Exploremetics
Journal of Innovative Engineering and Technology

Volume: 05 No: 01 | April -2024
ISSN (Online) 2636 – 590
ISSN (Print) 2636 - 591X

APPLICATION OF ARTIFICIAL INTELLIGENCE TECHNIQUE FOR ELECTRICITY THEFT DETECTION SYSTEM

Ezeji Nwamaka G*[1], Ezigbo Lucy I[1], Nwobodo-Nzeribe Nnenna H.[1]
1 Department of Computer Engineering, Enugu State University of Science and Technology
Author for correspondence: Ezeji N. G; **E-mail:** georgeniaezeji@yahoo.com

Abstract - Electricity theft is a significant global issue that costs power providers lots of funds annually in damages. This paper aims to apply Artificial Intelligence (AI) to detect electricity theft. Methodology used are data collection, data analysis, feature selection with Chi-Square, feature transformation with Principal Component Analysis (PCA), Support Vector Machine (SVM), and model for energy theft detection. To achieve this, a Whale Parameter Optimization Algorithm (WPOA) was proposed and applied to improve SVM training performance, using meter recharge data collected from Enugu Electricity Distribution Company (EEDC). The system results considering False Discovery Rate (FDR) reported that 0.1 was achieved for the whale-based SVM model. When the True Positive Rate (TPR) was considered for analysis, it was observed that whale-based SVM attained a score of 0.9. In addition, whale-based SVM attained a Positive Predictive Value (PPV) of 0.90. In terms of accuracy, the whale-based SVM reported an accuracy of 0.89. Overall, the result showed that the whale-based SVM outperformed other counterpart models from the system validation achieved through comparative analysis, hence it is recommended for the development of energy theft investigation application.

Keywords: Electric Theft; Artificial Intelligence; Support Vector Machine; Whale Parameter Optimization Algorithm; Feature Selection; Chi-Square Algorithm

1 Introduction

Electrical energy is a fundamental necessity for daily life and a crucial commodity upon which organizations and individuals rely to conduct their daily affairs. The energy is generated, transmitted, and distributed through the power system. These interconnected sections of the power system are managed through collaboration between public and private stakeholders and require significant capital investments. While the government often subsidizes power supply, it is not provided for free.

Regrettable, Gbenga (2023) submitted that certain private individuals, enterprises, and organizations resort to illegal means to access electricity without payment by bypassing the proper channels (power theft). Apart from the inherent dangers associated with illegal connections during power theft, such as the risk of electrocution, increased chances of faults, power system instability, and compromised safety, there are significant social and economic implications. One of these implications is the loss of revenue by distribution companies. According to Gbolahan (2017), power theft refers to the unlawful consumption of electrical energy without proper payment or permission.

The Nigerian Electricity Regulatory Commission (NERC) reports that more than 25% of the total energy transmitted by the Transmission Company of Nigeria to electricity distribution companies is lost due to theft and technical inefficiency. This alarming statistic underscores the need for effective measures to combat power theft, protect the integrity of the power system, ensure equitable distribution, and safeguard the financial sustainability of the energy sector.

Over the years, several administrative measures have been proposed to combat power

theft, including legal penalties under Sections 286(2) of the Penal Code and Section (PCS) 94(3) of the Electric Power Sector Reform Act (EPSRA). Additionally, the introduction of smart electricity meters with security tokens has been implemented to detect illegal tampering (Ebole, 2022). However, Gbenga (2023) argued that despite these efforts, meter bypass remains a prevalent issue in Nigeria.

Electricity theft not only causes revenue losses for utility companies but also leads to imbalanced grids, increased operational costs, and risks to the safety of individuals and properties. To combat electricity theft effectively, utilities employ various anti-theft measures, such as implementing advanced smart metering technologies, conducting regular inspections, and raising public awareness about the consequences of theft.

This paper on implementing an Artificial Intelligence (AI) electricity theft detection system is aimed to develop a system that can automatically identify and detect instances of electricity theft in real time. This would help to reduce financial losses for electricity companies and improve the reliability of the electricity grid.

2 Review of Relevant Literatures

Lezama, et al., (2022) developed a detection system for electricity theft using a machine learning technique. The work utilized data from smart electrical meters belonging to the State Grid Corporation of China, which supplies approximately1.1 billion users across the national territory. The investigation focused one valuating the performance of five machine learning models which are Support Vector Machine (SVM), K-Nearest Neighbor (K-NN), Random Forest (RF), Logistic Regression (LR), and Naïve Bayes (NB) and applied for the detection of electricity theft. The accuracy rates of the models were reported as follows: 81% for SVM, 79% for K-NN, 80% for RF, 69% for LR, and 68% for NB. Consequently, the study recommended the application of SVM in future research, even though the accuracy leaves room for improvement.

