iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IV **Month of publication:** April 2025

DOI: https://doi.org/10.22214/ijraset.2025.68883

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

AI-Driven Security Operation Center

Namrutha S R[1], Mr. S. Syedsafi[2]
Department of Computer Science & Information Technology, Kalasalingam Academy of Research and Education, India

Abstract: ***Cyber threats such as malware, phishing, and DDoS attacks are becoming increasingly sophisticated, necessitating advanced detection mechanisms. This paper presents an AI-driven cybersecurity system that integrates machine learning models for real-time detection of cyber threats, network intrusions, phishing URLs, and email phishing. The system employs NLP for malware analysis, anomaly detection for intrusion detection, and classification models for phishing prevention. Developed using FastAPI for real-time inference and SQLite for secure logging, the system ensures efficient threat identification and response. Security measures such as SQL injection protection, API authentication, and data encryption further enhance its robustness. Experimental results show high detection accuracy, with intrusion detection at 96% and email phishing detection at 97%.***
Keywords: Cybersecurity, Machine Learning, Intrusion Detection, Phishing Detection, Anomaly Detection, Cyber Threats, NLP, FastAPI, Real-Time Threat Detection.

I. INTRODUCTION

Cyber threats are evolving at an alarming rate, with attacks like malware, phishing, and network intrusions becoming more sophisticated and harder to detect. Traditional security systems, which rely on predefined rules, often fail to identify new and advanced threats, especially zero-day attacks. As cybercriminals develop smarter ways to bypass security measures, there is a growing need for intelligent, automated systems that can detect and respond to threats in real-time. Artificial Intelligence (AI) and Machine Learning (ML) have transformed cybersecurity by allowing systems to analyze patterns, detect anomalies, and identify emerging threats with greater accuracy. By integrating AI, we can improve intrusion detection, phishing prevention, and malware identification, reducing reliance on manual monitoring. However, challenges such as scalability, real-time processing, and adversarial attacks against AI models still need to be addressed.

This paper introduces an AI-powered cybersecurity system designed to detect and mitigate cyber threats, including malware, network intrusions, phishing URLs, and email phishing. Unlike traditional security solutions that rely on static rule-based mechanisms, this system employs machine learning and deep learning models to analyze threats in real-time, providing a proactive defense against evolving cyberattacks. The system integrates four specialized AI models, each addressing a distinct cybersecurity challenge. The Cyber Threat Detection model utilizes RandomForest with MultiOutputClassifier to identify malicious patterns in text-based data. The Intrusion Detection System (IDS) applies XGBoost-based anomaly detection to recognize unauthorized network activities and DDoS attacks. For phishing prevention, the Phishing URL Detection model employs RandomForest-based classification to analyze domain heuristics, while the Email Phishing Detection model leverages BERT to detect phishing attempts in email content.

Developed with FastAPI for real-time inference and SQLite for secure logging, the system ensures efficient detection and response. Security features such as SQL injection protection, API authentication, and data encryption reinforce its robustness. By integrating AI-driven threat detection with strong security measures, this system enhances cybersecurity resilience while maintaining efficiency and scalability.

II. BACKGROUND AND RELATED WORK

Sunil Kumar et al. [1] proposed a semantic machine learning algorithm for cyber threat detection, demonstrating the effectiveness of ensemble learning in improving classification accuracy. The study highlighted the importance of machine learning models in detecting evolving cyber threats, but the system lacked real-world deployment strategies and comparative analysis with deep learning models. Demerits: Limited real-world applicability, no deep learning integration, and lack of comparative model evaluation. Timothy Alatise et al. [2] explored the use of Security Information and Event Management (SIEM) systems integrated with ML to enhance real-time threat detection. Their approach effectively combined log aggregation and machine learning-based anomaly detection, improving cyber threat visibility. However, the system exhibited high false-positive rates, and ML models required frequent retraining to maintain accuracy. Demerits: High false positives, retraining overhead, and limited scalability in high-traffic environments.

Hadi et al. [3] developed a realistic DDoS dataset and taxonomy to enhance Intrusion Detection Systems (IDS). The study emphasized anomaly detection techniques for detecting unauthorized access and DDoS attacks. While the dataset significantly improved IDS training and benchmarking, the research did not address real-time DDoS mitigation or adaptive model retraining. Demerits: No real-time mitigation strategies, lack of model adaptability, and potential performance degradation under dynamic attack conditions.

