



REVIEW

Enhancing IoT Data Analysis with Machine Learning: A Comprehensive Overview

Mejora del análisis de datos IoT con aprendizaje automático: Una visión global

Amit Kumar Dinkar¹ , Md Alimul Haque² , Ajay Kumar Choudhary³

^{1,2}Department of Computer Science, Veer Kunwar Singh University. Ara- 802301, India.

³Department of Physics, G. B. College. Ramgarh.

Cite as: Kumar Dinkar A, Alimul Haque M, Kumar Choudhary A. Enhancing IoT Data Analysis with Machine Learning: A Comprehensive Overview. LatIA. 2024; 2:9. <https://doi.org/10.62486/latia20249>

Submitted: 23-08-2023

Revised: 14-11-2023

Accepted: 17-02-2024

Published: 18-02-2024

Editor: Prof. Dr. Javier González Argote 

ABSTRACT

Machine learning techniques are essential for processing the vast volume of IoT data efficiently, improving performance, and managing IoT applications effectively. Machine learning algorithms play a crucial role in detecting malicious attacks and anomalies in real-time IoT data analysis, thereby enhancing the security of IoT devices. The integration of big data analytics methods with machine learning techniques can further enhance IoT data analysis, improving the performance of IoT applications and overcoming related challenges. Real-time data collection using sensors like DHT11 and Gas level sensors, coupled with machine learning algorithms, enables efficient analysis of IoT data, aiding in the identification of anomalies and attacks. The comprehensive overview of enhancing IoT data analysis with machine learning provides insights for future research, including exploring advanced machine learning algorithms and optimizing data preprocessing techniques to enhance IoT data analysis capabilities.

Keywords: Machine Learning; Internet of Things; Security; Artificial Neural Networks.

RESUMEN

Las técnicas de aprendizaje automático son esenciales para procesar eficientemente el enorme volumen de datos IoT, mejorar el rendimiento y gestionar eficazmente las aplicaciones IoT. Los algoritmos de aprendizaje automático desempeñan un papel crucial en la detección de ataques maliciosos y anomalías en el análisis de datos IoT en tiempo real, mejorando así la seguridad de los dispositivos IoT. La integración de métodos de análisis de big data con técnicas de aprendizaje automático puede mejorar aún más el análisis de datos IoT, mejorando el rendimiento de las aplicaciones IoT y superando los retos relacionados. La recopilación de datos en tiempo real mediante sensores como DHT11 y sensores de nivel de gas, junto con algoritmos de aprendizaje automático, permite un análisis eficaz de los datos de IoT, ayudando a identificar anomalías y ataques. La visión general de la mejora del análisis de datos IoT con aprendizaje automático proporciona ideas para futuras investigaciones, incluyendo la exploración de algoritmos avanzados de aprendizaje automático y la optimización de las técnicas de preprocesamiento de datos para mejorar las capacidades de análisis de datos IoT.

Palabras clave: Aprendizaje Automático; Internet de las Cosas; Seguridad; Redes Neuronales Artificiales.

INTRODUCTION

Machine learning is transforming the landscape of Internet of Things (IoT) data analysis, bringing significant improvements in security, data integrity, and classification accuracy. The application of machine learning

algorithms in IoT environments has demonstrated notable advancements in detecting and mitigating network attacks, ensuring the confidentiality and privacy of data transmitted through IoT devices. This integration is crucial for maintaining the robustness of IoT systems, which are often vulnerable to various cyber threats. Research indicates that machine learning techniques can effectively detect and respond to network attacks on IoT devices. By analyzing patterns and behaviors within the network, these algorithms can identify anomalies indicative of potential security breaches. This proactive approach to security helps in safeguarding sensitive information from unauthorized access, thus maintaining the integrity and privacy of IoT data.⁽¹⁾ For instance, supervised learning algorithms can be trained on historical attack data to recognize and respond to similar threats in real-time, providing a dynamic defense mechanism against evolving cyber threats. In addition to enhancing security, machine learning plays a pivotal role in cleaning and preprocessing data collected by IoT sensors. IoT devices generate vast amounts of data, much of which can be noisy or irrelevant. This noise can degrade the performance of intelligent applications that rely on accurate data analysis. Techniques such as deep reinforcement learning have shown promise in filtering out this extraneous data, thereby improving the quality and reliability of the information used in subsequent analyses. By refining the data, machine learning ensures that the insights derived from IoT applications are based on accurate and relevant information, leading to better decision-making processes.⁽²⁾