Javaid, et al., (2022) developed a Recursive Feature Elimination (RFE) based feature selection and K-Nearest Neighbor Oversampling (KNNOR) based data balancing system for electricity theft detection. The system employed a Bidirectional Long-Short-Term Memory (BiLSTM) and Logit Boost stacking ensemble model. The system's functionality consisted of four stages: data preprocessing, feature extraction, data balancing, and electricity theft classification. The training and testing data sets were obtained from the state grid corporation of China's electricity consumption data. The performance of the proposed technique achieved an accuracy of 89.45% in electricity theft detection, however despite the success, there is room for improvement.

Mhaske, et al., (2022) presented an electricity theft detection system using the XGBoost algorithm. The system utilizes XGBoost and Optical Character Recognition (OCR) techniques to analyze data based on customers' consumption patterns recorded in the Advanced Metering Infrastructure (AMI). The goal is to detect non-technical losses in the electrical system and identify them as theft. The results of the system indicate that XGBoost, along with the OCR technique, provides accurate results with reduced time consumption. However, a practical validation of the model would have improved the reliability.

Onibonoje (2021) presented an IoT-based approach for the protection and billing of residential energy meters. The study focused on the real-time monitoring and management of residential, vendor, and consumer power systems through the meter using IoT technology. A user-friendly web-based platform was developed for various applications such as online meter recharge and remote control of user access to power for defaulting customers. However, the study did not mention electricity theft, which is the crust of the principal study.

Reshma, et al., (2022) conducted research on the implementation of machine learning for electricity theft detection. Their system utilized a green scheme to discover power robbery, computer payments, and display power loads

without compromising customer privacy. This was achieved by analyzing data from smart meters running on an advanced metering infrastructure. In addition, symmetric encryption utilizing the k-means algorithm was applied to encrypt meter readings and customer information. Furthermore, Convolutional Neural Network (CNN) model was utilized to learn features from the data on an hourly and daily basis and generated a model which was used for the calculation of energy consumption. The model was deployed into Atmega 328 using Octo-coupler device for the calculation of electricity units consumed. The result of the information generated from the system was stored to the cloud using NodeMcu. While this study made great contribution to energy management, it did not directly address the problem of energy theft.

Bohani, et al., (2021) conducted a comprehensive analysis of supervised learning techniques for the detection of electricity theft. The study aimed to evaluate the performance of various supervised learning techniques for detecting electricity theft. The dataset used for the study was acquired from the State Grid Corporation of China (SGCC). The techniques considered in the study were artificial neural network (Nwobodo-Nzeribe et al., 2022), decision tree, deep artificial neural network, and AdaBoost. The analysis revealed accuracies of 91.77% (decision tree), 92.74% (artificial neural network), 93.04% (deep artificial neural network), and 91.67% (Ada Boost). From the review of literatures, many researchers have applied diverse AI techniques for the detection of electricity theft. One of the most recent studies to address the issues of electricity theft was Lezama et al. (2022) who trained several machine learning algorithms such as SVM, K-NN, LR, NB, and RF for energy theft detection, and recommended SVM as the most suitable, however despite the successes, the author argues that prediction outcome of the model did not consider other factor which might affect the model training and customer behaviours. These factors include hyperparameter selection, load control, economic reasons, unavailability of the customers, non-experimental validation of the model and this has affected the reliability of the existing system. There is need for optimal hyperparameter selection so as to quickly get a reliable hyperplane decision boundary and a decision-based algorithm which complements the prediction model to address issues of false alarm and provide reliable solution for the monitoring and detection of customers suspected in energy theft. Deep artificial neural network was recommended as the best model; however, practical validation of the model can be used to enhance the trustworthiness.

3 Materials and Method

3.1 Materials

The materials required to model and implement an AI Based electricity theft detection system include data, hardware and software requirements.

Hardware requirements:

- LAPTOP to train and deploy the AI model (Hp Elite Book 840 G5 Intel Core I5-16GB RAM/256GB SSD/Backlit Keyboard/FP Reader Windows 11 Pro),
- 1.5KVA Inverter system for power supply;
- 5G router which provided internet access for data collection.
- IN-100RFsensor
- Tuya smart power sensor

Software requirements:

- Python – the programming language used for development
- Sklearn – python library for machine learning model training
- Matplotlib – python library for visualization and graphical analysis
- Data of customer meter research information
- Google Colab – online python Integrated Development Environment (IDE)

3.2 Methodology

The methodology for modelling and implementing an AI electricity theft detection system include data collection, data analysis and preparation, feature selection, feature transformation and the machine learning model

training and model evaluation. The machine learning algorithm considered for the detective system is the Support Vector Machine (SVM) algorithm which will be trained using the Whale Parameter Optimization Algorithm.