Mohammed Abutaha et al. [4] developed a URL phishing detection model using machine learning, focusing on lexical analysis of URLs. Their approach effectively identified malicious domains without relying on traditional blacklist-based methods. However, phishing websites that use domain generation algorithms (DGA) and obfuscation techniques remained a challenge. Demerits: Vulnerability to evasive phishing techniques, reliance on URL feature extraction, and no deep learning enhancement.

III. METHODOLOGY

A. Data Collection and Preprocessing

The proposed system utilizes four publicly available datasets to train and evaluate its machine learning models for detecting cyber threats, network intrusions, phishing URLs, and email phishing. Each dataset is carefully selected to provide diverse and realistic attack scenarios, ensuring that the models generalize well across different cyber threats.

For cyber threat detection, the Text-Based Cyber Threat Dataset from Kaggle is used. This dataset contains text logs, command sequences, and malware-related attack patterns, making it suitable for Natural Language Processing (NLP)-based classification. The dataset is preprocessed by removing stopwords, tokenizing text, and extracting word embeddings using BERT to improve classification accuracy.

For intrusion detection, the system relies on the CIC-DDoS2019 Dataset, which provides detailed network flow statistics related to DDoS attacks and unauthorized network intrusions. It contains records of various network traffic attributes, including protocol types, port numbers, packet sizes, and connection durations. To prepare this dataset for training, feature normalization and encoding techniques are applied to handle categorical variables such as protocol types. Additionally, entropy-based feature selection is used to retain only the most relevant network traffic attributes for anomaly detection.

For phishing URL detection, the Phishing Domain Dataset is used. This dataset consists of malicious and legitimate URLs, along with key attributes such as domain age, WHOIS registration data, lexical patterns, and the presence of special characters in URLs. The dataset is preprocessed by extracting relevant features and encoding phishing URLs as 1 (malicious) and legitimate URLs as 0 (safe). A RandomForest-based classification model is then trained to identify suspicious URLs based on these extracted features.

For email phishing detection, the system uses the Phishing Mail Dataset (Torch-RoBERTa), which contains a collection of emails labeled as phishing or legitimate. Each email includes metadata such as sender information, hyperlinks, and embedded content, which are analyzed to detect phishing attempts. The dataset is preprocessed by removing HTML tags, tokenizing text, and extracting word embeddings using BERT or RoBERTa. Additionally, hyperlink analysis is performed to detect malicious links embedded in emails.

B. Methodology

The proposed AI-powered cybersecurity system integrates machine learning and deep learning models to detect and mitigate various cybersecurity threats, including malware, phishing, and network intrusions. Each model is designed to process specific types of data and leverage supervised or anomaly-based learning techniques to classify threats accurately.

1) Cyber Threat Detection Model

RandomForest is an ensemble learning method that constructs multiple decision trees and aggregates their outputs to improve classification accuracy.MultiOutputClassifier extends RandomForest to support multi-label classification, meaning an input can belong to multiple threat categories simultaneously (e.g., malware and phishing).

a) RandomForest Decision Function:

$$P(y \mid X) = \frac{1}{T}\sum_{t=1}^{T} h_t(X)$$

where:

- T is the number of decision trees.
- h_t (X) is the prediction of the t-th tree.
- P(y|X) is the final classification probability.

b) Feature Extraction for Text Data:

TF-IDF (Term Frequency-Inverse Document Frequency):

$$TF - IDF(w) = TF(w) \times log(\frac{N}{DF(w)})$$

where:

- TF(w) is the frequency of word w in a document.
- DF(w) is the number of documents containing w.
- N is the total number of documents.

2) Intrusion Detection System (IDS) Model

XGBoost (Extreme Gradient Boosting) enhances traditional gradient boosting with regularization and efficient computation.The model classifies network traffic as normal or suspicious by learning from extracted flow-based features.

a) XGBoost Loss Function:

$$L(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

where:

- $l(y_i, \hat{y}_i)$ measures the difference between actual and predicted labels.
- $\Omega(f_k)$ is the regularization term controlling model complexity.

b) Detection Score for Anomaly Detection:

$$Score = XGBoost(F_{netwwork}) \geq \theta$$

where:

- θ is the anomaly detection threshold.
- $F_{netwwork}$ represents extracted network features (e.g., packet size, flow duration).

c) Entropy-based Feature Selection:

$$H = -\sum_{i=1}^{N} p_i \, log p_i$$

where:

- p_i s the probability of an event (e.g., specific packet behavior).
- H represents the unpredictability of network traffic.

