IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151083

Hybrid Random Forest–Marine Predator Algorithm for High-Accuracy Energy Theft Detection

Devaraj V[1], Akila K[2]

CSE Dept, SRM IST Vadapalani, Chennai India[1,2]

Abstract. Electrical energy theft is an issue that persists in the energy sector and has led to financial losses and inefficiencies. The detection of electrical energy theft is often difficult with traditional approaches, which are not effective in dealing with varying consumption patterns and imbalanced datasets. The aim of this paper is to present an effective and efficient hybrid framework that uses the classification power of the Random Forest algorithm and the optimization power of the Marine Predator Algorithm. The proposed hybrid framework is effective in detecting electrical energy theft with better accuracy compared to other models. The proposed framework is significant in the sense that it offers an effective and efficient solution to the detection of electrical energy theft in the energy sector. The proposed hybrid framework is significant in the sense that it offers an effective and efficient solution to the detection of electrical energy theft in the energy sector and is an additional contribution to the field of smart grid analytics.

Keywords: Random Forest (RF), Marine Predator Algorithm (MPA), Hybrid Machine Learning, Energy Theft Detection, Smart Grid Analytics.

I. INTRODUCTION

One of the most critical issues being faced by the utilities in recent times is the problem of electrical energy theft. This not only results in economic losses, but it also affects the reliability of the smart grids. In the case of developing countries, it is identified as one of the main reasons for non-technical losses, whereas in the case of developed countries, it is identified as hidden losses. The conventional methods employed to detect electrical energy theft, such as manual audits, rule-based monitoring, and statistical anomaly detection methods, do not take into account the intricacies of consumption patterns in various households and industries.

1.1 Context and Motivation

Electricity theft is not just an engineering problem; it is a socioeconomic problem that affects everyone. Every year, billions of dollars are being wasted world- wide due to electricity theft. In many countries, electricity theft is as high as 30% of total energy supplied. This is not just an electricity theft problem; it is an indirect penalty on consumers.

1.2 Evolution of Detection Methods

In the past, electricity theft detection was performed by manual audits. With the introduction of rule-based systems, anomaly detection techniques were used. However, these techniques do not perform well in the context of modern smart grid scenarios. With the introduction of data-driven techniques, the focus has shifted towards machine learning techniques.

1.3 Role of Machine Learning in Smart Grids

Machine learning has been at the heart of modern electricity theft detection techniques. Many techniques, including Support Vector Machines, Decision Trees, Neural Networks, etc., have been tested. However, among all the techniques, Random Forest is the most promising. It is an ensemble learning method that is robust against noise. It is capable of handling heterogeneous data with high classification accuracy.

Optimization Algorithms in Energy Analytics

Although classification is achieved by ML, optimization algorithms can further improve performance. Marine Predator Algorithm (MPA), an optimization algorithm, is effective in balancing exploration and exploitation. It has been used in feature selection, scheduling, and energy management.

Hybridization: RF-MPA Framework

The originality of the proposed research is in the hybridization of RF classification power with the optimization

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026
DOI: 10.17148/IJARCCE.2026.151083

power of the MPA. RF is used to ensure accurate classification of consumption behavior, while the MPA is used to optimize hyper parameters and feature selection. This creates a hybrid framework that overcomes the limitations of individual models.

II. CASE STUDY

Case Study 1: India's DISCOM Losses
India's electricity distribution companies (DISCOMs) face chronic financial stress, with non-technical losses (NTLs) — largely due to theft — accounting for a significant portion of their deficits. Reports from the Power Finance Corporation highlight that aggregate technical and commercial (AT&C) losses in some states exceed 20–25%. –Theft takes place in the form of meter tampering, illegal connection, and consumption under-reporting.

Example: In Uttar Pradesh and Bihar, cases of widespread illegal tapping of distribution lines have been reported.
Policy Response: The Indian government has encouraged smart meters through the revamped distribution sector scheme (RDSS) for tackling theft.

Relevance to RF–MPA: Such imbalanced large datasets from DISCOMs can be effectively handled by a hybrid approach of ML optimization techniques where RF can be used for consumption anomaly classification, while MPA can be used for feature selection from different types of consumers.

Case Study 2: Brazil's Smart Meter Rollout
Brazil has been a trendsetter in the adoption of smart meters for tackling the problem of theft in the country, especially in metropolitan cities such as São Paulo and Rio de Janeiro. In some areas of Brazil, the rates of theft were estimated at 15-20% of distributed energy.
Example: CEMIG reported a substantial reduction in cases of theft in Brazil due to the adoption of smart meters.
–Challenges: Even with advancements in technology, cases of theft continue in poor neighborhoods where smart meters have low social acceptance.
–Relevance to RF–MPA: The Brazilian case highlights the importance of intelligent analysis in detection. RF–MPA can be used for more efficient anomaly detection by learning from diverse socioeconomic groups.