3.2.1 Data Collection

Data of customer meter information was collected from the EEDC, headquarter office, at Okpara, avenue, Enugu State, Nigeria. The data was collected considering subscriptions made by customers in Obiagu district, Enugu in the years 2021 and 2022. The instruments used for the data collection are the IN-100 RF communications sensor and Tuya-smart power sensor used to monitor the energy consumption level of the user. The population size of data collection is 12,466 customers, with 16 attributes of meter information such as serial number, date, transaction ID, login, account, type of transaction, tariff, substation, token, power (w/h), electricity, vat, wallet, amount, and district. The total sample size of data collected is 2,119,322-meter records information.

3.2.2 Data Analysis and Preparation

The data preparation method applied was the integration of the dataset collected using the joining technique of data integration (Yevgeniya and Tumkovskiy 2022). According to Rebecca et al., (2019), the joining technique is specialized in merging datasets which has the same attributes as identifiers. This technique was applied considering the similar attributes of the customer meter recharge information and then merges as a single dataset used for the research. The flow chart in the figure 1 presented the process for the data integration.

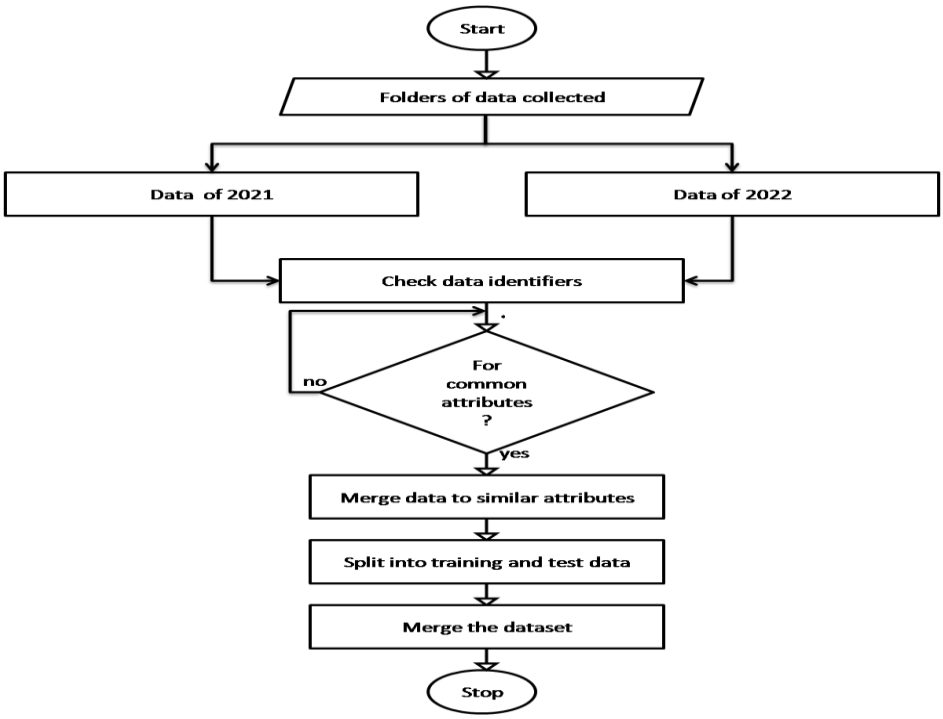

Figure 1: Flow chart of the data integration process

The figure 1 presented the flow chart operation of the data integration process used to merge the customer data collected from 2021 and 2022 as one dataset. The process identified common attributes of the data headers and then used to merge the data. Overall, the dataset was divided into training and test set in the ratio of 80:20 and then merged as the dataset for this research.

3.2.3 Feature Selection with Chi-square approach

Feature selection was utilized in this study to improve the quality of the data collected, model generation, cost reduction and address over-fitting during the training process of machine

learning algorithms through dimensionality reduction. The feature selection technique utilized in the study is the chi-square approach (Mengash et al., 2022). The step-wise approach of chi-square is presented as;

Stepwise Approach of the Chi-square algorithm

1. Chi-square test: measure dependence between observed and expected frequency variable
2. Data structure: Identify categorical values of data
3. Target variable: test target variables
4. Compute p value: compute probability of null hypothesis
5. Set significance value: compare p with set threshold
6. Select features: identify variables with significant p value
7. Return features: return output as selected features

This Chi-square focuses on the statistical dependencies of the data features to identify the most important and then prioritize in the dataset. This was achieved applying equation 1 to determine the features that are most relevant based on their chi-square probability score and then select as the most important attributes for energy theft detection (Tsehay et al., 2022).

$$Chi - square(X^2) = p = \sum \left[\frac{O-E^2}{E}\right] \quad \textbf{(1)}$$

Where O is the observed frequency of data in each cell, E is the expected frequency and p is the output of the chi-square test.

3.2.4 Feature Transformation

Feature transformation is a process used in this work to modify the data features and create new representation which is most suitable for identification by machine learning algorithms. The feature transformation algorithm adopted for the data is the Principal Component Analysis (PCA) (Navin et al., 2013). This PCA was applied to reduce the dimension of the dataset while retaining the quality of features using the steps in the algorithm 5;

Algorithm 5: Principal Component Analysis (PCA) Algorithm (Basna et al., 2021).

