchain technology, while also ensured the privacy of energy use data. Unlike the current machine learning and state-based methods, this strategy is computationally viable because it does not necessitate model training, previous proof of energy theft, or any extra resources. Moreover, the incorporation of blockchain technology in the presented system provides a decentralized and reliable environment for processing energy consumption data, effectively eliminating the risk of a single source of failure. The implied system is verified utilizing a test-bed at NIT Goa, specifically examining the test scenario of power fraud and malfunctioning smart meters. It attained an accuracy rate of over 98% in detecting energy theft, while

maintaining a satisfactory throughput value & an acceptable latency value for energy data transactions. Additionally, it successfully met the power theft detection criteria in the smart grid NAN.

2.5 Synopsis of the Literature Review

Approaches	References	Objectives	Applied Methods	Dataset Details	Evaluation Metrics	Outcome
Deep Learning	Eddin et al. (2022)	Detect electricity theft attacks on smart meters	Fine-Tuned RNN-Based Detector	A real smart meter dataset from Ontario, Canada	They used precision, accuracy, f1-score for the evaluation	Models trained with lower perturbations exhibit more robustness compared to models trained with reasonable variations in ETA, a measure of malevolent behaviour.
	Lin et al. (2021)	To identify instances of illicit electricity appropriation in time-series data of power usage.	TSRNN, SMOTE algorithm	Time-series data of electricity consumption	The metrics employed in this context include the classification accuracy (ACC), false alarm rate (FAR), and true positive rate (TPR).	The suggested model is capable of extracting data features to monitor aberrant electricity theft actions.
	Zheng et al. (2017)	To address the limitations of prior approaches that just relied on one-dimensional (1-D) data	Deep CNN model, Dropout Layer	Official data from the State Grid Corporation of China (SGCC) regarding authentic electricity consumption.	The evaluation leveraged the AUC and MAP metrics.	Surpassed other established methodologies
Machine Learning	Khan et al. (2020)	To rectify the improper categorization of imbalanced data, as well as the overfitting problems and the high false positive rate (FPR) associated with current methods.	Firefly Algorithm based Extreme Gradient Boosting (FA-XGBoost)	Official data from the State Grid Corporation of China (SGCC) regarding authentic electricity consumption.	The measures employed include precision, recall, F1-score, MCC, ROC-AUC, and PR-AUC.	Accurately identified electrical thieves. In addition, the model handles big time series data and classifies better than other advanced models.
	Hussain et al., 2021	To identify fraudulent activity in power usage data.	Novel feature-engineered NGBoost machine-learning model	Official data from the State Grid Corporation of China (SGCC) regarding authentic electricity consumption.	Accuracy, recall & precision are used	Confirmed the effectiveness and importance of the models in the researched field.
Hybrid Models	Bian et al. (2021)	Targeting users' atypical electricity use patterns	PSO-Attention-LSTM	The public data set of the University of Massachusetts	RMSE, MAE, MAPE and AE, Confusion matrix, PR and FPR are applied here.	Has enhanced capability in detecting anomalies
	Li et al. (2019)	To address the issues of ineffective electricity inspection and inconsistent power use	CNN-RF	The dataset name is Electric Ireland/SEAI Smart Meter Data	ROC curve, AUC, TPR, FPR, precision, F1-score, recall are used	It is a highly prospective classification approach in the realm of electricity fraud detection.
	Bitirgen and Filik (2023)	Detect false data injection attacks	Optimized CNN-LSTM using PSO	Phasor measurement unit data (PMU) dataset	Binary class, three-class, and multi-class classification, accuracy	Effective detection of FDIA
Hybrid Models	Almazroi and Ayub, (2021)	To detect instances of electrical fraud	CNN-LSTM-BMO and CNN-LSTM-BWO	Official data from the State Grid Corporation of China (SGCC) regarding authentic electricity consumption.	F-score, accuracy, recall, precision, RMSE, MSE, and MAPE are utilized here	Exceed all the current techniques being compared
	Ullah et al. (2020)	Distinguish between honest and fraudulent customers	CNN-GRU-PSO HDNN	Official data from the State Grid Corporation of China (SGCC) regarding authentic electricity consumption.	AUC, precision, accuracy, recall and F1-Score are used	Show that the proposed model outperforms the existing techniques in terms of ETD and class imbalanced issues. Moreover, the proposed model is also more robust and accurate than the existing methods.
	Xia et al. (2022)	Employ automated methods to identify and classify possibly questionable consumers.	Combination of heterogeneous strong classifiers LG, LSTM, and KNN	Official data from the State Grid Corporation of China (SGCC) regarding authentic electricity consumption.	AUC value is used	Promising solution for more accurate theft detection
	Miao et al., (2022)	Devised a model for identifying electricity theft	LSTM-SAE and SVM-PSO	Official data from the State Grid Corporation of China (SGCC) regarding authentic electricity consumption.	FPR and DR	Potential to enhance accuracy and efficiency
	Pereira et al. (2016)	In order to alleviate the substantial computational load associated with the SVM training process	Social-Spider Optimization-based Support Vector Machine	Two proprietary datasets provided by a Brazilian electrical power utility.	Accuracy, Feature selection and parameter tuning.	SSO has achieved really encouraging outcomes in various applications.
	Huang and Xu (2021)	Robust theft detection leveraging honest user data	Stacked sparse denoising autoencoder and PSO	Residential electricity consumption of Fujian, China.	Stacked Sparse Denoising Autoencoder (SSDAE), ROC and Optimal Error Threshold (OET) are used.	Robust solution for electricity theft identification
	Gu et al. (2022)	To lower FPR of sophisticated metering infrastructure-based electricity theft detecting methods	LFPR-DNN and PSO	Open Irish data set	FPR, TPR, AUC, Bayesian Detection Rate (BDR) are used here	Can efficiently address data imbalance and outperformed other traditional methods
	Asif et al. (2021)	To detect power system electricity theft caused by Non-Technical Losses	Alexnet-Adaboost-ABC	Official data from the State Grid Corporation of China (SGCC) regarding authentic electricity consumption.	Precision, Recall, F1-score, MCC and AUC are applied here.	Demonstrated that the proposed model surpasses the existing methodologies.