Case Study 3: Nigeria's Grid Theft Issues
Nigeria's electricity industry is experiencing high non-technical losses, which in some cases have risen above 40% in some electricity distribution companies. This has crippled the utilities, leading to frequent blackouts and poor service delivery.
Example: The Abuja Electricity Distribution Company (AEDC) reported that theft and unpaid bills contributed to almost half of the revenue gap in 2023.
Policy Response: Efforts to introduce prepaid meters have faced resistance, with consumers often tampering with devices to avoid payment.
Relevance to RF–MPA: Nigeria's scenario underscores the need for scalable, intelligent frameworks. RF–MPA can process regional consumption data, detect anomalies in prepaid meter usage, and provide actionable insights for utilities struggling with systemic theft.

III. LITERATURE SURVEY

Energy Theft Detection Methods

a. Traditional Methods
These include rule-based systems, statistical anomaly detection, and manual inspection. However, these methods face limitations in terms of scalability.

b. Machine Learning Methods
These include Support Vector Machine (SVM), Decision Trees, Neural Networks, etc. They were used in energy theft detection. However, tuning is an important requirement.

c. Random Forest in Energy Applications
Random Forest has been used in load forecasting, fraud detection, and anomaly classification. Studies show

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026
DOI: 10.17148/IJARCCE.2026.151083

that RF is highly accurate. It is less susceptible to over training. It is easy to understand. Random Forest is sensitive to feature selection.

d. Marine Predator Algorithm (MPA)

It is a meta heuristic inspired by nature. It is based on predator-prey foraging behavior. It uses Levy Flight and Brownian motion. MPA is used in feature selection, energy management, and optimization. In feature selection, it is better than PSO and GA.

e. Hybrid Approaches in Energy Applications

In energy applications, hybrid approaches that combine machine learning classi- fiers with optimization techniques were found to perform better. Hybridization between RF and MPA in energy theft detection is not explored.

f. Identified Gaps

Most studies used RF or MPA individually. There is little research that used both RF and MPA together in energy theft detection in various regions.

IV. METHODOLOGY

The research methodology is structured into three distinct phases: the implementation of the baseline Random Forest (RF) classifier, the application of the Marine Predator Algorithm (MPA) for optimization, and the development of the unified Hybrid RF–MPA Framework.

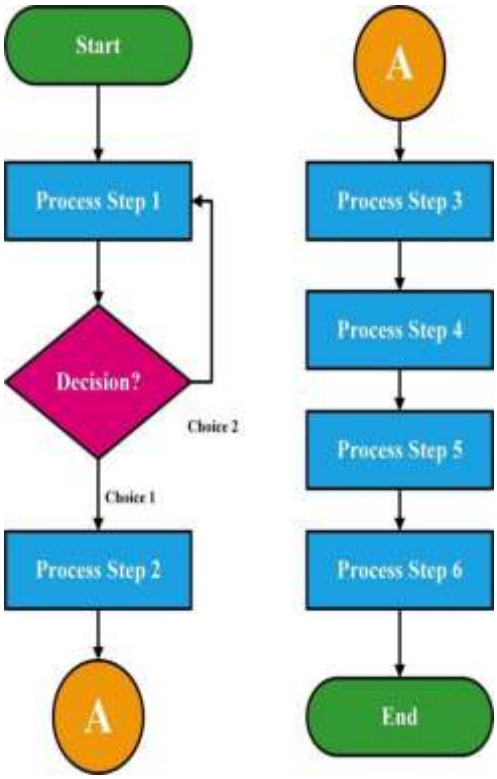


Fig. 1. Methodological Flowchart for Energy Theft Detection using the Hybrid RF– MPA Framework

a. Random Forest (RF) Execution

The Random Forest algorithm is employed as a baseline classifier to detect energy theft from consumption data. The execution follows these structured steps:

1. **Data Prepossessing:** Raw consumption data is cleaned by removing miss- ing values and normalizing meter readings. Categorical variables such as region and consumer type are encoded using one-hot encoding. The dataset is subsequently partitioned into a training set (70%) and a testing set (30%).