1. Input: Dataset X
2. Output: Reduced-dimensional data Y
3. Input: X, a dataset or data points.
4. For each feature, subtract the mean and divide by the standard deviation
5. Calculate the covariance matrix, C, of the standardized data X.
6. Compute the eigen-values (λ) and eigenvectors (v) of the covariance matrix C
7. Define the number of components to keep k.
8. Select the first k eigen-values and their corresponding eigenvectors.
9. Calculate the new data Y by multiplying X by the selected eigenvectors.
10. End

3.2.5 The Support Vector Machine (SVM) Algorithm

The SVM algorithm operates by searching for the perfect hyper-plane between the data points, while maximizing the class margins. It identified the supports vectors from the dataset sampled which are the closest data to the decision boundary and then build hyper-plane based using the equation 2.

$$f(x) = \beta_0 + \beta_1.x_1 + \beta_2.x_2 + \cdots \ldots\ldots\ldots\ldots\ldots + \beta_n.x_n \quad \textbf{(2)}$$

Where $f(x)$ is the decision function which defined the class data point x, β_0 is the intercept that adjusts the hyper-plane along the y-axis, $\beta_1, \beta_2, \ldots \ldots \beta_n$ are the hyper-plane coefficients for each data samples of x_1, $x_2,,,,,,,,,,x_n$.The equation 2 was used to determine the best hyper-plane between the classes of dataset. The trained SVM with the fixed hyper-plane was used to solve future classification or regression problem. The algorithm is presented as algorithm 1.

Algorithm 1: SVM Algorithm

1. Start
2. Initialize parameters, such as learning rate (δ) and regularization strength (C).
3. Randomly initialize weights (w) and bias (b).
4. Iterate until convergence:
5. For each training data sample (x, y):
6. Score = $w * x + b$;Margin = $y * score\%$ Compute score and margin
7. If

Margin < 1

$Applyw = w - \delta * (w - C * y * x)$ % Update weights:

$Applyb = b + \delta * C * y\%$ Update bias

8. Else:

Apply w = w - δ * w #Update weights

9. 4. Repeat the iteration process until convergence
1. Optimize weights (w) and bias (b)
2. Determine the decision boundary of the data points
3. Determine the separating hyper-plane with equation 2
10. End

The optimization algorithm that was used to train machine learning algorithm

To this end, the research employs the optimization of hyper-parameters considering the whale parameter optimization algorithm.

3.2.6 Whale Parameter Optimization Algorithm (WPOA)

The WPOA was used in this study for the selection of hyper-parameters and their optimal values to improve the training of machine learning algorithms. In the study WPOA was applied due to its ability to solve optimization problems modeling the smart behavior of whales which is generally categorized as intelligent animals which applied coordinated strategies to haunt preys. The algorithms is tailored towards optimal selection of parametric values during machine learning training, while striking a balance between known good solution and possible future solution. The WPOA operated by first identifying the best prey position and circulation of bubbles in a 9-shape is applied until the prey is captured and killed as shown in the figure 2 (Mirjalili and Lweis, 2016).

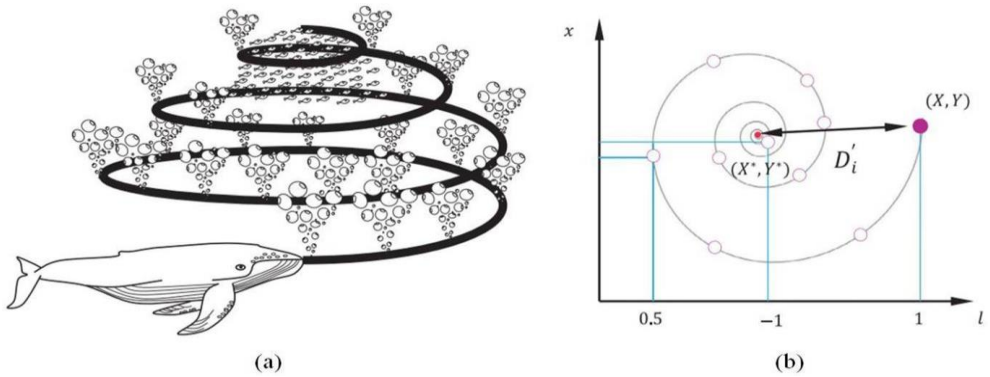

Figure 2: (a) Bubble-net feeding behaviour of humpback whales (b) Spiral updating position

The figure 2 was used to model the behaviour of the whales during the hunting process for prey. The (a)depicts the bubbling behaviour of the whale as it detects the prey in a circulation form, while updating its position to match the best search operator. The algorithm WPOA is as stated in algorithm 2 as follows;

Algorithm 2: The Whale Parameter Optimization Algorithm

1. Start
2. Initialize a population of whales: $X_i for\ i = 1,2,3,\ldots,n$:%% this involves random population of whale initialization
3. Defining objective function%% Here the key parameters such as position of prey, time of circulation until the prey is captured.
4. Create circling behavior around the prey %% this involves the creation of bubbles which confuses the prey state at the time.
5. Update the position of each whale %% this is based on three stages which are;
    a. Encircle prey %% this involves the adjustment of the whale position while moving towards the prey
    b. Search for prey %% this involved position best to capture the prey
    c. Update encircling %% adjustment of the whale

encirculation by computing the distance between whales.

6. Check for termination condition %% iteratively adjust the position of the whale until prey is captured and then adjustment terminated
7. Output %% once the algorithm converges and reached maximum iteration, the best result found which is the prey captured is returned as output

End

4 Model Training

This part of the work focused on training the optimized SVM algorithms developed for the generation of energy theft classification model. To achieve this, the data collected was transformed using the PCA algorithm and then feed to the SVM in algorithm 1, which then initializes the hyper-parameters and then applied the optimization based hyper parameter selection algorithm developed with WPOA to train the SVM. During the training phase, the score of the weights and margin was used to determine the decision boundary hyper plane and the support vectors.

Whale based SVM

The new SVM with WPOA is presented as;

Algorithm 3: Whale based SVM

1. Start
2. Initialize parameters, such as learning rate (δ), weights (w) and bias (b) and regularization strength (C)
3. Apply WPOA algorithm % for optimal hyper-parameter selection
4. Start training iteration
5. For each training data sample (x, y):

Score = $w * x + b; Margin = y * score\%$ Compute score and margin

6. If Margin < 1

$Applyw = w - \delta * (w - C * y * x)$ % Update weights:

$Applyb = b + \delta * C * y\%$ Update bias

7. Else: $Applyw = w - \delta * w$ #Update weights
8. Repeat the iteration process until convergence
9. Optimize weights (w) and bias (b)
10. Determine the separating hyper-plane
11. End

In the algorithm 3; the WPOA based SVM pseudocode was presented, showing how the SVM after the initialization of the hyper-parameters applied the WPOA for the optimal selection and adjustment while training to determine the best hyper-plane which separated the support vectors and then generated the classification model. The flow chart in the figure 3 showed the process work flow of the WPOA based SVM as it applies the whale hunting strategy to select the best hyper-parameters and generate the SVM model.

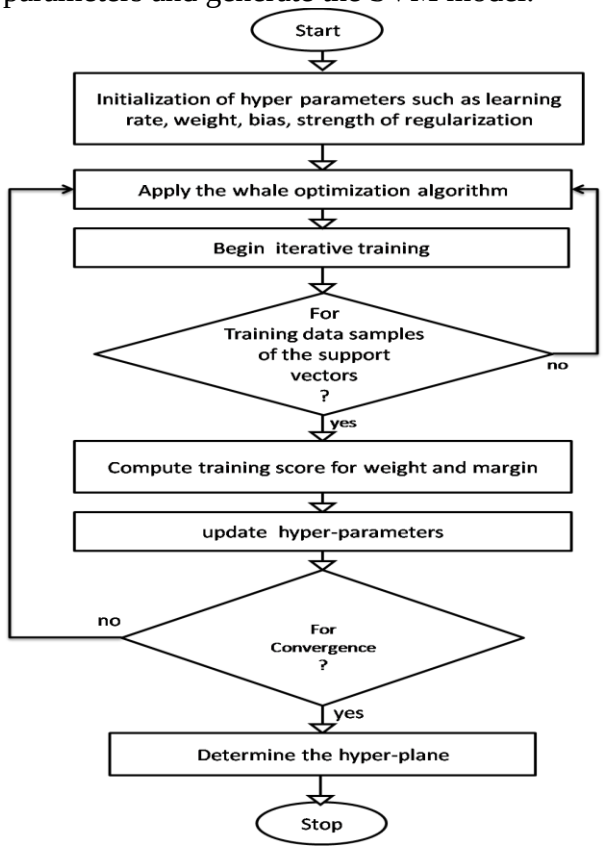

Figure 3: Flow chart of WPOA based SVM

The block diagram of the optimized SVM training

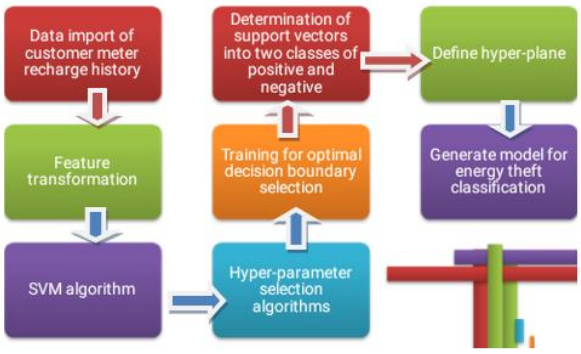

Figure 4: Block diagram of the Optimized SVM training for energy theft classification