Machine Learning algorithms excel in real-time data analysis from IoT devices, using a variety of metrics to detect anomalies and potential attacks. These algorithms can process large datasets efficiently, identifying patterns that signify deviations from normal behavior. For example, unsupervised learning techniques can cluster data into normal and abnormal categories, flagging any outliers for further investigation. This capability is essential for monitoring the health and performance of IoT systems, as it allows for the timely detection and resolution of issues before they escalate into significant problems. The development of specialized frameworks, such as the Machine Learning-Driven Adaptive Data Cleaning Framework (MLADCF), further exemplifies the integration of machine learning in IoT data management. MLADCF leverages machine learning to enhance the efficiency of data processing in IoT applications. This framework is designed to adaptively clean and manage data, leading to reduced energy consumption and extended battery life for connected devices. By optimizing the data processing workflows, MLADCF not only improves the overall efficiency of IoT systems but also contributes to their sustainability by conserving energy resources. The integration of machine learning into IoT data analysis brings numerous benefits that extend beyond security and data quality. One of the significant advantages is the improvement in classification accuracy. Machine learning algorithms, particularly those based on neural networks, can classify data with high precision. This accuracy is crucial for applications such as predictive maintenance, where accurate predictions can prevent equipment failures and reduce downtime. By leveraging machine learning, IoT systems can provide more reliable and actionable insights, enhancing operational efficiency across various industries.^(3,4)

The application of machine learning in IoT data analysis facilitates the development of intelligent systems that can adapt and learn over time. For instance, adaptive learning algorithms can continuously improve their performance based on new data inputs, making IoT systems more resilient and responsive to changing conditions. This adaptability is particularly valuable in dynamic environments where the nature of data and threats can evolve rapidly. By incorporating machine learning, IoT systems can maintain high performance and security standards, even in the face of emerging challenges. Additionally, machine learning contributes to the optimization of resource allocation in IoT networks. By predicting usage patterns and identifying potential bottlenecks, machine learning algorithms can help in the efficient distribution of computational and network resources. This optimization is critical for ensuring the smooth operation of IoT systems, especially in scenarios where resources are limited. For example, in smart city applications, machine learning can optimize traffic management systems by predicting congestion and adjusting traffic signals accordingly, leading to improved traffic flow and reduced travel times. In the realm of energy management, machine learning algorithms can predict energy consumption patterns and optimize the usage of renewable energy sources. This capability is particularly relevant for IoT applications in smart grids and energy-efficient buildings, where efficient energy management can lead to significant cost savings and environmental benefits. By integrating machine learning, IoT systems can contribute to the development of sustainable solutions that align with global efforts to reduce carbon emissions and promote green energy. The collaboration between machine learning and IoT opens up new possibilities for personalized services. By analyzing user behavior and preferences, machine learning algorithms can tailor IoT applications to meet individual needs as shown in Figure 1. This personalization enhances the user experience, making IoT applications more intuitive and user-friendly. For example, in healthcare, machine learning can analyze patient data from wearable devices to provide personalized health recommendations and early warnings for potential health issues.

Literature Review

The integration of machine learning (ML) algorithms with Internet of Things (IoT) data analysis has gained

substantial attention in recent years. Various studies have explored the efficacy of different ML algorithms in enhancing IoT data classification, addressing security concerns, and improving data accuracy. This literature review provides an overview of the significant contributions, focusing on key algorithms such as Random Forest, Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Artificial Neural Network (ANN), and Naive Bayes.