Approaches	References	Objectives	Applied Methods	Dataset Details	Evaluation Metrics	Outcome
Hybrid Models	Banga et al. (2021)	Deal with data imbalance in consumption records	Ensemble methods - 12 classification models then 2 ensemble techniques based on best five performing algos.	Official data from the State Grid Corporation of China (SGCC) regarding authentic electricity consumption.	Accuracy, MCC, f1-Score, log-loss, and ANOVA	Achieved higher performance rate.
	Javaid et al., 2023	Address energy theft in smart grids	SSA–GCAE–CSLSTM	Official data from the State Grid Corporation of China (SGCC) regarding authentic electricity consumption.	Accuracy, precision, AUC, F1-score, recall are used	Performed effective energy theft detection
	Akram et al. (2021)	Develop novel models for theft detection	RUS-MRFO and rus-BSA	NYISO	F1-score, recall, precision and accuracy.	Demonstrated encouraging results and possess substantial potential for future application.
Others	Muzumdar et al. (2022)	Accurately detecting energy theft while safeguarding the privacy of consumers	Blockchain-enabled system for detecting energy theft	Smart grid NAN data	Accuracy, throughput, affordable latency are used	Incorporated blockchain for theft detection and privacy

Table 1: Synopsis of the literature review

3. Methodology

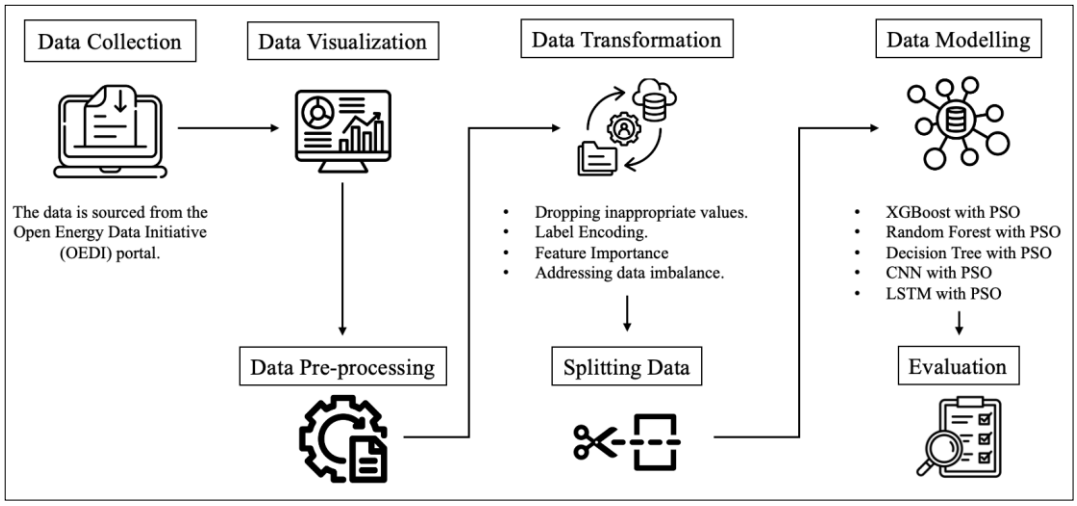


Fig.2: Research Methodology

In order to accomplish the objective of the study, a specific approach (Fig.2) was employed. The process involves gathering data from a reliable and publicly available source that contains sufficient information for analysis. Subsequently, the data is visualized to extract meaningful insights and enhance comprehension. Following the data pre-processing step to cleanse and prepare the data for subsequent investigation, the next step involves data transformation. This includes removing improper values, converting categorical input into numerical format, extracting relevant features, and addressing any data imbalance. After that, the data is divided into features and target variables. Then, the optimized machine learning and deep learning models are constructed and their performances are compared. Finally, the models are evaluated using assessment metrics to determine which models are ideal for the chosen topic.

The dataset is obtained from the Mendeley Data (Zidi et al., 2022) which is derived from the Open Energy Data Initiative (OEDI) portal. The collection encompasses energy consumption statistics for 16 distinct categories of users. It comprises many power consumption estimates for various consumers over a span of one year. Six distinct categories of fraudulent activities are incorporated into the existing dataset. They encompass several forms of theft that certain

consumers can perpetrate. The initial form of theft involves a significant decrease in electricity usage during daylight hours. The reduction is determined by multiplying the usage by a randomly selected amount ranging from 0.1 to 0.8. In the second form of theft, electricity usage abruptly and unpredictably reduces to zero for an unspecified duration. The third kind of stealing bears resemblance to the initial form, with the distinction that each value of consumption is multiplied with a randomized factor. An arbitrary fraction of the average consumption is created for the fourth category of stealing. The fifth category presents the average usage, while the final category of stealing reverses the sequence of readings. The data owners created a fraud generator that allowed them to randomly produce these six sorts of theft.

The data visualization area offers significant insights into the dataset by presenting meaningful graphical representations of the distribution of target classes, the count of values in the class columns, and the distribution of power facilities.

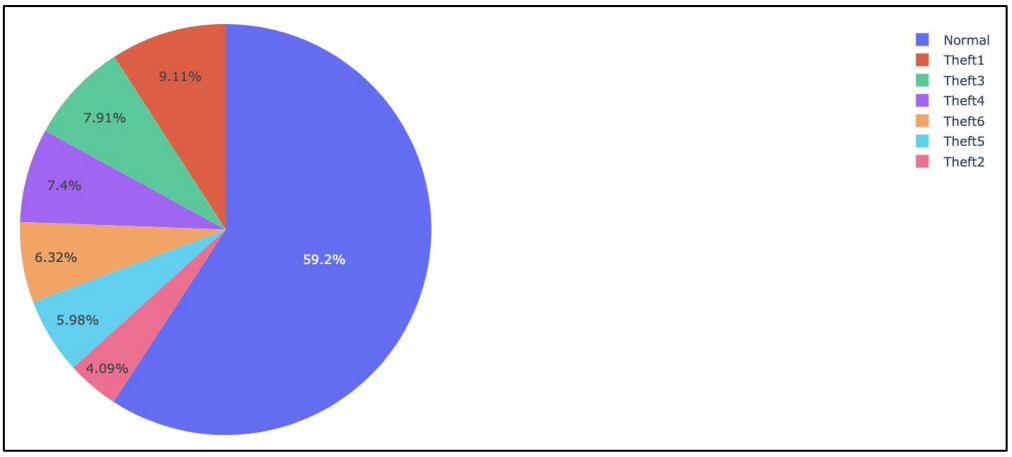


Fig.3: Energy Consumption Data Class Distribution Featuring Fraudulent Patterns

A pie chart illustrating the distribution of target class values is presented in Figure 3. Depicted in cyan, is the "normal" class, comprising 59.2% of the instances. The residual segments represent occurrences of fraudulent activities, which are classified as theft 1 through theft 6. This pattern underscores the differing percentages of each form of fraud present in the dataset.