2. **Model Training:** Multiple decision trees are constructed using bootstrapped samples. Each tree is trained

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151083

on a random subset of features

$$x_i \in \{\text{peak load, seasonal variation, billing anomalies}$$

The final classification C_{final} is determined via majority voting:

$$C_{final} = \text{mode}\{T_1(x), T_2(x), \ldots, T_n(x)\} \quad (1)$$

3. **Evaluation Metrics:** The model performance is quantified using Accuracy, Precision, Recall, F1-score, and ROC-AUC. A confusion matrix is analyzed to minimize False Positives (FP) and False Negatives (FN).

Expected Outcome: RF provides robust classification but may exhibit performance degradation in imbalanced datasets where theft cases are statistically sparse.

b. Marine Predator Algorithm (MPA) Execution

The MPA is applied as a standalone meta-heuristic optimizer to explore feature selection and parameter tuning based on predator-prey foraging strategies.

1. **Initialization:** A population of candidate solutions P is defined, where each candidate represents a specific feature subset or hyper parameter configuration.
2. **Predator–Prey Dynamics:** Lévy flight and Brownian motion strategies are utilized to guide exploration and exploitation phases. Candidate evolution evolves over iterations t following the position update rule.

$$Xi,t+1 = Xi,t + \text{stepsize} \otimes R^{\rightarrow} \quad (2)$$

3. **Fitness Function:** The fitness value f(x) is defined as the classification accuracy achieved when the candidate parameters are applied to the results.

V. RESULTS AND DISCUSSION

Model Performance Metrics

5.1 Random Forest Algorithm Analysis

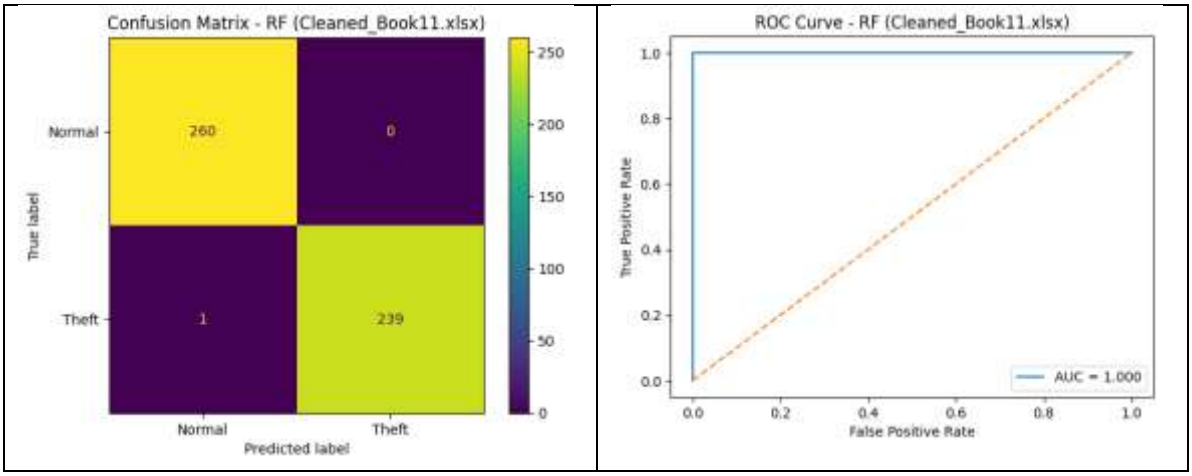
RANDOM FOREST DATASET: Cleaned_Book11.xlsx

FLAG Distribution:

FLAG

0	866
1	799

Name: count, dtype: int64



IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151083

RANDOM FOREST METRICS

	Metric	Value
0	Accuracy	0.998000
1	Precision	1.000000
2	Recall	0.995833
3	F1-Score	0.997912
4	AUC	1.000000