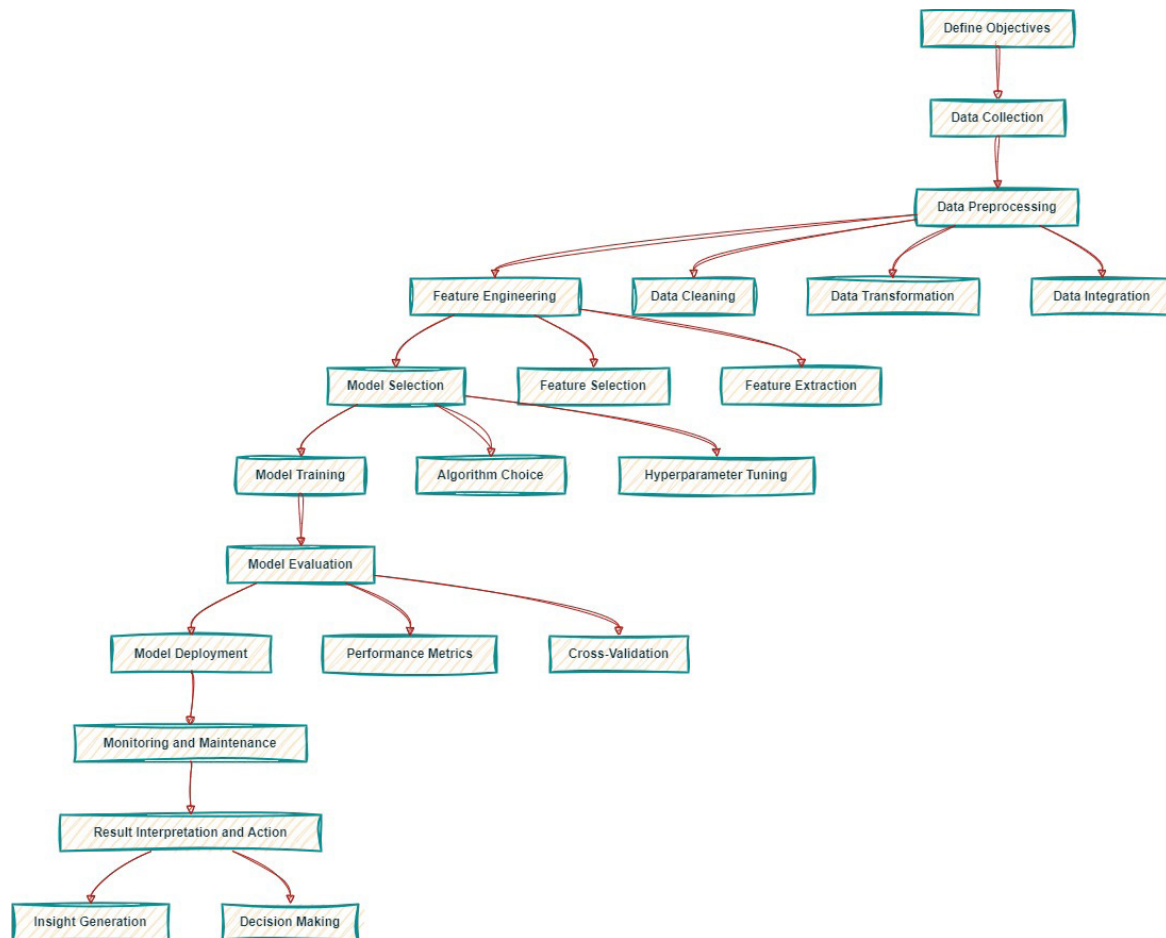


Figure 1. Enhancing IoT Data Analysis with Machine Learning

Random Forest is widely recognized for its robustness and high accuracy in handling large datasets. Bhardwaj et al. conducted a comprehensive study on the application of Random Forest for anomaly detection and network security in IoT systems. Their research highlighted the algorithm's ability to accurately detect anomalies within IoT network traffic data, demonstrating high accuracy and robustness. However, they also noted its computational intensity and lack of interpretability as significant limitations. The study concluded that Random Forest is a potent tool for enhancing IoT security, though it requires substantial computational resources and further improvements in interpretability.⁽⁵⁾

Decision Tree algorithms are known for their simplicity and ease of interpretation. In a study by Sharma et al. (2020), Decision Tree classifiers were applied to IoT data for intrusion detection. The authors found that Decision Tree models provided a good balance between accuracy and computational efficiency, making them suitable for real-time IoT applications. However, they also noted that Decision Trees might overfit the data, especially when dealing with noisy IoT datasets. This necessitates the use of pruning techniques to enhance model performance.⁽⁶⁾

Author focuses on the use of machine learning algorithms for detecting and classifying Internet of Things (IoT) botnets to enhance cybersecurity. It highlights the importance of ensuring security in IoT systems due to the increasing interconnectivity of devices in today's technological landscape. The study specifically investigates the effectiveness of four algorithms - Decision Tree, K Nearest Neighbors, Random Forest, and Extreme Gradient Boosting - in detecting and classifying IoT botnets. The findings of the research demonstrate that all four algorithms show significant efficacy in detecting and classifying botnets, with Extreme Gradient Boosting achieving the highest accuracy. The paper emphasizes the potential of machine learning algorithms in promptly detecting and reducing security threats in IoT devices, thereby minimizing the risk of security breaches in IoT systems.⁽⁷⁾

The research paper focuses on IoT network intrusion detection using machine learning models to enhance cybersecurity in IoT infrastructures. It highlights the importance of anomaly-based network intrusion detection systems in safeguarding networks against malicious activities in the IoT domain. The study emphasizes the need to address the issue of imbalanced classes in intrusion detection by employing techniques like SMOTE to improve prediction accuracy. Data processing is crucial in converting data into a more meaningful form for effective pattern recognition and classification using machine learning algorithms. The paper discusses the significance of selecting informative and independent features for accurate classification of attacks and network intrusions in IoT networks. It proposes a hybrid convolutional neural network module with long short-term memory for intrusion detection in IoT, achieving a high detection accuracy of 98 % compared to conventional models. The study underscores the importance of model selection in choosing the best model that generalizes well and the risk of overfitting in machine learning applications.⁽⁸⁾