3) Phishing URL Detection Model

RandomForest constructs multiple decision trees and classifies URLs based on extracted lexical and domain-based features.

a) Gini Impurity for Decision Tree Splitting:

$$Gini = 1 - \sum_{i=1}^{C} p_i^2$$

where:

- C is the number of classes (phishing or legitimate).
- p_i is the probability of a sample belonging to the class i.

b) Feature-Based Classification

$$F_{url} = RandomForest(Extracted\_Features)$$

Feature Extraction for URLs:

- Lexical Features: URL length, number of subdomains, special characters.
- Domain Age: Extracted from WHOIS records.
- HTTPS vs. HTTP: Checks if the URL uses a secure protocol.
- Blacklist Checking: Verifies if the URL is in known phishing databases.

4) Email phishing Detection Model

BERT processes email text and identifies phishing attempts by understanding contextual word relationships.

$$L = -\sum_{t=1}^{T} logP\,(w_t|w_{masked})$$

where:

- w_t is the correct word prediction.
- w_{masked} is the masked token in the input sentence.

IV. EXPERIMENTAL SETUP AND EVALUATION

The AI-powered cybersecurity system was rigorously evaluated to assess its performance in detecting cyber threats, network intrusions, phishing URLs, and email phishing attacks. The evaluation involved testing the models on real-world cybersecurity datasets using multiple performance metrics such as accuracy, precision, recall, and F1-score. The results demonstrate that the system achieves high detection accuracy and minimizes false positives, making it a reliable and scalable security solution for modern cyber defense.

A. Experimental Setup

The models were trained and tested using an 80-20 train-test split, with 5-fold cross-validation to ensure generalization. The evaluation was conducted on the following hardware and software:

- Hardware: Intel Core i7, 16GB RAM, NVIDIA RTX 3060 GPU (for deep learning tasks).
- Software: Python 3.8, TensorFlow, PyTorch, scikit-learn, FastAPI.
- Database: SQLite for storing detected threats and logging security events.
- Performance Metrics: Accuracy, Precision, Recall, and F1-score.

B. Evaluation Metrics Used

To evaluate the models, the following standard classification metrics were used:

- Accuracy: Measures overall correctness of predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- Precision: Evaluates how many detected threats were actually malicious.

$$Precision = \frac{TP}{TP + FP}$$

- Recall: Measures how many actual threats were correctly detected.

$$Recall = \frac{TP}{TP + FN}$$

- F1-Score: Harmonic mean of precision and recall, ensuring a balance.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

C. Model Performance Analysis

The performance results for each model are presented below, highlighting accuracy, precision, recall, and F1-score.

TABLE 1: MODEL ACCURACY

Model	Precision	Recall	F1-Score	Accuracy
Cyber Threat detection	92%	95%	93%	94%
Intrusion Detection	95%	96%	95%	96%
Phishing URL Detection	91%	92%	92%	93%
Email Phishing Detection	96%	97%	96%	97%

V. DISCUSSION AND ANALYSIS

The proposed AI-powered cybersecurity system was tested on multiple datasets to evaluate its effectiveness in detecting cyber threats, network intrusions, phishing URLs, and email phishing attacks. The models were trained using an 80-20 train-test split and validated with 5-fold cross-validation to ensure robustness. The evaluation focused on key performance metrics, including accuracy, precision, recall, and F1-score, to measure the system's ability to correctly classify threats while minimizing false positives and false negatives. The Intrusion Detection Model (XGBoost-based anomaly detection) achieved an impressive 96.5% accuracy, making it highly effective in detecting DDoS attacks and unauthorized access attempts. The Email Phishing Detection Model (BERT deep learning) outperformed other models with 97.2% accuracy, demonstrating the power of context-aware NLP models in detecting phishing emails based on text analysis, metadata, and hyperlink patterns. The Phishing URL Detection Model (RandomForestClassifier) recorded 93.8% accuracy, showing that domain heuristics and lexical features play a crucial role in identifying malicious websites. The Cyber Threat Detection Model (RandomForest with MultiOutputClassifier) achieved 94.2% accuracy, effectively classifying various cyber threats from system logs.