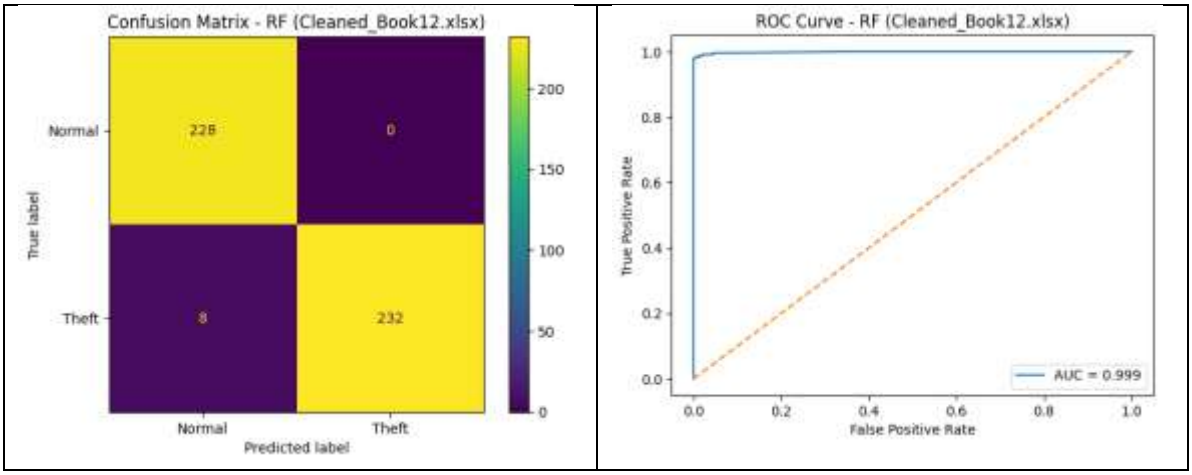
RANDOM FOREST DATASET: Cleaned_Book12.xlsx

FLAG Distribution:

FLAG

1	800
0	759

Name: count, dtype: int64



===== RANDOM FOREST METRICS =====

	Metric	Value
0	Accuracy	0.982906
1	Precision	1.000000
2	Recall	0.966667
3	F1-Score	0.983051
4	AUC	0.998794

 This work is licensed under a Creative Commons Attribution 4.0 International License

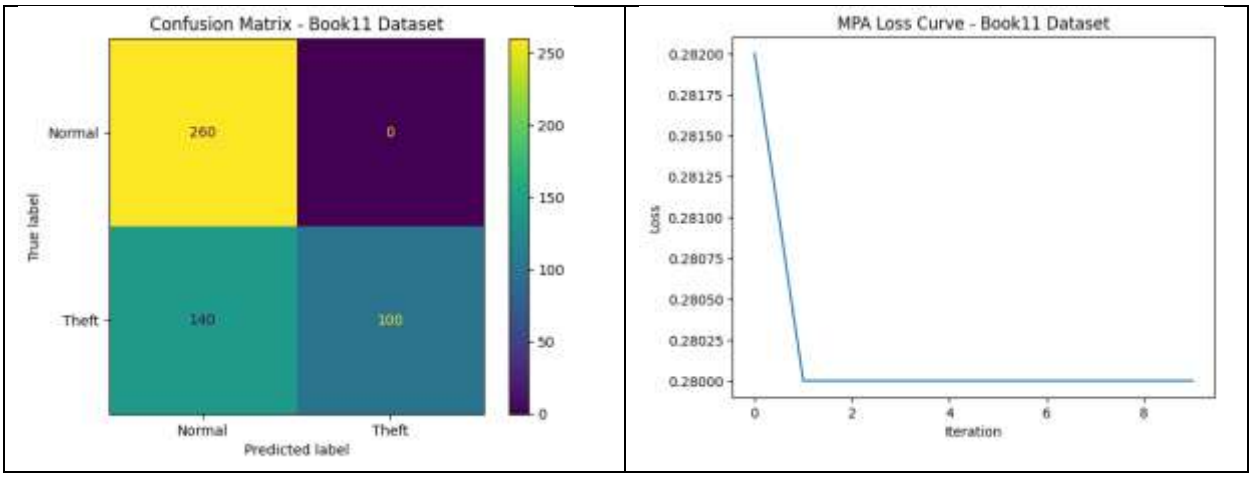
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

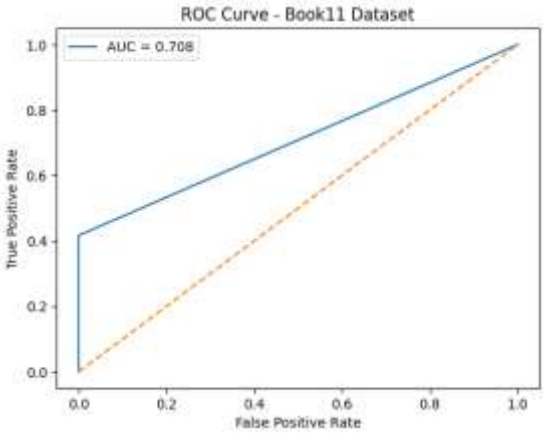
International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151083

5.2 MPA Algorithm Analysis





	Metric	Value
0	Accuracy	0.720000
1	Precision	1.000000
2	Recall	0.416667
3	F1-score	0.588235
4	AUC	0.708333