The figure 4 presents the block diagram of the SVM training using the three optimization algorithms for the selection of hyper-parameters during the training process. While the parameters are optimally selected, the score of the weights and margins are used to determine the decision boundary which is the hyper-plane and then determination of the support vectors. While the training continued, the scores are monitored considering the positive and negative classes, until the hyper-parameters converge and then the SVM model generated for the classification of energy theft. The flow chart of the model was presented in the figure 5;

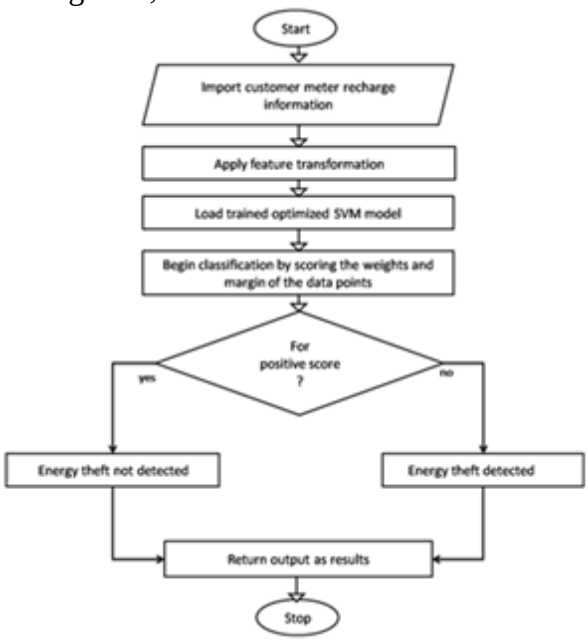

Figure 5: Flowchart of SVM based model for Intelligent Energy Theft Investigation Model (IETIM)

The figure 5 presents the IETIM achieved with the trained SVM algorithm. In the beginning of the flow chart, the data of customer meter recharge information was imported and the features transformed into the trained SVM algorithm for classification of energy theft. To achieve this, the trained SVM assigned scores to the data points that are used to determine the class of the hyper-plane it belongs. When these score are positive, then it belongs to the class of none meter theft, while when the scores are negative, meter theft is detected and returned as output.

5 Discussion of Results

The result of the SVM training displayed the training performance of the Whale based SVM, considering accuracy, false positive rate, false detection rate, true positive and confusion matrix respectively.

5.1 Result of the whale optimization based-SVM

The whale based SVM was applied to optimize hyper-parameter selection of SVM. The process involves defining a fitness function that evaluates the SVM model's performance given specific hyper-parameters. The Whale optimization algorithm iteratively updates a population of potential solutions based on the performance evaluation until a satisfactory solution is found or a stopping criterion is met. To evaluate the training performance, PPV, FDR, accuracy, regression, FPR and TPR were employed as shown in the figure 6.

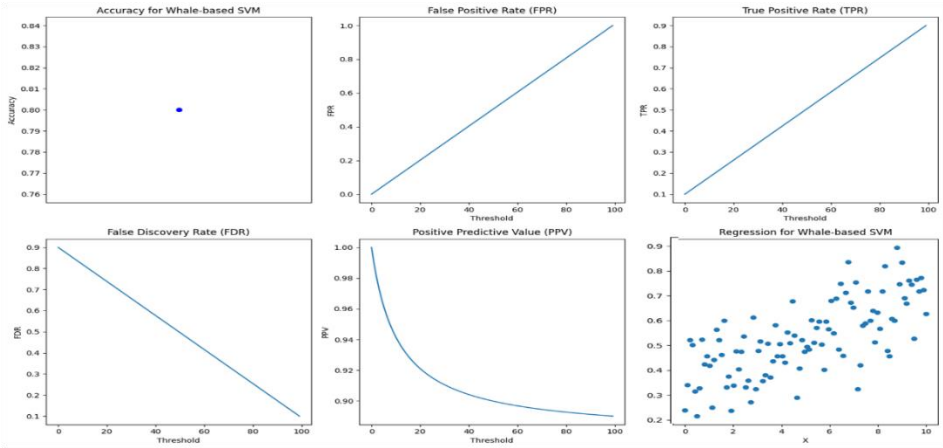

Figure 6: Results of the whale based SVM training

Table 1: Result of the SVM training performance

SVM WPO based	Accuracy	FPR	TPR	FDR	PPV	Regression
	89%	1.0	0.9	0.1	0.9	0.895

From the figure 6, the results for accuracy reported 0.89, TPR recorded 0.9, FPR recorded 1.0, PPV scored 0.9 respectively. This results overall, implied that the application of whale optimization based SVM in the classification of customers involved in energy theft suggests that the model exhibits strong predictive capabilities, particularly in correctly identifying instances of energy theft (as indicated by the high TPR and PPV scores) while maintaining a relatively high overall accuracy. However, the high FPR could indicate a potential issue with false alarms, which may require further investigation or fine-tuning of the model parameters to improve its precision. Nonetheless, these initial findings highlight the potential effectiveness of utilizing whale optimization based SVM for the crucial task of detecting and mitigating energy theft among customers.