Following the removal of inappropriate values during pre-processing, the researcher employed label encoding to turn the categorical data, namely the class and theft variables, into numerical data. Next, a correlation matrix (Fig.4) was computed for the full data-frame and then displayed visually using a heatmap.

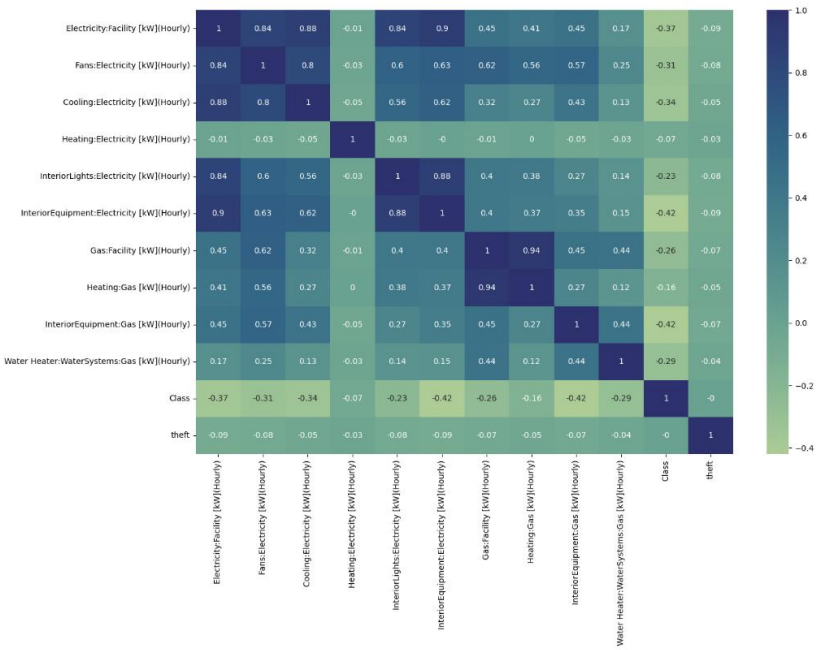


Fig.4: Correlation Matrix

Upon examining the prevalence of classes in the 'theft' column to ascertain whether the data exhibits imbalance, the Extra-Trees regressor model is utilized to calculate the feature importance. Afterwards, a bar chart (Fig.5) is utilized to visually represent the outcome, and the top 10 significant features were chosen based on the model's feature importance. Then, the data was divided into features and a target variable, with theft being the designated target.

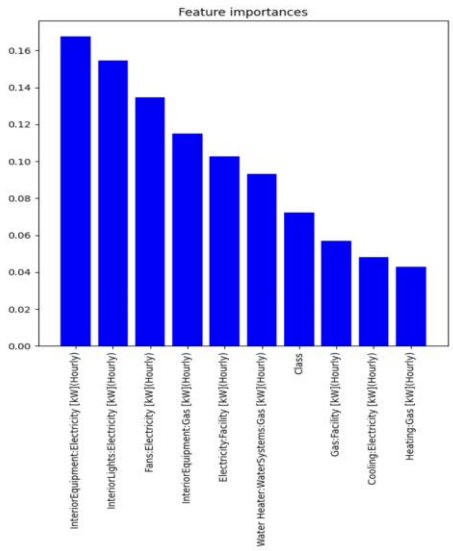


Fig.5: Bar chart of Feature Importance

In the Data Modelling phase, 5 unique models were created for the detection of electricity theft. Each and every model utilized a different machine or deep learning approach, which was then optimized through Particle Swarm Optimization (PSO) to provide a better result.

The combined use of the XGBoost and PSO models represents a robust ensemble learning approach that utilizes gradient boosting to develop a dependable and accurate classifier specifically designed for the identification of electricity theft (Jiang et al., 2020). PSO's

optimization capabilities significantly enhance the capability of XGBoost to manage complex data linkages and interactions. Random Forest model integrated with PSO is an another powerful approach. By collectively capturing intricate patterns and iteratively adjusting the model's parameters, enhanced precision and dependability can be achieved. The application of the Decision Tree model in conjunction with PSO represents a foundational yet efficacious methodology for discerning occurrences of energy theft. In contrast, the implementation of the Convolutional Neural Network (CNN) enhanced with PSO offers a more refined strategy for detecting electricity theft. Long Short-Term Memory (LSTM) models exhibit remarkable proficiency in capturing distant dependencies, rendering them ideally suited for time series-oriented endeavours such as detecting instances of electrical theft. By integrating PSO, the parameters of the LSTM are refined, resulting in enhanced capability to capture intricate patterns within the dataset (Yao et al., 2020).

Various evaluation metrics, including Accuracy, F1-score, Precision, Recall, and Confusion matrix, were employed to assess each model in the concluding phase. Through the conducted experiments, it was discovered that XGBoost with PSO surpasses all other methods in terms of performance. Both the Random Forest model with PSO and the LSTM with PSO obtained an accuracy exceeding 80% whereas the Decision Tree attained an accuracy of 67%. The CNN had limited capability in predicting only one class and produced relatively inferior outcomes compared to alternative methods.

4. Design

The primary objective of this research is to construct hybrid models, as opposed to traditional models, by incorporating a relatively novel approach known as machine learning or deep learning models coupled with particle swarm optimization (Fig.6 represents the basic structure of the PSO algorithm).

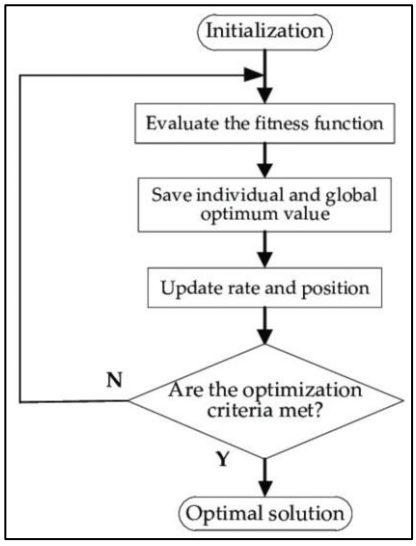


Fig.6: Basic structure of the particle swarm optimization (PSO) algorithm (Xiao et al., 2018)

The design of the framework encompasses data acquisition, data visualization, data pre-processing, data transformation, data splitting, model construction, and model assessment. A

comparative analysis was conducted using five deployed models to determine the most effective one. First, the dataset is retrieved from Mendeley Data and it becomes easily comprehensible after conducting the data visualization. Following data pre-processing and transformation, the resulting data was partitioned into two distinct components: features and target. During the model construction phase, a total of three machine learning models and two deep learning models using PSO were developed. The PSO technique was incorporated into every model to improve their convergence and forecast accuracy. The confusion matrix is employed in the evaluation process to analyse the predicted and actual labels. The performance of all the models is evaluated by deriving the precision value, f1-score, accuracy, and recall value. After evaluating and comparing the models, it is observed that XGBoost with PSO yielded the best accuracy. The entire procedure was carried out using the Python programming language and the researcher opted for Google Colab due to its convenience.

5. Implementation

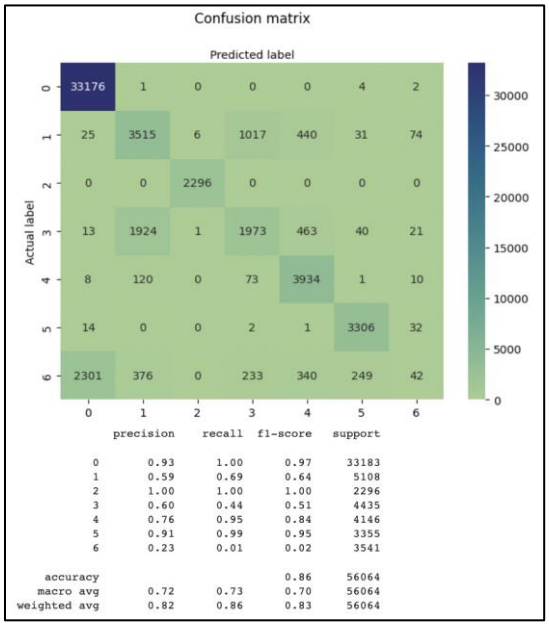
In this paper, a comparative analysis of optimized machine learning and deep learning models for electricity fraud detection is presented. The Python programming language has been leveraged to thoroughly implement this coding segment, and the entire coding process was completed on the Google Colab interface. The GPU that was utilized was a Tesla T4 which comes by default with the free version. For the purpose of implementation, a handful of essential libraries, were utilized, including NumPy and Pandas, which are used for numerical computations as well as data processing and analysis. Apart from these, a few more libraries were also used. Matplotlib.pyplot, plotly and seaborn were used for generating plots and visualizing statistical data. Joblib was utilized for storing and loading massive NumPy arrays. Imblearn was employed to handle imbalanced data. Sklearn.metrics was used for evaluating performance metrics and Keras was deployed for implementing deep learning models. Upon importing the csv file into the Pandas data frame, the data underwent pre-processing, visualization, and exploration in order to gain insights on its diversity. Following this, the data was divided into training and testing sets. The X_train and y_train as training sets and the X_test and y_test as testing sets were utilized for the subsequent coding step. The Extra Trees Regressor was utilized for the purpose of conducting the feature importance analysis in order to identify the most impactful features. The analysis focused on optimizing the performance of five distinct machine learning & deep learning models using PSO. The machine learning models, including XGBoost, Random Forest, and Decision Tree, were constructed and evaluated using Python's Sklearn module. On the other hand, the deep learning models, namely CNN and LSTM, were developed and assessed using Keras with TensorFlow backend. The classification reports and confusion matrices have been generated for all five models. The evaluation was conducted using metrics such as accuracy, precision, recall, and f1-score. The comparison of the results revealed that the XGBoost with PSO model attained the highest accuracy of 86% among all other models. The Random Forest coupled with PSO, achieved an accuracy of 83%. Similarly, the LSTM model with PSO, achieved an accuracy of 82%. However, the Decision Tree with PSO and the CNN with PSO models had poor outcomes in comparison to the other three models.

6. Results & Evaluation

The Evaluation section comprises an explanation of the designed models, including their architecture and their assessed performances. A total of five models were implemented and compared in order to determine the most effective one for detecting electricity theft. The results present vital insights into the effectiveness of the models across several optimal ML and DL strategies. The model evaluation was conducted utilizing a collection of classification criteria, which yielded a comprehensive comprehension of the individual capabilities of each model.

6.1 Experiment 1

At first the XGBoost machine learning model was implemented, which was optimized using PSO. XGBoost, a scalable ensemble learning algorithm based on tree boosting, was introduced by (Chen & Guestrin, 2016). It is renowned for its exceptional performance and extensive adoption in the field of machine learning. In recent times, it has garnered significant interest because to its exceptional efficiency and remarkable accuracy in prediction (Song et al., 2019). During the model development, first, the necessary libraries such as xgboost and pyswarm were imported. Next, the input features were standardized using the StandardScaler. The objective function utilized a collection of hyperparameters, such as max_depth or learning_rate, and trained the XGBoost model on the specified training data. Subsequently, it calculated on the testing data to check the inverse accuracy. Then the PSO technique performed a search within the defined lower and upper bounds, with a swarm size of 10 and a maximum iteration of 1 to find out the best hyperparameter. Once the ideal hyperparameter was determined, it was employed to train the ultimate XGBoost model. Afterwards, the code was employed to make predictions on the test data and calculated the accuracy of the absolute model.



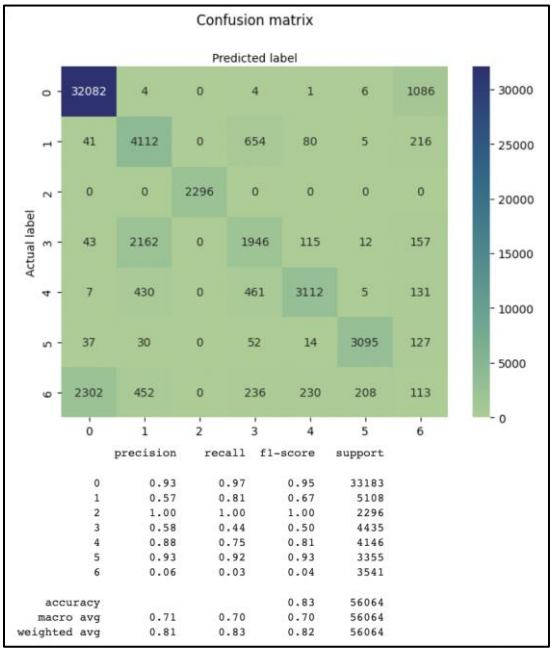
	precision	recall	f1-score	support
0	0.93	1.00	0.97	33183
1	0.59	0.69	0.64	5108
2	1.00	1.00	1.00	2296
3	0.60	0.44	0.51	4435
4	0.76	0.95	0.84	4146
5	0.91	0.99	0.95	3355
6	0.23	0.01	0.02	3541
accuracy			0.86	56064
macro avg	0.72	0.73	0.70	56064
weighted avg	0.82	0.86	0.83	56064

Fig.7: Confusion Matrix of XGBoost with PSO Model

Figure 7 demonstrates that the XGBoost model with PSO achieved a high level of performance, resulting in an accuracy of 86%.

6.2 Experiment 2

The Random Forest algorithm is often regarded as the most preferred and robust supervised machine learning technique. It is capable of efficiently handling both the regression and classification jobs (Ahire et al., 2019). Where, PSO is a stochastic method that has been successfully implemented in the resolution of numerous engineering as well as technological issues. It has been employed (García-Gonzalo & Fernández-Martínez, 2012) to solve a variety of inverse and continuous optimization problems, including combinatorial, different or distinct, dynamic, and multi optimal problems with or without further constraints. In this study, the initial stage in model construction involved importing the RandomForestClassifier from the scikit-learn library to train and develop the RF model. Similarly to the preceding model, this model likewise intended to determine the optimal hyperparameters for training the final Random Forest model.



	precision	recall	f1-score	support
0	0.93	0.97	0.95	33183
1	0.57	0.81	0.67	5108
2	1.00	1.00	1.00	2296
3	0.58	0.44	0.50	4435
4	0.88	0.75	0.81	4146
5	0.93	0.92	0.93	3355
6	0.06	0.03	0.04	3541
accuracy			0.83	56064
macro avg	0.71	0.70	0.70	56064
weighted avg	0.81	0.83	0.82	56064

Fig.8: Confusion Matrix of Random Forest with PSO Model

Based on the data presented in Fig. 8, it is evident that the Random Forest with PSO model had significant performance, achieving an accuracy of 83% as well as favourable recall, f1-score, and precision values.

6.3 Experiment 3

In this experiment the Decision Tree classifier with PSO model was employed. The fitness function was defined subsequent to verifying the dimensions of the training and assessment data. Similar to the objective function, the fitness function incorporates a set of

hyperparameters in order to train the classifier model. The fitness function was calculated and negative accuracy was derived from the test dataset. The final Decision Tree models are trained by extracting and utilizing the optimal hyperparameters, similar to the process employed for other two models. Once the predictions were made and evaluated, the confusion matrix was generated to examine the classification report and class prediction.

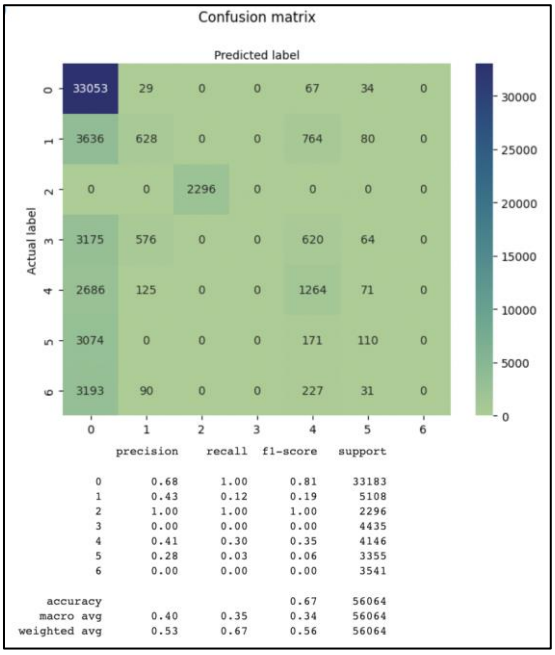


Fig.9: Confusion Matrix of Decision Tree Classifier with PSO Model

Figure 9 indicates that the accuracy offered by the Decision Tree with PSO model is inferior to the accuracy of the other two hybrid machine learning models that were previously implemented. Also, the precision value is unsatisfactory. So, among the three machine learning models optimized using PSO, XGBoost emerges as the top performer.

6.4 Experiment 4

In addition to machine learning models, two deep learning models were implemented in this research. This experiment involved constructing a model called CNN with PSO. After the dimension was checked and the required libraries were imported, the fitness function was called. The fitness function utilized a series of hyperparameters, including learning_rate, num_filters, kernel_size, and num_neurons, to train the CNN model on the training data. Subsequently, the function explored within the designated limits, employing a 10 particles swarm-size, and executed for a maximum of 1 iteration in order to identify the ideal hyperparameters. The PSO algorithm determined the optimal hyperparameters, which were then used to construct the CNN model using the training dataset. The performance of the model was subsequently assessed using the test dataset.

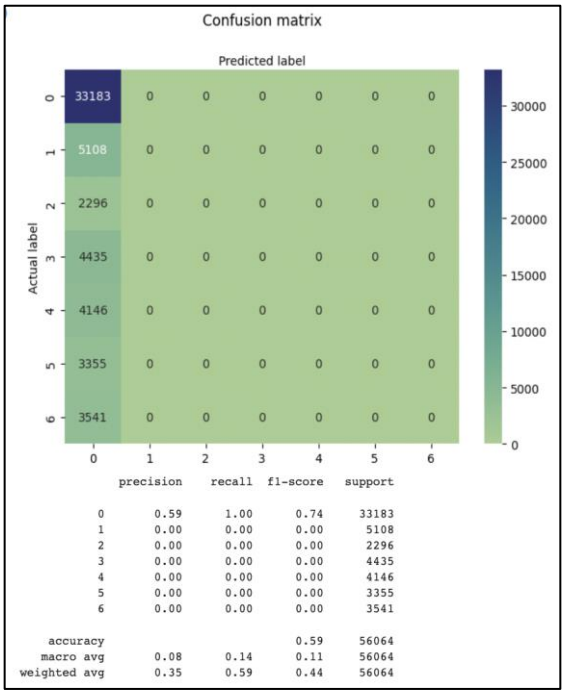


Fig.10: Confusion Matrix of CNN with PSO Model

The confusion matrix (Fig.10) reveals that the model exhibited a complete lack of accurate predictions, as it did not generate any True Positive (TP) or True Negative (TN) values. Furthermore, the attained accuracy of 59% is deemed unsatisfactory.

6.5 Experiment 5

The final experiment showcased the integration of LSTM with the PSO optimizer. The LSTM models improved with PSO represented a substantial step forward in the detection of electrical theft. The enhanced model demonstrated a notable aptitude for distinguishing between "normal" and "theft" instances. Several functions were defined for visualization and plotting, in addition to library importation which included plotCostHistory, plotPositionHistory, plot3D, plotTrainValAcc, plotTrainValLoss, and confclassif. Beyond these functions, LSTM function, loss function, particlesLoop function and applyLSTM_PSO function were defined to create and optimize the model.

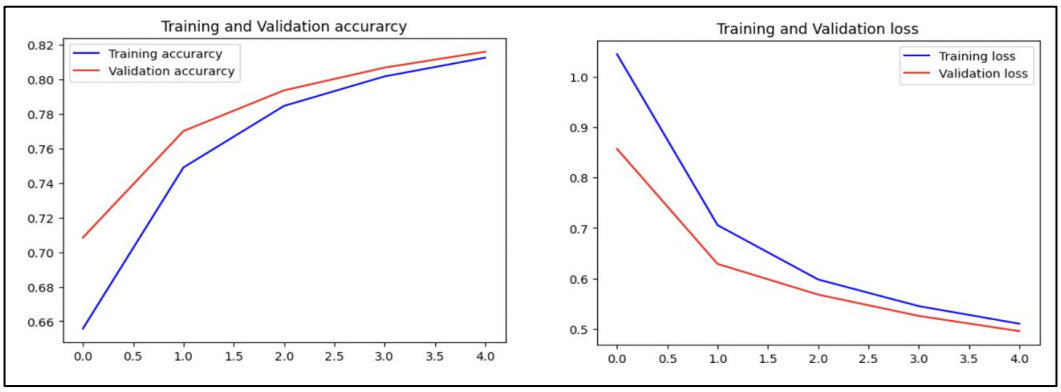


Fig.11: Loss and Accuracy graph of LSTM with PSO model

The Confusion matrix and the classification report (Fig.12) visually represents the model's performance.

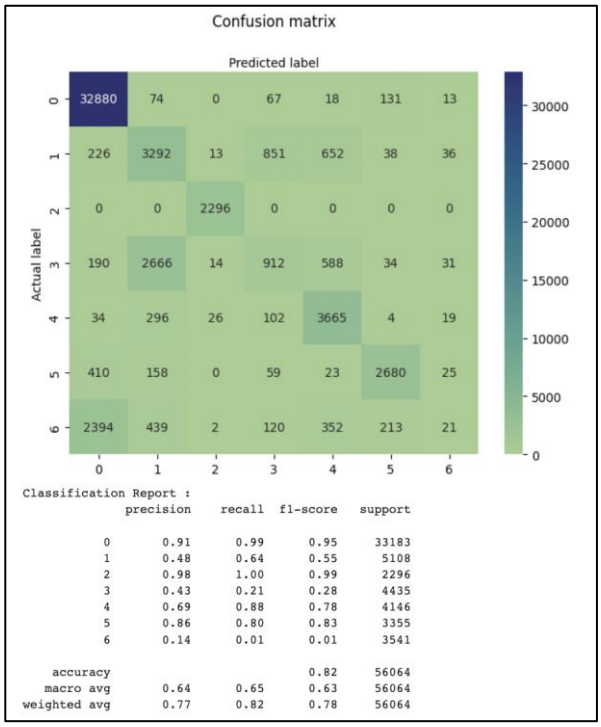


Fig.12: Confusion Matrix of LSTM with PSO model

The classification report showcases the precision, recall, and F1-score, indicating enhanced accuracy in comparison to the CNN with PSO model. The LSTM model with PSO achieved a superior accuracy of 82% compared to the other deep learning model, establishing it as one of the most effective choice for detecting electricity theft.

Upon conducting a comprehensive comparison (Table.2) of all the implemented models, it has been determined that the XGBoost with PSO model demonstrates its efficacy in real-world situations. As a result, it becomes a significant option for improving energy security and reducing losses caused by theft.

Approach	Model Name	Precision	Recall	F1 Score	Accuracy
Machine Learning	XGBoost with PSO	0.93	1	0.97	0.86
	Random Forest with PSO	0.93	0.97	0.95	0.83
	Decision Tree with PSO	0.68	1	0.81	0.67
Deep Learning	CNN with PSO	0.59	1	0.74	0.59
	LSTM with PSO	0.91	0.99	0.95	0.82