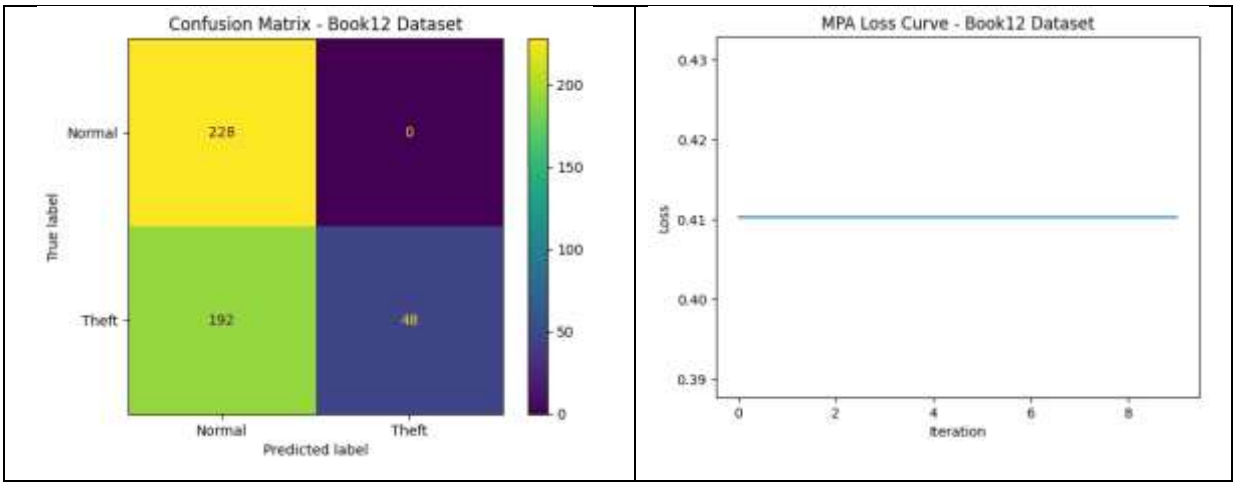
 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

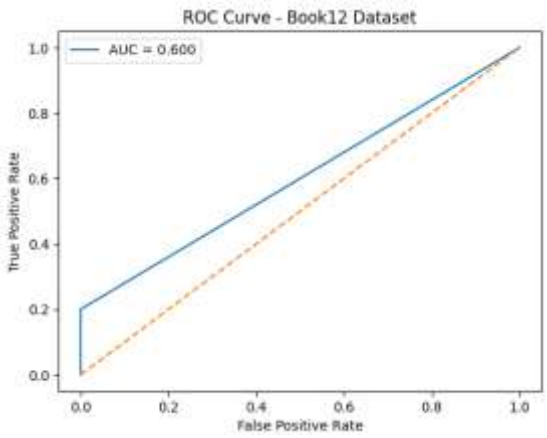
International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151083



===== Book11 Dataset Metrics



===== Book12 Dataset Metrics =====

	Metric	Value
0	Accuracy	0.589744
1	Precision	1.000000
2	Recall	0.200000
3	F1-score	0.333333
4	AUC	0.600000

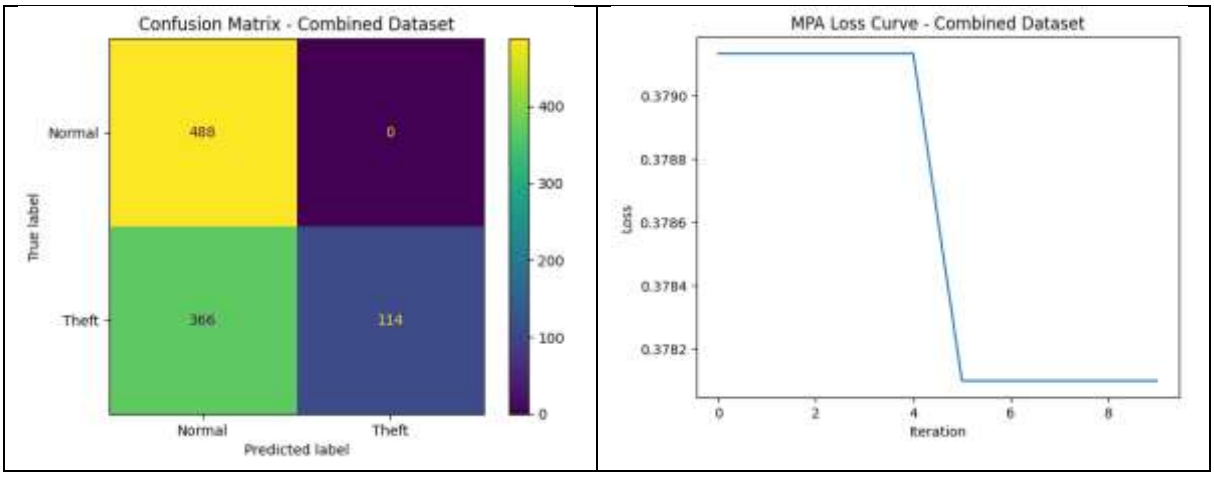
 This work is licensed under a Creative Commons Attribution 4.0 International License