5.2 Comparative analysis with other state of the art algorithms

The comparative analysis considered the new systems developed with other state of the art existing algorithms for energy theft investigation. In the analysis, accuracy was singled out as a key metrics to assess the model classification performances as showcased in the table 1.

Table 2: Comparative analysis with other state of the Art Algorithms

Authors	Technique	Accuracy (%)
Javaid, et al., (2022)	Ensemble Model	89.45
Bohani, et al., (2021)	DT	91.77
	ANN	92.74
	D-NN	93.04
	AdaBoost	91.67
Dong, et al., (2021)	Privacy-Preserving Method	92.86
Ullah, et al., (2021)	Hybrid Deep Neural Network	87.00
Chen, et al., (2020)	Residual Neural Network	97.90
Madhure, et al., (2020)	CNN-LSTM	97.00
Ayub, et al., (2020)	CNN-GRU	91.10
Syed, et al., (2020)	Deep Learning	92.69
Nabil, et al., (2018)	Deep Recurrent Network	93.00
New system	Whale based SVM	89.00

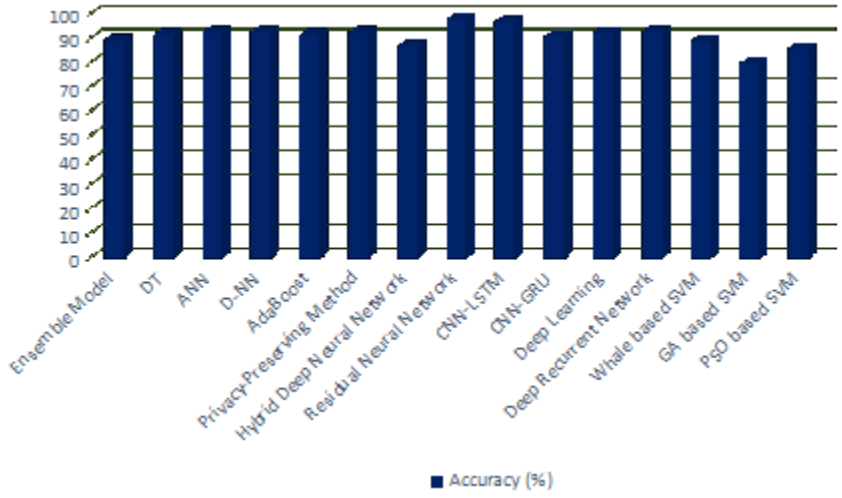

Figure 7: Comparative accuracy results

The table 2 compared the performance of the new system with the existing model developed for the classification of customers involved with electricity theft, considering the accuracy of correct suspected customer classification. These results were graphically presented in figure 5 for better analysis.

The figure 7 showcased the comparative analysis of the existing systems with the new system considering accuracy. From the analysis, it was observed that the new system which recorded encouraging accuracy values competes with the existing deep learning-based model. In addition, the new system supersedes the existing models because it was experimentally validated with real-world customer data. This experimental validation of the new system showcased how reliable the system is in a real-world energy theft scenario.

6 Conclusion

The work on the application of Artificial Intelligence (AI) technique for electricity theft detection system is aimed to develop a system that can automatically identify and detect instances of electricity theft in real time. It has been observed that while AI and SVM algorithm is a popular algorithm for solving regression and classification problems, issues of optimal hyper-parameter selection to facilitate the best hyper-plane decision boundary has remained a major challenge. To solve this problem, optimization algorithm was proposed and applied to train SVM. The system result considering FDR reported that 0.1 was achieved for the whale based SVM model. When TRP was considered for analysis, it was observed that whale based SVM attained a score of 0.9. In addition, whale based SVM attained PPV of 0.90. In terms of accuracy, the whale based SVM reported an accuracy of 0.89. Overall, the result showed that the whale based SVM outperformed other counterpart models from the system validation achieved through comparative analysis, hence it is recommended for use to develop the new software for energy theft investigation.

6.1 Recommendation

The developed system is recommended for adoption by the power distribution companies of Nigeria, particular the EEDC to facilitate energy theft investigation.

References

Ayub N., Aurangzeb K., Awais M., & Ali U. (2020). "Electricity Theft Detection using CNN-GRU and Manta Ray Foraging Optimization Algorithm", 2020 IEEE 23rd *International Multitopic Conference* (INMIC) | 978-1-7281-9893-4/20/$31.00 ©2020 IEEE | DOI: 10.1109/INMIC50486.2020.9318196

Basna S. H.. Abdulazeez A. (2021). A Review of Principal Component Analysis Algorithm for Dimensionality Reduction. , 2. https://doi.org/10.30880/JSCDM.2021.02.01.003

Bohani F., Suliman A., Saripuddin M., Sameon S., Salleh S., & Nazeri S. (2021). "A Comprehensive Analysis of Supervised Learning Techniques for Electricity Theft Detection", Hindawi: *Journal of Electrical and Computer Engineering Volume 2021, Article* ID 9136206, 10 pages https://doi.org/10.1155/2021/9136206

Chen Y., Hua G., Feng D., Zang H., Wei Z., & Sun G. (2020). "Electricity Theft Detection Model for Smart Meter Based on Residual Neural Network", 12th IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC).

Dong S., Zeng Z., & Liu Y., (2021). "FPETD: Fault-Tolerant and Privacy-Preserving Electricity Theft Detection", *Hindawi: Wireless Communications and Mobile Computing Volume* 2021, Article ID 6650784, 11 pages https://doi.org/10.1155/2021/6650784

Ebole Alpha F. (2020). "Electricity Power Theft Detection Using Wireless Prepaid Meter" http://sites.google.com/site/ijcsis/ ISSN 1947-5500; pp.36-78.

Gbenga, Y. (2023). "The crime of electricity theft and punishment for offenders"https://tribuneonlineng.com/the-

crime-of-electricity-theft-and-punishment-for-offenders.

Gbolahan O.O. (2017). "Electricity Theft and Power Quality in Nigeria" *International Journal of Engineering Research & Technology* (IJERT) http://www.ijert.org ISSN: 2278-0181 IJERTV6IS060492 Published by : www.ijert.org Vol. 6 Issue 06, June - 2017

Javaid P., Almogren A., Adil M., Javed M., & Zuair M. (2022). "RFE Based Feature Selection and KNNOR Based Data Balancing for Electricity Theft Detection Using BiLSTM-LogitBoost Stacking Ensemble Model", IEEE Access Digital Object Identifier 10.1109/ACCESS.2022.3215532 Pp 112948-112963

Madhure R., Raman R., & Singh S. (2020). "CNN-LSTM based Electricity Theft Detector in Advanced Metering Infrastructure", 11th ICCCNT 2020 July 1-3, 2020 - IIT - Kharagpur

Mengash H, Lal Hussain, Hany Mahgoub, A. Al-Qarafi, Mohamed K. Nour, Radwa Marzouk, S. A. Qureshi, A. Hilal. (2022). Smart Cities-Based Improving Atmospheric Particulate Matters Prediction Using Chi-Square Feature Selection Methods by Employing Machine Learning Techniques. Applied Artificial Intelligence, 36. https://doi.org/10.1080/08839514.2022.2067647

Mirjalili, S. & Lewis, A. (2016). The Whale Optimization Algorithm. *Adv. Eng. Softw.* **95**, 51–67 (2016).

Nabil M., Ismail M., Mahmoud M., Shahin M., Qaraqe K., & Serpedin E. (2018). "Deep Recurrent Electricity Theft Detection in AMI Networks with Random Tuning of Hyper-parameters", 24th International Conference on Pattern Recognition (ICPR) Beijing, China, August 20-24 978-1-5386-3788-3/18/$31.00 ©2018 IEEE Pp 740-745

Navin Goyal, S. Vempala, Ying Xiao. (2013). Fourier PCA and robust tensor decomposition. https://doi.org/10.1145/2591796.2591875

Rebecca E. Johnson, A. Grove, A. Clarke (2019). Pillar Integration Process: A Joint Display Technique to Integrate Data in Mixed Methods Research. *Journal of Mixed Methods Research*, 13. https://doi.org/10.1177/1558689817743108

Syed D., Abu-Rub H., Refaat S., & Xie L. (2020). "Detection of Energy Theft in Smart Grids using Electricity Consumption Patterns", 2020 IEEE *International Conference on Big Data* (Big Data) | 978-1-7281-6251-5/20/$31.00 ©2020 IEEE | DOI: 10.1109/BigData50022.2020.9378190 Pp 4059-4063

Tsehay Admassu Assegie, R. Tulasi, V. Elanangai, N. .. Kumar (2022). Exploring the performance of feature selection method using breast cancer dataset. https://doi.org/10.11591/ijeecs.v25.i1.pp232-237

Ullah A., Javaid N., Yahaya A., Sultana T., Al-Zahrani F., & Zaman F. (2021). "A Hybrid Deep Neural Network for Electricity Theft Detection Using Intelligent Antenna-Based Smart Meters", Hindawi: Wireless Communications and Mobile Computing Volume 2021, Article ID 9933111, 19 pages https://doi.org/10.1155/2021/9933111

Yevgeniya Tyryshkina, S. Tumkovskiy. (2022). Method for accelerating the joining of distributed datasets by a given criterion. https://doi.org/10.31799/1684-8853-2022-5-2-11