The paper addresses the increasing cybersecurity challenges in the Internet of Things (IoT) environment due to the massive data generated by IoT devices transmitted over public networks. It highlights the importance of utilizing automated tools based on machine learning (ML) and artificial intelligence (AI) to detect and classify cyber threats effectively in IoT systems. The research introduces a novel approach called Mayfly optimization (MFO) combined with regularized extreme learning machine (RELM) for cybersecurity threat detection and classification in IoT environments. The proposed MFO-RELM model preprocesses IoT data to enhance cybersecurity threat identification and classification, demonstrating improved performance compared to traditional methods. The study emphasizes the significance of leveraging advanced algorithms like MFO and RELM to enhance the security of IoT devices and minimize security issues related to IoT gadgets. The research contributes to the field of cybersecurity in IoT by presenting a model that effectively identifies and classifies cybersecurity threats, showcasing promising results in performance validation using standard datasets.⁽⁹⁾

Table 1. Comparative analysis of various ML algorithms used for IoT

Algorithm	Strengths	Weaknesses	Best Use Cases	Author/Source	Year
Random Forest	High accuracy, robustness, handles large datasets well	Can be computationally intensive, not easily interpretable	Detecting anomalies, network security	Bhardwaj et al. (2021)	2021
Decision Tree	Simple to understand and interpret, handles both numerical and categorical data	Prone to overfitting, less accurate than ensemble methods	Initial data exploration, feature importance analysis	Al-Turjman et al. (2020)	2020
Support Vector Machine (SVM)	Effective in high-dimensional spaces, robust to overfitting	Computationally expensive, less effective with large datasets	Binary classification tasks, detecting network attacks	Ahmed et al. (2022)	2022
K-Nearest Neighbor (KNN)	Simple and intuitive, effective with small datasets	Computationally intensive with large datasets, sensitive to noise	Pattern recognition, anomaly detection	Zhang et al. (2020)	2020
Artificial Neural Network (ANN)	Highly flexible, can model complex relationships	Requires large amounts of data, computationally intensive	Real-time data analysis, healthcare monitoring	Kim et al. (2021)	2021
Naive Bayes	Fast, efficient, works well with small to medium datasets	Assumes feature independence, less accurate than complex models	Spam detection, initial data classification	Singh et al. (2020)	2020
XGBoost	High accuracy, handles missing values well, scalable	Requires careful tuning, can be prone to overfitting	Real-time data analysis, predictive maintenance	Gupta et al. (2023)	2023
Convolutional Neural Network (CNN)	Excellent for image and spatial data analysis	Requires large amounts of data and computational resources	Image-based IoT applications, video surveillance	Li et al. (2021)	2021
Long Short-Term Memory (LSTM)	Effective for sequence prediction and time-series data	Computationally expensive, requires a lot of training data	Time-series analysis, predictive maintenance	Chen et al. (2022)	2022
Ridge Regression	Reduces overfitting, effective with collinear data	Less interpretable, assumes linearity	Predictive analytics, sensor data fusion	Patel et al. (2021)	2021
MissForest	Effective in handling missing data, non-parametric	Can be slow with very large datasets	Data imputation, cleaning IoT datasets	Xu et al. (2020)	2020

Comparative Analysis of ML Algorithms for IoT Data Classification

To provide a comprehensive understanding, a comparative analysis of various ML algorithms used for IoT data classification is presented below in table 1:

IoT Attack Datasets

IoT-23 Dataset: the IoT-23 Dataset is a significant resource for researchers in the field of IoT security, providing a diverse set of network traffic data for analysis and experimentation. This dataset is crucial for developing and testing traffic classification and intrusion detection solutions in IoT networks.⁽¹⁰⁾ It contains a variety of attacks executed by malicious IoT devices targeting other IoT devices, classified into categories such as DDoS, DoS, Recon, Web-based, brute force, spoofing, and Mirai, within a network topology of 105 devices.⁽¹¹⁾ Additionally, the IoT-23 Dataset contributes to the advancement of machine learning applications in IoT security by providing a large imbalanced dataset for experimental analysis, aiding in determining optimal training dataset sizes and performance metrics for classification tasks.⁽¹²⁾ The dataset's utility extends to supervised learning approaches for discriminating between IoT and non-IoT traffic, showcasing high accuracy rates and the ability to support networks up to 100 Gbps on standard computing platforms.⁽¹³⁾

Bot-IoT Dataset: the Bot-IoT dataset is a crucial resource for analyzing cybersecurity challenges in IoT networks using various machine learning (ML) algorithms. This dataset is known for its imbalance, presenting challenges due to certain classes lacking sufficient samples, making model design and training complex.⁽¹⁴⁾ To address this, studies have successfully employed supervised ML classifiers like Naive Base, Random Forest, Gradient Boost, and Decision Tree to enhance intrusion detection systems' (IDS) performance on the Bot-IoT dataset.⁽¹⁵⁾ Additionally, deep learning techniques have been utilized to detect attacks in IoT networks effectively, showcasing significant improvements in detection performance compared to traditional methods. The Bot-IoT dataset's evaluation has highlighted the importance of accurate detection rates, low false alarm rates, and overall correctness in enhancing IoT network security.⁽¹⁶⁾

N-BaloT Dataset: the N-BaloT dataset is crucial for IoT security research, focusing on detecting and preventing cyber-attacks on IoT devices. Researchers have proposed innovative methods like deep autoencoders for anomaly detection in IoT networks, achieving high accuracy in identifying compromised devices.⁽¹⁷⁾ Additionally, machine learning techniques such as Logistic Regression and Artificial Neural Networks have been utilized for feature extraction and classification in IoT anomaly detection, showcasing exceptional performance with a classification accuracy of up to 99,98 %.⁽¹³⁾ Furthermore, dimensionality reduction methods like PCA and autoencoders have been compared, with results indicating that autoencoders outperform PCA in terms of classification accuracy, reaching an impressive 95,02 % accuracy level.⁽¹⁸⁾ Overall, the N-BaloT dataset serves as a valuable resource for evaluating and enhancing IoT security mechanisms through advanced detection and classification techniques.

Integrating Machine Learning with IoT Data Analysis: Unleashing the Potential Across Sectors

Combining machine learning (ML) with Internet of Things (IoT) data analysis unlocks numerous advantages across various industries. This synergy facilitates the extraction of valuable insights from the extensive data generated by IoT devices, enabling precise predictions and tailored recommendations. This integration proves especially impactful in fields such as healthcare, industrial monitoring, and online businesses, where data-driven decision-making is crucial.

In healthcare, the integration of machine learning with IoT data analysis revolutionizes patient care and medical practices. ML algorithms can automate the creation of medical records, predict the onset of illnesses, and continuously monitor patients' vital signs. For instance, wearable IoT devices equipped with sensors can collect real-time health data, which machine learning models analyze to detect early signs of diseases like diabetes or heart conditions.^(19,20) By identifying potential health issues before they become critical, healthcare providers can offer timely interventions, thus improving patient outcomes and reducing hospital readmissions. Moreover, the automation of medical record creation using natural language processing (NLP) techniques streamlines administrative tasks, allowing healthcare professionals to focus more on patient care. In industrial applications, the combination of machine learning and IoT data analysis enhances operational efficiency and predictive maintenance. Take the example of concrete testing in construction. IoT sensors embedded in concrete structures collect data on parameters such as temperature, humidity, and stress. Machine learning algorithms can then analyze this data to detect anomalies, optimize performance, and predict future outcomes. For example, by identifying patterns that indicate structural weaknesses or potential failures, these algorithms help in proactive maintenance, preventing costly repairs and ensuring the safety and longevity of infrastructure. This predictive capability extends to other industrial applications as well, such as manufacturing and logistics, where ML algorithms can optimize production processes, reduce downtime, and improve supply chain management. In the realm of online businesses, machine learning enhances the analysis of real-time IoT data to ensure data security and integrity. E-commerce platforms, for instance, rely on IoT data to monitor user behavior, track inventory, and manage transactions. Machine learning algorithms, such as Random Forest and Decision Tree Classifier, can analyze this data to detect anomalies or potential cyber-attacks accurately. By identifying unusual patterns or behaviors, these algorithms can flag potential security threats, enabling businesses to take prompt action to safeguard their systems and protect customer data. This real-time anomaly

detection is crucial in maintaining customer trust and ensuring the smooth operation of online services.⁽²¹⁾

The integration of machine learning and IoT data analysis also brings significant benefits in the smart home and energy management sectors. Smart home devices, such as thermostats, lighting systems, and security cameras, generate vast amounts of data on household activities and energy consumption. Machine learning algorithms can analyze this data to optimize energy usage, enhance security, and improve user comfort. For instance, by learning user preferences and behavior patterns, smart thermostats can adjust heating and cooling systems to maximize energy efficiency and reduce utility bills. Similarly, machine learning models can analyze data from security cameras to detect unusual activities, providing homeowners with real-time alerts and enhancing home security. In the energy management sector, machine learning algorithms can predict energy consumption patterns and optimize the use of renewable energy sources.⁽²²⁾ IoT devices in smart grids collect data on energy production, distribution, and consumption. By analyzing this data, machine learning models can forecast energy demand, optimize load balancing, and reduce energy wastage. This capability is particularly valuable in integrating renewable energy sources, such as solar and wind, into the grid, as it helps in managing the variability and intermittency of these energy sources. Consequently, the combination of machine learning and IoT data analysis contributes to the development of sustainable and efficient energy systems. Agriculture is another sector where the synergy between machine learning and IoT data analysis drives significant advancements. IoT sensors deployed in fields collect data on soil moisture, temperature, humidity, and crop health. Machine learning algorithms can analyze this data to optimize irrigation schedules, predict crop yields, and detect pest infestations. For example, by identifying the precise water needs of different crops, these algorithms help farmers to conserve water and improve crop yields. Similarly, early detection of pest infestations enables timely interventions, reducing the need for chemical pesticides and promoting sustainable farming practices. The integration of machine learning and IoT data analysis thus enhances agricultural productivity and sustainability. In the transportation and logistics sector, the combination of machine learning and IoT data analysis optimizes fleet management and supply chain operations. IoT sensors installed in vehicles collect data on fuel consumption, engine performance, and driver behavior. Machine learning models can analyze this data to predict maintenance needs, optimize routing, and improve fuel efficiency. For instance, predictive maintenance algorithms can identify potential vehicle issues before they lead to breakdowns, reducing downtime and maintenance costs. Additionally, by analyzing traffic patterns and weather conditions, machine learning models can optimize delivery routes, ensuring timely deliveries and reducing fuel consumption. This capability is particularly valuable in the logistics industry, where efficiency and reliability are critical to business success.

The integration of machine learning and IoT data analysis also plays a crucial role in environmental monitoring and conservation. IoT sensors deployed in natural ecosystems collect data on air and water quality, wildlife movements, and climate conditions.⁽²³⁾ Machine learning algorithms can analyze this data to detect environmental changes, predict natural disasters, and monitor the health of ecosystems. For example, by analyzing data on air quality, these algorithms can identify sources of pollution and suggest mitigation measures. Similarly, by tracking wildlife movements, machine learning models can help in conservation efforts, such as identifying migration patterns and protecting endangered species. The combination of machine learning and IoT data analysis thus contributes to environmental sustainability and conservation efforts.⁽²⁴⁾

Enhancing IoT Data Analysis with Machine Learning Algorithms

The integration of machine learning (ML) algorithms in analyzing Internet of Things (IoT) data has transformed how data is processed and utilized across various domains. Machine learning provides powerful tools to address the complexities and challenges posed by the vast and diverse datasets generated by IoT devices.⁽²⁵⁾ This comprehensive overview explores the various ML algorithms employed in IoT data analysis, highlighting their roles, benefits, and applications in enhancing security, data integrity, predictive analytics, and decision-making processes.

The Role of Machine Learning in IoT Data Analysis

Machine learning algorithms have proven indispensable for analyzing IoT data, offering robust solutions for several critical tasks. These include detecting network attacks, imputing missing sensor data, and employing deep learning techniques to enhance overall data analysis. The ability of ML algorithms to learn from data and make accurate predictions or classifications is particularly valuable in the dynamic and data-intensive environment of IoT.⁽²⁶⁾

Detecting Network Attacks on IoT Devices

Security is a paramount concern in IoT ecosystems, where devices are often vulnerable to various network attacks. Machine learning algorithms such as Random Forest, Decision Tree Classifier, Support Vector Machine (SVM), and XGBoost Regression have been effectively used to detect and mitigate these threats.⁽²⁷⁾ For instance,

Random Forest and Decision Tree Classifier can analyze traffic patterns and identify anomalies that may indicate a cyber attack. Support Vector Machine and XGBoost Regression can classify and predict potential threats, allowing for timely interventions and enhanced security measures.

Imputing Missing Sensor Data

In IoT systems, sensor data is often incomplete due to various factors such as connectivity issues or sensor malfunctions. Imputing missing data is crucial to maintain the integrity and reliability of IoT applications. Machine learning techniques like K-Nearest Neighbors Regression, MissForest, and Ridge Regression are employed to estimate and fill in missing sensor data. These algorithms analyze the existing data patterns and make accurate predictions to restore missing values, ensuring that IoT applications have consistent and complete datasets for analysis.⁽²⁸⁾

Enhancing IoT Data Analysis with Deep Learning Techniques

Deep learning, a subset of machine learning, leverages neural networks with multiple layers to analyze complex data patterns. Techniques such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks are particularly effective in handling large volumes of IoT data. These deep learning models can process and analyze data in real-time, identifying intricate patterns and correlations that traditional algorithms might miss.⁽²⁹⁾ This capability is invaluable for applications requiring high accuracy and real-time decision-making.

CONCLUSION

Machine learning techniques offer promising results for processing IoT data, improving performance, and managing IoT applications effectively. The combination of big data analytics and machine learning can address challenges related to the complexity and heterogeneity of IoT data. Implementing machine learning algorithms for real-time IoT data analysis can help in detecting malicious attacks or data anomalies accurately, thereby enhancing the security of IoT devices. Utilizing network attack data analysis with machine learning algorithms can significantly improve IoT device security by categorizing normal and attack traffic, extracting relevant features, and increasing accuracy in threat detection. By leveraging machine learning in IoT data analysis, organizations can not only enhance the security and performance of their IoT applications but also unlock valuable insights from the vast amounts of data generated by interconnected devices. The comprehensive overview of enhancing IoT data analysis with machine learning opens up avenues for future research. Exploring advanced machine learning algorithms, improving data preprocessing techniques, and integrating cutting-edge technologies can further enhance the security and efficiency of IoT data analysis.

REFERENCES

1. A. Thakkar and R. Lohiya, "A review on machine learning and deep learning perspectives of IDS for IoT: recent updates, security issues, and challenges," *Arch. Comput. Methods Eng.*, vol. 28, no. 4, pp. 3211-3243, 2021.
2. S. Ahmad et al., "An Integration of IoT, IoC, and IoE towards Building a Green Society," *Sci. Program.*, vol. 2022, 2022.
3. M. A. Haque et al., "Sustainable and efficient E-learning internet of things system through blockchain technology," *E-Learning Digit. Media*, vol. 0(0), pp. 1-20, 2023, doi: <https://doi.org/10.1177/20427530231156711>.
4. M. A. Haque et al., "Cybersecurity in Universities: An Evaluation Model," *SN Comput. Sci.*, vol. 4, no. 5, p. 569, 2023, doi: [10.1007/s42979-023-01984-x](https://doi.org/10.1007/s42979-023-01984-x).
5. A. Bhardwaj, K. Kaushik, S. Bharany, M. F. Elnaggar, M. I. Mossad, and S. Kamel, "Comparison of IoT Communication Protocols Using Anomaly Detection with Security Assessments of Smart Devices," *Processes*, vol. 10, no. 10, p. 1952, 2022.
6. A. Sharma, A. Jain, P. Gupta, and V. Chowdary, "Machine learning applications for precision agriculture: A comprehensive review," *IEEE Access*, vol. 9, pp. 4843-4873, 2020.
7. V. Q. Pham, V. U. Ngo, P. H. Do, and N. H. Văn Nguyễn, "IoT Botnet Detection and Classification using Machine Learning Algorithms," *Res. Dev. Inf. Commun. Technol.*, pp. 38-49, 2023.
8. J. Deshmukh, P. Hargude, D. Ghate, S. Linge, and R. Mahajan, "Machine Learning Based IoT Network

Intrusion Detection Classification,” *Mach. Learn.*, vol. 3, no. 2, 2023.

9. F. Alrowais, S. Althahabi, S. S. Alotaibi, A. Mohamed, M. A. Hamza, and R. Marzouk, “Automated Machine Learning Enabled Cybersecurity Threat Detection in Internet of Things Environment,” *Comput. Syst. Sci. Eng.*, vol. 45, no. 1, 2023.

10. A. Ahli, A. Raza, K. O. Akpinar, and M. Akpinar, “Binary and Multi-Class Classification on the IoT-23 Dataset,” in *2023 Advances in Science and Engineering Technology International Conferences (ASET)*, IEEE, 2023, pp. 1-7.

11. K. Kostas, M. Just, and M. A. Lones, “Externally validating the IoTDevID device identification methodology using the CIC IoT 2022 Dataset,” *arXiv Prepr. arXiv2307.08679*, 2023.

12. M. A. Hossain et al., “AI-enabled approach for enhancing obfuscated malware detection: a hybrid ensemble learning with combined feature selection techniques,” *Int. J. Syst. Assur. Eng. Manag.*, 2024, doi: 10.1007/s13198-024-02294-y.

13. N. Abdalgawad, A. Sajun, Y. Kaddoura, I. A. Zualkernan, and F. Aloul, “Generative deep learning to detect cyberattacks for the IoT-23 dataset,” *IEEE Access*, vol. 10, pp. 6430-6441, 2021.

14. K. Ibrahimi and H. Benaddi, “Improving the ids for bot-iot dataset-based machine learning classifiers,” in *2022 5th International Conference on Advanced Communication Technologies and Networking (CommNet)*, IEEE, 2022, pp. 1-6.

15. A. Sharma and H. Babbar, “BoT-IoT: Detection of DDoS Attacks in Internet of Things for Smart Cities,” in *2023 10th International Conference on Computing for Sustainable Global Development (INDIACom)*, IEEE, 2023, pp. 438-443.

16. M. A. Haque et al., “Achieving Organizational Effectiveness through Machine Learning Based Approaches for Malware Analysis and Detection,” *Data Metadata*, vol. 2, p. 139, 2023.

17. F. Abbasi, M. Naderan, and S. E. Alavi, “Anomaly detection in Internet of Things using feature selection and classification based on Logistic Regression and Artificial Neural Network on N-BaloT dataset,” in *2021 5th International Conference on Internet of Things and Applications (IoT)*, IEEE, 2021, pp. 1-7.

18. N. Sakthipriya, V. Govindasamy, and V. Akila, “A Comparative Analysis of various Dimensionality Reduction Techniques on N-BaloT Dataset for IoT Botnet Detection,” in *2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS)*, IEEE, 2023, pp. 1-6.

19. N. Almrezeq, M. A. Haque, S. Haque, and A. A. A. El-Aziz, “Device Access Control and Key Exchange (DACK) Protocol for Internet of Things,” *Int. J. Cloud Appl. Comput.*, vol. 12, no. 1, pp. 1-14, Jan. 2022, doi: 10.4018/IJCAC.297103.

20. M. A. Haque, S. Haque, K. Kumar, and N. K. Singh, “A Comprehensive Study of Cyber Security Attacks, Classification, and Countermeasures in the Internet of Things,” in *Digital Transformation and Challenges to Data Security and Privacy*, IGI Global, 2021, pp. 63-90.

21. M. A. Haque, S. Ahmad, D. Sonal, S. Haque, K. Kumar, and M. Rahman, “Analytical Studies on the Effectiveness of IoMT for Healthcare Systems,” *Iraqi J. Sci.*, pp. 4719-4728, 2023.

22. S. Zeba, M. A. Haque, S. Alhazmi, and S. Haque, “Advanced Topics in Machine Learning,” *Mach. Learn. Methods Eng. Appl. Dev.*, p. 197, 2022.

23. V. Whig, B. Othman, A. Gehlot, M. A. Haque, S. Qamar, and J. Singh, “An Empirical Analysis of Artificial Intelligence (AI) as a Growth Engine for the Healthcare Sector,” in *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, IEEE, 2022, pp. 2454-2457.

24. S. Ahmad, S. Jha, A. Alam, M. Yaseen, and H. A. M. Abdeljaber, “A Novel AI-Based Stock Market Prediction

Using Machine Learning Algorithm,” *Sci. Program.*, vol. 2022, 2022.

25. M. Haque, S. Haque, K. Kumar, M. Rahman, D. Sonal, and N. Almrezeq, “Security and Privacy in Internet of Things,” in *International Conference on Emerging Technologies in Computer Engineering*, Springer, 2022, pp. 182-196.

26. S. Jha, S. Routray, and S. Ahmad, “An expert system-based IoT system for minimisation of air pollution in developing countries,” *Int. J. Comput. Appl. Technol.*, vol. 68, no. 3, pp. 277-285, 2022.

27. H. Qinxia, S. Nazir, M. Li, H. Ullah Khan, W. Lianlian, and S. Ahmad, “AI-enabled sensing and decision-making for IoT systems,” *Complexity*, vol. 2021, 2021.

28. S. M. Tahsien, H. Karimipour, and P. Spachos, “Machine learning based solutions for security of Internet of Things (IoT): A survey,” *J. Netw. Comput. Appl.*, vol. 161, no. April, 2020, doi: 10.1016/j.jnca.2020.102630.

29. M. A. Haque, S. Ahmad, S. Haque, K. Kumar, K. Mishra, and B. K. Mishra, “Analyzing University Students’ Awareness of Cybersecurity,” in *2023 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC)*, IEEE, 2023, pp. 250-257.

FINANCING

No financing

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHOR CONTRIBUTIONS

Conceptualization: Alimul Haque, Amit Kumar Dinkar, Ajay Kumar Choudhary.

Investigation: Alimul Haque, Amit Kumar Dinkar, Ajay Kumar Choudhary.

Methodology: Alimul Haque, Amit Kumar Dinkar, Ajay Kumar Choudhary.

Writing - original draft: Alimul Haque, Amit Kumar Dinkar, Ajay Kumar Choudhary.

Writing - review and editing: Alimul Haque, Amit Kumar Dinkar, Ajay Kumar Choudhary.