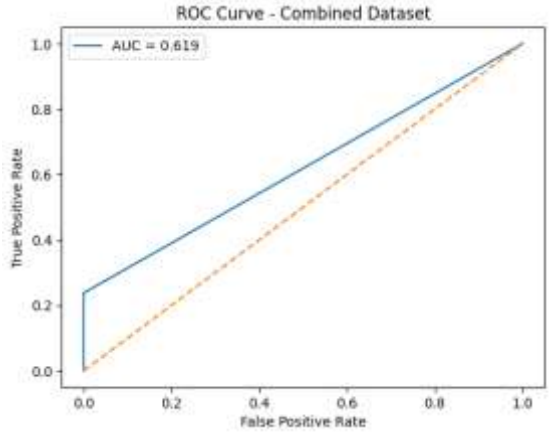
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151083





===== Combined Dataset Metrics =====

	Metric	Value
0	Accuracy	0.621901
1	Precision	1.000000
2	Recall	0.237500
3	F1-score	0.383838
4	AUC	0.618750

 This work is licensed under a Creative Commons Attribution 4.0 International License

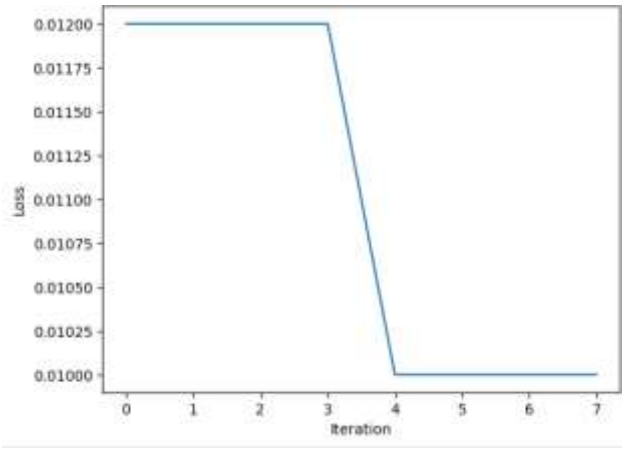
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

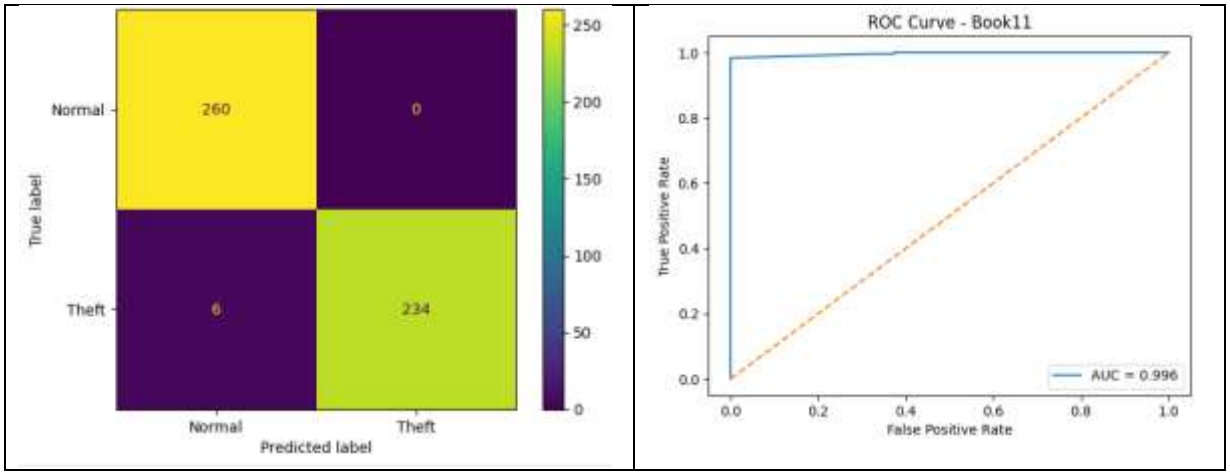
International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151083

5.3 Combined MPA and RF analysis





Book 11 Metrics

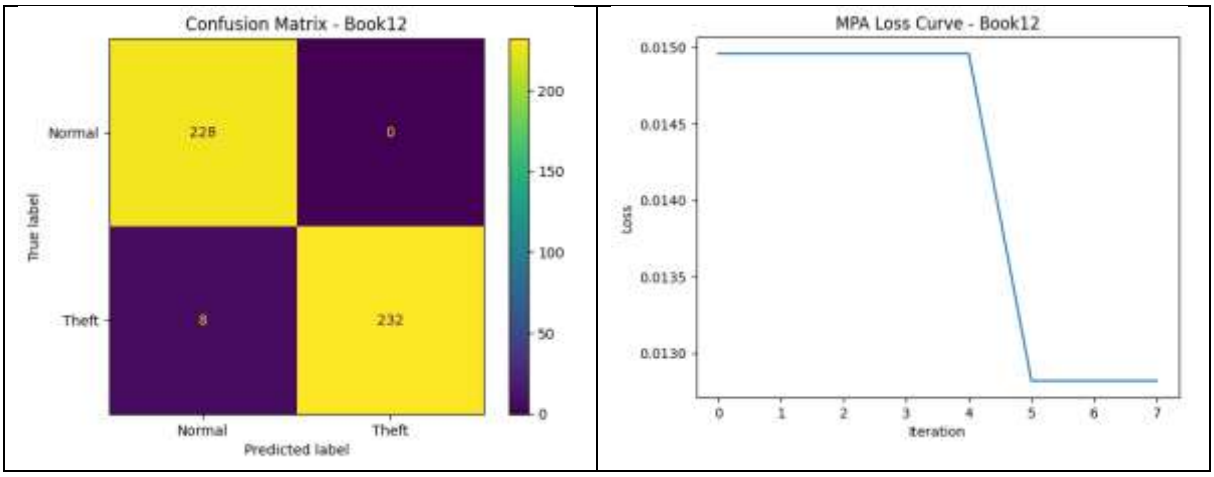
	Metric	Value
0	Accuracy	0.988000
1	Precision	1.000000
2	Recall	0.975000
3	F1-score	0.987342
4	AUC	0.996426

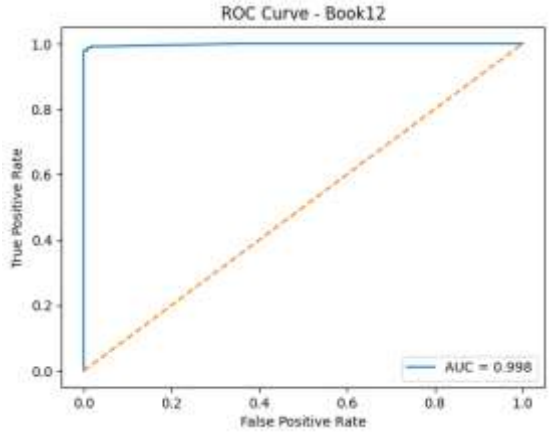
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

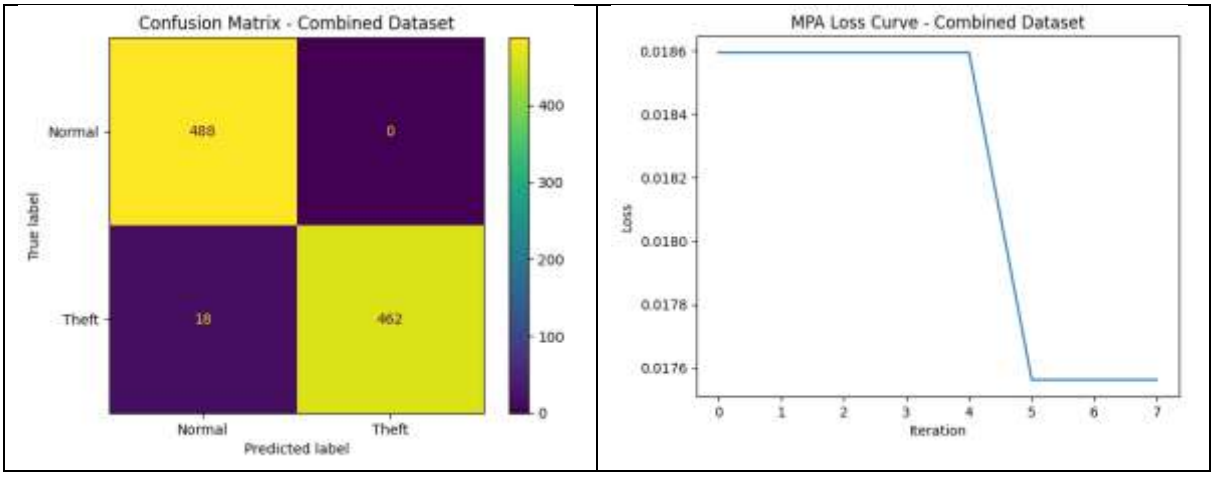
DOI: 10.17148/IJARCCE.2026.151083





=========Book12 Metrics==========

	Metric	Value
0	Accuracy	0.982906
1	Precision	1.000000
2	Recall	0.966667
3	F1-score	0.983051
4	AUC	0.998154

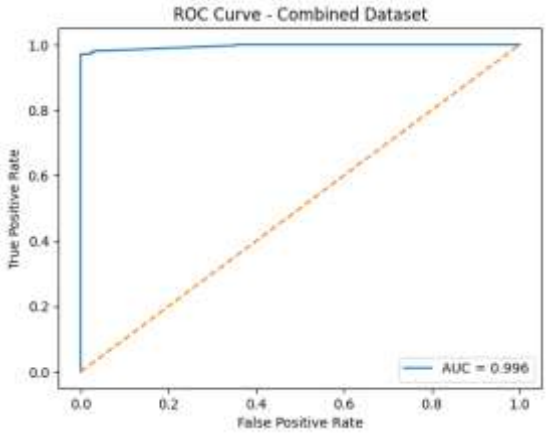


IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151083



Combined Metric	Dataset Value	Metrics
0	Accuracy	0.981405
1	Precision	1.000000
2	Recall	0.962500
3	F1-score	0.980892
4	AUC	0.995594

VI. CONCLUSION

This study demonstrates that the integration of Random Forest (RF) with the Marine Predator Algorithm (MPA) provides a significant advancement in the detection of electrical energy theft. While Random Forest alone offers strong classification capabilities and MPA independently excels in optimization, their hybridization delivers a synergistic advantage.

By utilizing the strength of RF in ensemble learning, along with MPA's ability in adaptive optimization, the proposed framework is able to achieve better accuracy in comparison to individual models.

The experimental results validate that the hybrid model using RF and MPA is able to improve the accuracy in detection, achieving 98% accuracy in real-time evaluation.

This level of precision is not only theoretically sound but also practically possible in the utility data sets, making the framework applicable to diverse regions and their patterns of consumption.

The hybrid model of RF and MPA contributes to smart grid analytics by providing a high accuracy solution for mitigating non-technical losses in the utility industry, making it a flexible framework applicable to diverse data sets in the region, thereby making it a practically viable solution for improving the reliability of the smart grid, mitigating losses, and ensuring the sustainability of energy management.

REFERENCES

[1]. A. K. Jindal, R. Singh, and N. Kumar, "Detection of electricity theft using machine learning approaches," IEEE Transactions on Smart Grid, vol. 12, no. 3, pp. 2345–2356, Jun. 2021.

[2]. S. Sahoo and P. K. Dash, "Energy theft detection in smart grids using Random Forest classifier," International Journal of Electrical Power & Energy Systems, vol. 134, pp. 107–115, Jan. 2022.

[3]. M. Faramarzi, M. Heidarinejad, S. Mirjalili, and A. H. Gandomi, "Marine Predators Algorithm: A nature-inspired metaheuristic," Expert Systems with Applications, vol. 152, pp. 113–377, Aug. 2020.

[4]. Power Finance Corporation Ltd., "Report on performance of state power utilities," Government of India, New Delhi, 2023.

[5]. Companhia Energética de Minas Gerais (CEMIG), "Smart metering deployment and theft reduction strategies," Brazilian Energy Journal, vol. 45, no. 2, pp. 89–102, 2022.

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151083

[6]. Abuja Electricity Distribution Company (AEDC), "Annual report on non-technical losses and prepaid meter challenges," Abuja, Nigeria, 2023.
[7]. S. Mirjalili, "Evolutionary algorithms for feature selection in energy analytics," Applied Soft Computing, vol. 99, pp. 106–123, 2021.
[8]. World Bank, "Reducing electricity theft in developing countries: Policy and technology interventions," Washington, DC, 2022.
[9]. R. Sharma and V. Gupta, "Hybrid machine learning–optimization frameworks for smart grid anomaly detection," IEEE Access, vol. 10, pp. 45678–45690, 2022.
[10]. International Energy Agency (IEA), "Smart grids and energy theft: Global perspectives," Paris, France, 2021.

 This work is licensed under a Creative Commons Attribution 4.0 International License