Table.2 Model evaluation comparison

7. Discussion

This section highlights the efficacy of the proposed hybrid models, the outcomes of the comparative analysis - whether they are satisfactory or not and the potential future opportunities and prospects of leveraging the successful models. Furthermore, this being an academic research project, there were certain limitations which were taken into account. Vandalism of electricity has been an everlasting problem. Numerous scholars have extensively examined and explored potential solutions to mitigate or avert these concerns, although they have not managed in halting it yet. The level of hardship is intensified in third world countries, particularly in densely populated areas. This research was an attempt to make a significant contribution in this space by utilizing three machine learning models and two deep learning models which were optimized using the PSO optimizer. The dataset was obtained from Mendeley Data and it is a synthetic version of the original data augmented with six distinct sorts of fraudulent activities. The original dataset is acquired from the Open Energy Data Initiative (OEDI) website. The synthetic dataset was utilized in this context as it could alleviate the privacy issues. Prior to constructing the models, data pre-processing, data visualization, feature extraction were carried out to optimize their efficiencies. Three of the five optimized models performed remarkably well, achieving high values for accuracy along with precision, recall, and f1-score. The XGBoost with PSO model demonstrated its superiority as the best-performing model with an accuracy of 86%. XGBoost is an algorithm for ensemble learning in which the forecasts of numerous decision trees are combined. This approach has the potential to generate predictions that are more precise and resilient in comparison to employing a solitary decision tree. Furthermore, the PSO is a robust optimization algorithm capable of searching a vast hyperparameter set in an efficient manner. It potentially aids in identifying the optimal hyperparameter configuration that enhances the accuracy of the model. The adoption of this strategy not only protects the revenue of service providers but also empowers consumers by providing them with knowledge about their electricity consumption and encourages them to participate in environmental sustainability.

8. Conclusion & Future Work

The intention of this research was to assist electrical service providers in mitigating electricity theft through the adoption of innovative hybrid models. Five optimized deep learning and machine learning models – CNN with PSO, LSTM with PSO, XGBoost with PSO, Random Forest with PSO and Decision Tree with PSO were employed to carry out a comparative analysis to determine the most effective one. Each of these models were implemented in a meticulously organized, processed, planned, and structured manner. Based on a comprehensive comparative analysis, it is apparent that XGBoost with PSO surpasses all other models in terms of performance. It achieved 86% accuracy which is remarkable. The Random Forest with PSO and LSTM with PSO models also demonstrated exceptional accuracy. However, the Decision Tree model combined with PSO only provided an accuracy rate of 67%. The CNN model using PSO algorithm failed to predict

any class and attained a mere accuracy of 59%. Therefore, utility companies may find successful models advantageous, as they can offset financial losses and enhance their customer service.

In forthcoming endeavours, it is intended to substitute PSO with alternative optimizers in order to assess their superiority. Also, the aim will be to enhance the CNN models through the adjustment of hyperparameters, such as fine-tuning, or by increasing the model's complexity with the addition of extra layers or filters. The Decision Tree Model with PSO also can be improved by adjusting the maximum number of iterations or specifying the particle representation. The dataset used in this study has synthetic data, so in future research can be extended with real life dataset, incorporating more diverse information from various geographical places which could improve the model's ability to handle different situations and make more accurate predictions. In summary, this research establishes a strong basis for further investigation and advancement in safeguarding electrical systems from theft and other hostile actions.

References

1. Ahire, M.C. *et al.* (2019) *Analysis of data using PSO model and Random Forest*, *International Journal for Research Trends and Innovation*. Available at: http://www.ijrti.org/ (Accessed: 05 November 2023).
2. Ahmad, T. *et al.* (2017) *Review of various modeling techniques for the detection of electricity theft in smart grid environment*, *Renewable and Sustainable Energy Reviews*. Available at: https://www.sciencedirect.com/science/article/abs/pii/S1364032117314090 (Accessed: 03 November 2023).
3. Akram, R. *et al.* (2021) *Towards big data electricity theft detection based on improved RUSBoost classifiers in smart grid*, *MDPI*. Available at: https://www.mdpi.com/1996-1073/14/23/8029 (Accessed: 01 November 2023).
4. Almazroi, A.A. and Ayub, N. (2021) *A Novel Method CNN-LSTM Ensembler Based on Black Widow and Blue Monkey Optimizer for Electricity Theft Detection*, *IEEE Xplore*. Available at: https://ieeexplore.ieee.org/abstract/document/9568902 (Accessed: 01 November 2023).
5. Asif, M. *et al.* (2021) *Alexnet-Adaboost-ABC Based Hybrid Neural Network for Electricity Theft Detection in Smart Grids*, *SpringerLink*. Available at: https://link.springer.com/chapter/10.1007/978-3-030-79725-6_24 (Accessed: 01 November 2023).
6. Banga, A., Ahuja, R. and Sharma, S.C. (2021) *Accurate Detection of Electricity Theft Using Classification Algorithms and Internet of Things in Smart Grid*, *SpringerLink*. Available at: https://link.springer.com/article/10.1007/s13369-021-06313-z (Accessed: 01 November 2023).
7. Bian, J. *et al.* (2021) *Abnormal Detection of Electricity Consumption of User Based on Particle Swarm Optimization and Long Short Term Memory With the Attention*

Mechanism, *IEEE Xplore*. Available at: https://ieeexplore.ieee.org/abstract/document/9366425 (Accessed: 01 November 2023).

8. Bitirgen, K. and Filik, Ü.B. (2022) *A hybrid deep learning model for discrimination of physical disturbance and cyber-attack detection in smart grid, International Journal of Critical Infrastructure Protection*. Available at: https://www.sciencedirect.com/science/article/abs/pii/S187454822200066X?casa_token=xzbh-j9FBqcAAAAA%3ApDy8Sxkm3qgcGa_kf4keoxmZ8zkr55751zCXooqX4IT2ROepmCzLGhZWQ_PWfFdSJcfgI9jZH_Mq (Accessed: 01 November 2023).
9. Chen, T. and Guestrin, C.E. (2016) *XGBoost: A Scalable Tree Boosting System, The ACM Digital Library*. Available at: https://dl.acm.org/doi/abs/10.1145/2939672.2939785 (Accessed: 05 November 2023).
10. Eddin, M.E. *et al.* (2022) *Fine-Tuned RNN-Based Detector for Electricity Theft Attacks in Smart Grid Generation Domain, IEEE Xplore*. Available at: https://ieeexplore.ieee.org/document/9964082 (Accessed: 01 November 2023).
11. García-Gonzalo, E. and Fernández-Martínez, J.L. (2012) *A brief historical review of particle swarm optimization (PSO), Latest TOC RSS*. Available at: https://www.ingentaconnect.com/contentone/asp/jbic/2012/00000001/00000001/art00002 (Accessed: 05 November 2023).
12. Gu, D. *et al.* (2022) *Electricity Theft Detection in AMI With Low False Positive Rate Based on Deep Learning and Evolutionary Algorithm, IEEE Xplore*. Available at: https://ieeexplore.ieee.org/abstract/document/9709670 (Accessed: 01 November 2023).
13. Huang, Y. and Xu, Q. (2021) *Electricity theft detection based on stacked sparse denoising autoencoder, International Journal of Electrical Power & Energy Systems*. Available at: https://www.sciencedirect.com/science/article/abs/pii/S014206151933666X (Accessed: 01 November 2023).
14. Hussain, S. *et al.* (2021) *A novel feature-engineered–NGBoost machine-learning framework for fraud detection in electric power consumption data, MDPI*. Available at: https://www.mdpi.com/1424-8220/21/24/8423 (Accessed: 01 November 2023).
15. Ismail, M. *et al.* (2020) *Deep Learning Detection of Electricity Theft Cyber-Attacks in Renewable Distributed Generation, IEEE Xplore*. Available at: https://ieeexplore.ieee.org/abstract/document/8998142/ (Accessed: 03 November 2023).
16. Javaid, P., Nadeem *et al.* (2023) *Electricity theft detection for energy optimization using deep learning models, WILEY ONLINE LIBRARY*. Available at: https://onlinelibrary.wiley.com/doi/10.1002/ese3.1541 (Accessed: 01 November 2023).
17. Jiang, H. *et al.* (2020) *Network Intrusion Detection Based on PSO-Xgboost Model, IEEE Xplore*. Available at: https://ieeexplore.ieee.org/abstract/document/9044370/ (Accessed: 04 November 2023).

18. Khan, Z.A. *et al.* (2020) *Electricity theft detection using supervised learning techniques on Smart Meter Data, MDPI*. Available at: https://www.mdpi.com/2071-1050/12/19/8023 (Accessed: 01 November 2023).
19. Li, S. *et al.* (2019) *Electricity theft detection in power grids with deep learning and random forests, Journal of Electrical and Computer Engineering*. Available at: https://www.hindawi.com/journals/jece/2019/4136874/ (Accessed: 01 November 2023).
20. Lin, G. *et al.* (2021) *Electricity Theft Detection in Power Consumption Data Based on Adaptive Tuning Recurrent Neural Network, Frontiers in Energy Research*. Available at: https://www.frontiersin.org/articles/10.3389/fenrg.2021.773805/full (Accessed: 01 November 2023).
21. Miao, X. *et al.* (2022) *Studies on Electricity Theft Detection Approach Based on LSTM-SAE and Support Vector Machine, IEEE Xplore*. Available at: https://ieeexplore.ieee.org/document/9949789/ (Accessed: 01 November 2023).
22. Muzumdar, A., Modi, C. and Vyjayanthi, C. (2022) *Designing a blockchain-enabled privacy-preserving energy theft detection system for smart grid neighborhood area network, Electric Power Systems Research*. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0378779622001146?casa_token=8VSDiePqHs4AAAAA%3AQFsTIE9uDrnI6LHLKyUZ3fkVXXKO9-Lh_otjosc3S72KmJfW4TH37GXR5PeA8FcdRY7rGdlNT4Oq (Accessed: 01 November 2023).
23. Pereira, D.R. *et al.* (2016) *Social-spider optimization-based support vector machines applied for energy theft detection, Computers & Electrical Engineering*. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0045790615003572?casa_token=TUyVa4CmcggAAAAA%3A5mR1P_QAKbshdNvgziu6PgY7ZhdN89Jn5Bk6e_q5A0_WQT2vcAFjymDsKKA2Bq8bmG7o3eg34qqB (Accessed: 01 November 2023).
24. Razavi, R. *et al.* (2019) *A practical feature-engineering framework for electricity theft detection in smart grids, Applied Energy*. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0306261919300753 (Accessed: 03 November 2023).
25. Song, K. *et al.* (2019) *A steel property optimization model based on the XGBOOST algorithm and improved PSO, Computational Materials Science*. Available at: https://www.sciencedirect.com/science/article/pii/S0927025619307712#b0105 (Accessed: 05 November 2023).
26. Ullah, A. *et al.* (2020) *CNN and GRU based Deep Neural Network for Electricity Theft Detection to Secure Smart Grid, IEEE Xplore*. Available at: https://ieeexplore.ieee.org/document/9148314 (Accessed: 01 November 2023).
27. Wallace, S. and -T. Lu, K. (2016) *Phasor Measurement Unit, Phasor Measurement Unit - an overview | ScienceDirect Topics*. Available at: https://www.sciencedirect.com/topics/computer-science/phasor-measurement-unit (Accessed: 01 November 2023).
28. Xia, R. *et al.* (2022) *An efficient method combined data-driven for detecting electricity theft with stacking structure based on Grey Relation Analysis, MDPI*. Available at: https://www.mdpi.com/1996-1073/15/19/7423 (Accessed: 01 November 2023).

29. Xiao, Y., Wang, Y. and Sun, Y. (2018) *Reactive Power Optimal Control of a Wind Farm for Minimizing Collector System Losses*, *ResearchGate*. Available at: https://www.researchgate.net/publication/329007429_Reactive_Power_Optimal_Control_of_a_Wind_Farm_for_Minimizing_Collector_System_Losses/fulltext/5bef76d9299bf1124fd82972/Reactive-Power-Optimal-Control-of-a-Wind-Farm-for-Minimizing-Collector-System-Losses.pdf (Accessed: 05 November 2023).
30. Yao, Y., Han, L. and Wang, J. (2020) *LSTM-PSO: Long Short-Term Memory Ship Motion Prediction Based on Particle Swarm Optimization*, *IEEE Xplore*. Available at: https://ieeexplore.ieee.org/abstract/document/9018688/ (Accessed: 04 November 2023).
31. Zheng, Z. *et al.* (2017) *Wide and Deep Convolutional Neural Networks for Electricity-Theft Detection to Secure Smart Grids*, *IEEE Xplore*. Available at: https://ieeexplore.ieee.org/document/8233155 (Accessed: 01 November 2023).
32. Zidi, S. *et al.* (2022) *Theft detection in smart grid environment*, *Mendeley Data*. Available at: https://data.mendeley.com/datasets/c3c7329tjj/3 (Accessed: 04 November 2023).
33. Zou, D. *et al.* (2010) *Novel global harmony search algorithm for unconstrained problems*, *Neurocomputing*. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0925231210003371 (Accessed: 01 November 2023).