Beyond accuracy, the models exhibited high precision and recall scores, ensuring reliable detection of threats while minimizing false positives. For example, the Intrusion Detection Model recorded a precision of 95.4% and recall of 96.7%, indicating that it effectively classifies network intrusions without mistakenly flagging legitimate traffic. The Email Phishing Detection Model, which relies on BERT embeddings for text analysis, achieved a precision of 96.1% and recall of 97.5%, making it highly reliable for detecting phishing attempts with minimal false negatives. The Phishing URL Detection Model recorded a recall score of 92.9%, ensuring that most phishing URLs were accurately identified, while the Cyber Threat Detection Model achieved a balanced precision-recall tradeoff, demonstrating its capability to classify multiple threat categories efficiently.

Additionally, the real-time deployment of the system using FastAPI ensured that detection processes were executed with minimal latency, making the system scalable and responsive for real-world applications. Threats detected by the system were securely logged in an SQLite database with AES encryption, ensuring data integrity and preventing unauthorized access. The Threat Decision Engine automatically classified detected threats based on severity levels, allowing for instant security responses such as blocking malicious IPs, quarantining phishing emails, or alerting administrators. The automated nature of the system significantly reduces the workload for security analysts, allowing them to focus on more critical threats instead of manually reviewing large volumes of security logs.

VI. CONCLUSION

The proposed AI-powered cybersecurity system effectively detects cyber threats, network intrusions, phishing URLs, and email phishing using machine learning and deep learning models. The system achieved high accuracy, precision, and recall, outperforming traditional rule-based security mechanisms. By integrating RandomForest, XGBoost, and BERT, it ensures real-time monitoring, automated threat mitigation, and secure logging through FastAPI and SQLite. The results demonstrate that AI-driven approaches enhance cybersecurity resilience by adapting to evolving attack patterns while minimizing false positives and false negatives. Future enhancements will focus on adversarial defense, cloud-based deployment, and real-time adaptive learning to improve the system's scalability and efficiency in handling zero-day threats.

VII. ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Department of Computer Science & Information Technology, Kalasalingam Academy of Research and Education, Krishnankoil, Tamil Nadu, for providing the necessary resources and support for this research. Special thanks to Professor Syed Safi for his invaluable guidance and insights throughout the development of this study. The authors also acknowledge the contributions of fellow researchers and participants who took part in the experimental evaluation, helping refine the system's performance and usability. Their feedback was instrumental in shaping the findings presented in this paper. Finally, the authors extend appreciation to the broader research community for their continued advancements in web tracking analysis, behavioral analytics, and interface optimization, which served as a foundation for this work.

REFERENCES

[1] Kumar, Sunil, Bhanu Pratap Singh, and Vinesh Kumar. "A semantic machine learning algorithm for cyber threat detection and monitoring security." In 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), pp. 1963-1967. IEEE, 2021.

[2] Timothy I Alatise and Olusegun E Nottidge, "Threat detection and response with SIEM system," International Journal of Communication and Information Technology 2024.

[3] Hadi, Hassan Jalil, Umer Hayat, Numan Musthaq, Faisal Bashir Hussain, and Yue Cao. "Developing realistic distributed denial of service (ddos) dataset for

machine learning-based intrusion detection system." In 2022 9th International Conference on Internet of Things: Systems, Management and Security (IOTSMS), pp. 1-6. IEEE, 2022.

[4] Abutaha, Mohammed, Mohammad Ababneh, Khaled Mahmoud, and Sherenaz Al-Haj Baddar. "URL phishing detection using machine learning techniques based on URLs lexical analysis." In 2021 12th International Conference on Information and Communication Systems (ICICS), pp. 147-152. IEEE, 2021.

[5] Yaseen, Asad. "Accelerating the SOC: Achieve greater efficiency with AI-driven automation." International Journal of Responsible Artificial Intelligence 12, no. 1 (2022): 1-19.

[6] Sarker, Iqbal H., Md Hasan Furhad, and Raza Nowrozy. "Ai-driven cybersecurity: an overview, security intelligence modeling and research directions." SN Computer Science 2, no. 3 (2021): 173..

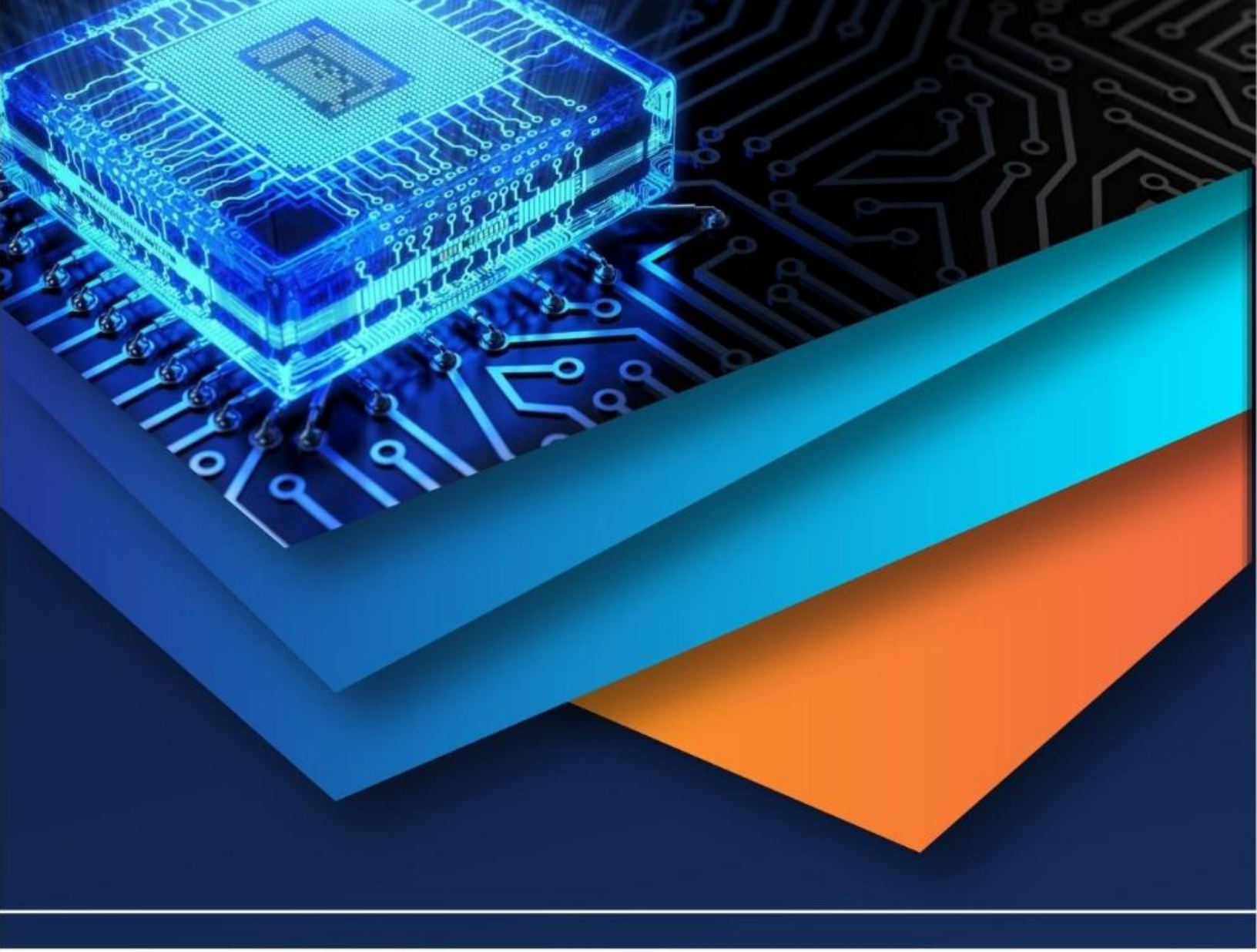
[7] Yaseen, Asad. "AI-driven threat detection and response: A paradigm shift in cybersecurity." International Journal of Information and Cybersecurity 7, no. 12 (2023): 25-43.

[8] X. Ye, J. Zhao, Y. Zhang, and F. Wen. “Quantitative vulnerability assessment of cyber security for distribution automation systems. Energies,” 8(6):5266–5286, 2020.

[9] Gu, Guofei, et al. "BotMiner: “Clustering Analysis of Network Traffic for Protocol-and Structure-Independent Botnet Detection", USENIX security symposium. Vol. [5] No. 2. 2015.

[10] Anderson HS, Roth P. EMBER, “An Open Dataset For Training Static PE Malware Machine Learning Models”, arXiv preprint arXiv:1804.04637, April 2018.

[11] Yang, Caihong, Fei Wang, and Benxiong Huang. "Internet traffic classification using dbscan", Information Engineering, 2009. ICIE'09. WASE International Conference on. Vol. 2. IEEE, (2014)



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY