INTERNATIONAL JOURNAL OF SUSTAINABLE DEVELOPMENT IN COMPUTING SCIENCE
OPEN ACCESS, PEER REVIEWED, REFEREED JOURNAL
ISSN: 3246-544X

Using AI to Identify and Mitigate Cybersecurity Threats in Fog Computing Environments

[1]Mohan Harish Maturi, [2]Srikar Podicheti, [1]Karthik Meduri, [1]Geeta Sandeep Nadella

[1]Department of Information Technology, University of the Cumberlands, Williamsburg, KY, USA

[2]Department of Computer Science, University of the Pacific, Stockton, CA

mmaturi9213@ucumberlands.edu

* corresponding author

JOURNAL INFO

Double Peer Reviewed
Impact Factor: 5.6 (SJR)
Open Access
Refereed Journal

ABSTRACT

Fog computing represents a transformative paradigm shift, bridging the gap between centralized cloud architectures and distributed edge computing environments. This research explores the cybersecurity challenges introduced by the dispersed nature of fog computing and evaluates the effectiveness of artificial intelligence (AI) in mitigating these risks. Traditional security solutions are often inadequate for these dynamic environments. This study examines the applicability of AI-driven techniques of AI machine-learning Models like (Random Forest, Gaussian Naive Bayes, Logistic Regression, and KNN) in addressing these cybersecurity challenges and producing a high accuracy score of prediction of threats. Machine learning models are employed to detect anomalies and potential security breaches in algorithms that analyze complex data streams for subtle deviations, and natural language processing techniques improve the analysis of security logs and communication patterns. The research highlights the strategic value of AI in safeguarding decentralized computing resources and provides a foundation for future exploration and practical implementation of AI-driven cybersecurity measures in fog computing contexts.

ISSN: 3246-544X

Introduction

The evolution of computing paradigms has continually shaped the landscape of digital infrastructure, from the early days of mainframe computing to the ubiquitous presence of cloud services today. In this progression of fog consumes arose the pivotal advancement of bridging the gap between centralized cloud architectures and distributed edge computing environments [1]. The Cisco 2012 for fog computing to spread the competencies of foggy calculating of information handing out and storage to the network edge is nearly faster towards production and expending. This distributed approach alone augments real-time facts dispensation abilities to reduce latency and conserve network bandwidth, which is well-suited for applications in IoT of smart cities and industrial automation. Distributing computing resources to the edge predates the formalization of fog computing [2]. Edge computing principles were outlined in the days of computing; localized processing was essential due to limited connectivity and computing power. Their computing capabilities advanced in centralized cloud architectures became predominant to scalability and cost-efficiency at the expense of latency-sensitive applications and data privacy concerns [1-3]. The resurgence of interest in edge computing in the late 2000s resulted in fog totaling as a strategic extension for addressing the limits of centralized mist and traditional edge computing models.

Adopting fog computing introduces unique cybersecurity challenges for effectively managing and realizing its full benefits. Edge devices and fog nodes activate in miscellaneous and hostile surroundings because centralized data centers lack robust security measures [4]. These vulnerabilities expose edge networks to a spectrum of cyber threats, such as counting malware inoculations for illegal access efforts, information exfiltration, and service disruptions. The lively and varied nature of advantageous surroundings complicates security management by necessitating adaptive and scalable cybersecurity solutions capable of effectively safeguarding distributed computing resources [5]. In their challenges, artificial intelligence (AI) has a keystone tech for boosting the security carriage of fog-calculating environments and using AI/techniques such as machine learning, bottom-less learning, and usual linguistic dispensation for main governments to enhance their cybersecurity defenses with practical danger discovery of anomaly detection, and automated event answer competences. Machine-learning procedures edge devices to examine their data to discern normal behavior patterns and notice deviations revealing possible safety openings [6]. Deep models process high-dimensional data streams and empower real-time anomaly detection across dynamic edge networks. Natural-language dispensation techniques augment the examination of security logs and communication patterns, which are the initial discovery and extenuation of sophisticated cyber dangers in traditional detection methods.

The propagation of the IoT plans and the advent of 5G systems have accelerated the acceptance of fog computing within an era where data is treated and examined at their data source. This shift addresses critical concerns for bandwidth consumption of data privacy, which is supreme in applications requiring real-time decision-making abilities [7]. Their

ISSN: 3246-544X

distributed nature of fog-computing introduces inherent cybersecurity risks that are managed to safeguard the honesty and privacy of data transmitted and treated at an advantage. The decentralized deployment of computing resources across a myriad of edge devices and fog nodes amplifies their exposing organizations to a spectrum of cyber dangers, ranging from urbane malware bouts to insider threats and public industrial exploits.

To mitigate these cybersecurity challenges, organizations increasingly turn to AI methods and AI techniques Like ML and DL [8]. These AI/technologies empower fog environments with advanced threat detection capabilities for identifying and mitigating security incidents before they escalate. Machine models trained on historical data can notice irregularities and designs revealing potential cyber use of deep algorithms for processing and analyzing complex data streams to uncover subtle deviations in evading traditional security measures [9]. Using natural-language processing techniques enhances the analysis of unstructured security logs and communication patterns, facilitating rapid incident response and threat intelligence gathering across distributed edge networks.

1.1 Problem statement

The implementation of fog computing has presented significant cybersecurity challenges to the decentralized countryside of edge devices and fog nodes, which function in varied, aggressive surroundings. These challenges include vulnerabilities to malware injections for illegal admission attempts and data breaches, and facility disruptions exacerbated by edge networks' dynamic and heterogeneous nature [10]. Traditional cybersecurity solutions designed for centralized cloud architectures are inadequate in mitigating these risks and require innovative approaches to improve the security posture of fog-computing atmospheres.

1.2 Research Objective

The main objective of this investigation is to explore and evaluate the efficacy of AI-driven and machine-learning techniques to detect and mitigate cybersecurity threats inside fog computing environments. There are few key objectives are given below:

1. Investigate current cybersecurity challenges in fog-computing and identify common vulnerabilities and threats in edge devices and fog nodes.
2. Assess the applicability and effectiveness of machine-learning representations used for irregularity uncovering algorithms to identify abnormal behaviors and possible safety openings in real-time-edge environments.
3. Proposing and validating a comprehensive framework for integrating AI for cyber security to events tailored to fog computing is a unique challenge to enhance the complete system resilience and protect critical data assets at the system advantage.

2. Literature Review

ISSN: 3246-544X

Fog is a postponement of the cloud to address their latency boundaries and bandwidth associated with centralized data centers for computational assets earlier to the advantage of the system. This devolved approach is helpful for needful real-time dispensation, low dormancy replies, and independent vehicles in the distributed and heterogeneous nature of fog computing to significant cybersecurity experiments [10]. Unlike centralized fog, nodes stay organized in varied and less secure atmospheres for various cyber threats that necessitate healthy security events, data honesty, discretion, and obtainability. The inherent vulnerabilities in fog arise primarily from its decentralized architecture system, which acts in miniature centers at the edge of the networks to become a potential target for attackers. Physical attacks are a significant concern because fog nodes are frequently placed in remote or publicly accessible locations. Attackers can physically tamper with nodes to extract sensitive information about malicious hardware components [9-11]. The frequent announcement is to be made in different nodes for disposed to Man-in-the-Middle occurrences aimed at an attacker captures to alter the numbers being diffused. This not only compromises data integrity but also poses severe risks to the privacy and security of the users. Their Distributed Denial-of-Service (DDoS) bouts for foggy nodes per excessive traffic produce disruptions and render the system incapable of processing authentic requests [12].

Confirmation and access control mechanisms are vital for upholding safety in milieus. Traditional centralized authentication methods are impractical due to fog nodes' distributed nature and resource constraints. Researchers have proposed various decentralized authentication schemes to address these challenges. Blockchain technology offers a dispersed and tamper-proof method for managing identities to secure authentication without relying on a central authority. Attribute-Based Access Control (ABAC) is another approach to fine-grained access control for permissions founded on the attributes of operators; then, assets remain that only official entities container for information or services [13]. These mechanisms are to be lightweight and minimize the computational and energy overhead on fog nodes to maintain their robust security.

Confidentiality is supreme, given the subtle flora of the statistics processed at the edge. Encryption is a fundamental technique for protecting facts together in shipment and repose. These tasks lie in implementing efficient encryption methods that do not significantly impact resource-constrained fog nodes' performance [14]. Homomorphic encryption allows calculations on secret information without waiting for decryption and is emerging as a feasible alternative. This approach protects data privacy in fog nodes while processing data safely. Safe Multi-Party Computing (SMPC) is a different method that enables many parties to calculate input functions to maintain the parameters secret collaboratively. Those that preserve privacy approaches are vital for securing private information in fog language and staying secret to handle unknown nodes.

The dynamic and disseminated wildlife of fog requires advanced Intrusion Discovery and prevention systems (IDPS) that can operate efficiently in settings. Traditional IDPS helps

ISSN: 3246-544X

are intended for centralized systems and are near to being suitable for immediate processing and the resource limitations of fog nodes. Researchers are developing learning and artificial intelligence for anomaly detection [15]. These systems are remarkable in detecting and vindicating cyber intimidations and adapting to evolving attack patterns. Collaborative intrusion detection to share threat intelligence to enhance the overall security of a coordinated response to detected threats. This collaborative approach improves detection accuracy and mitigates attacks more effectively. Secure communiqué among fog nodes and IoT devices is a serious feature of fog computing safety [16].

Standard procedures are Transports-Layers-Security (TLS) and Data-gram Transports-Layer-Security (DTLS), commonly secondhand to secure statistics transmission. The varied and resource-constrained nature of fog environments requires optimized communication protocols. Researchers are helping to develop frivolous and energy-efficient secure communication protocols that are specifically future for fog computing. These protocols aim for robust security to significant computational and energy overhead on the devices [17]. Optimized key management schemes and efficient communication in fog networks are being explored for inconsequential cryptographic algorithms.

2.1 AI Application in Cyber Security

The application in their AI, including machine learning plus DL, NLP consumes revolutionized cybersecurity, gives more dynamic and adaptive for effective solutions to combat cyber threats; figure 1 shows the Application of AI in Cybersecurity. Machine learning enhances anomaly detection and predictive threat intelligence in deep learning, which improves the detection of sophisticated attacks and malware analysis. The efficient analysis of textual data helps in the empathy of threats and the extraction of valuable insights from unstructured information [16-17]. As threats evolve, integrating AI techniques will be central to popular emerging advanced sanctuary measures that can protect against increasingly sophisticated adversaries.

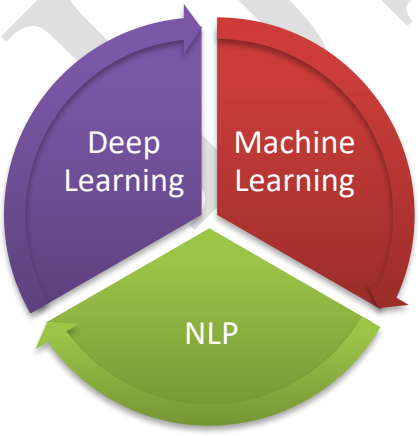

Figure 1: Application of AI in Cybersecurity

ISSN: 3246-544X

2.1.1 Machine Learning

A subset of AI is extensively used in cybersecurity to improve discoveries and preventions besides response to cyber threats. Old-style cyber-security is events on static rules and autographs that are insufficient against sophisticated and evolving threats. Machine learning systems learn from records to classify patterns and make predictions to improve the accuracy and efficiency of threat detection. The unique chief application of ML in cybersecurity is anomaly detection. Their training models on historical network traffic data for algorithms can classify eccentricities from usual conduct that might show potential safety incidents [17]. Clustering and classification techniques are employed to differentiate between benign and spiteful actions. Similar data points can be grouped in gathering algorithms to identify unusual ones, and classification algorithms can label data as normal or anomalous based on learned characteristics. Some further employ applications involve the creation of models of prediction for threat information. It can analyze information from many places in risk streams and forecast prospective threats. A proactive plan in organizations to prepare for and reduce hazards earlier creates disaster. It is utilized in automating tasks for malware detection besides classifying files as malicious or benign from known malware signatures and behaviors [18].

2.1.2 Deep Learning

A more sophisticated version of ML-artificial neural networks are systems that use numerous layers to simulate complicated info and pattern presentations. Its capacity to handle and use informal information makes it useful in cyber-security. One of the key areas for learning excels in detecting advanced persistent threats (APTs) and zero-day attacks. Traditional security measures often struggle with these sophisticated threats due to their novel characteristics and stealthy nature. Deep uses convolutional networks and recurrent neural nets to investigate categorizations of events and recognize shapes that indicate an ongoing or imminent attack [19]. Forecasting systems effectively analyze time-series data and are suitable for detecting anomalous behaviors over time in network traffic.

It is instrumental in ensuring the accuracy of incursion discovery schemes with deep neural networks on massive data sets of network traffic. Besides, attacks can achieve high detection rates with low false positives. They improve their robustness and reliability over time by adapting to new and evolving threats through continuous learning. Threat detection is used in malware analysis within deep belief networks (DBNs), and auto-encoders can mechanically cut topographies from malware to identify previously unknown malware variants. This automatic feature extraction and taxonomy capability significantly improves the productivity and helpfulness of malware recognition procedures.

2.1.3 Natural Language Processing (NLP)

ISSN: 3246-544X

it is an additional significant component of AI studies, the interface between machines and human speech. NLP approaches are used to analyze and understand written information from security records given to warnings in security-related briefings. A popular use of NLP in cybercrime is the evaluation of safety logs and warnings. Safety operation centers (SOCs) overrun data from various sources, including routers for intrusion detection systems of app files [20]. NLP approaches may interpret and analyze unorganized textual input to find useful data to connect incidents and spot abnormalities with recognition of named entities (NER). They may extract Internet Protocol (IP) addresses and URLs of even file hash codes from records to help identify suspect actions.

Another significant application is in the analysis of threat intelligence feeds. Threat intelligence reports often contain valuable information about emerging threats, vulnerabilities, and attack techniques. NLP techniques in topic modeling and sentiment analysis process these reports to extract actionable insights and identify relevant trends of threat intelligence analysis for organizations to give knowledge around the newest intimidations and then regulate their safety measures [11]. NLP is also employed to detect phishing attacks. Phishing emails id subtle linguistic cues that can be challenging for traditional rule-based systems to detect. NLP models can help analyze emails' text to identify suspicious language for urgency and requests for sensitive information or discrepancies in writing style. These practices are text classification and semantic analysis in NLP models, which can accurately distinguish between legitimate and phishing emails to protect users from potential scams.

2.2 Role of AI in Cyber Threat Detections

AI has fundamentally transformed the land of cybersecurity threat detection with progressive techniques beyond traditional methods. Traditional threat detection systems rely on predefined signatures and heuristics to identify malicious activities. While successful versus recognized networks suffer against zero-day vulnerability assaults and advanced threats that change over time. They solve these constraints by examining their massive sizes of numbers to find trends and abnormalities that indicate cyber dangers. Machine-learning methods are trained prior to encompass both typical and harmful to create models capable of detecting departures from predicted trends in methods of strategy detection, grouping, and finding outliers that aid in distinguishing unexpected activity in network data that might show an assault. These models learn from the data, continuously improve their detection capabilities, and are exposed to new information [21]. This adaptive learning process is decisive for identifying emerging threats that do not easily match any known signature.

It improves the detection of threats by using neural networks made up of neurons to represent complex connections in the information in question. Convolutional neural systems (CNNs) and recurrent neural networks (RNNs) are excellent at analyzing time series of information and distinguishing patterns across sequences to ideal near-identify breaches or persistent advanced threats (APTs). Statistical models are to learn from huge quantities of detecting

ISSN: 3246-544X

subtle and complex methods of attack that standard systems could miss. AI's capacity to interpret information in context transforms threat identification management. The time frame for recognizing and reacting to threats has free delay actions that can cause severe harm. AI systems may analyze data flows immediately to determine possible dangers that may arise and initiate quick actions [22].

Malware remains one of the most pervasive threats in cybersecurity, evolving constantly to evade traditional detection methods. AI significantly enhances malware detection and analysis by employing sophisticated techniques to identify malicious software. Traditional antivirus solutions depend on signature uncovering, easily circumvented by polymorphic and metamorphic malware that alter their code to avoid detection. AI is used to identify such evasive malware through advanced feature extraction and pattern recognition [8-23]. During their deep belief in networks (DBNs) and auto-encoders, they automatically learn representations of malware characteristics from raw data. These models can distinguish between benign and malicious files based on learned features, enabling the detection of previously unknown malware variants. AI-powered sandboxing solutions provide dynamic analysis of suspicious files. When a file is executed in a controlled monitor, its behavior is to detect malicious unauthorized access to system resources or attempts to connect to command-and-control servers. This behavioral analysis with static analysis techniques delivers an inclusive tactic for malware uncovering.

An AI-powered system can automatically initiate containment measures to isolate affected systems from the network, initiate data backups, and notify relevant stakeholders when detecting a potential breach. Machine learning models can also assist in forensic analysis by correlating data from different sources and reconstructing attack timelines while identifying the basic origin of the happening. AI's ability to learn from past incidents enables continuous improvement of response strategies. In the usefulness of previous responses, AI systems can refine their actions and adapt to new threat scenarios [24]. This iteration learning procedure in incident response mechanisms endures health besides effectiveness against evolving threats.

2.3 Existing Research on Cybersecurity

The integration of fog computing into the cybersecurity landscape has garnered significant research attention due to its potential to enhance security measures through localized data processing and reduced latency. Various studies have explored the apps of machine-learning systems in detecting and extenuating cyber-fears within fog-computing environments [25]. Table 1 summarizes notable research efforts in this domain, highlighting key aspects like study focus, results, employed models, and identified limitations. The aim is to offer a comprehensive overview of the existing state-run research and identify gaps in the upcoming survey on ML-based cybersecurity for fog computing.

Table 1: Survey of ML-based Cybersecurity

ISSN: 3246-544X

Author Name	Year	Study	Results	Models	Limitations
Yu et al.	2020	Safety besides privacy problems in fog-computing	Identified key sanctuary and secrecy challenges in cutting-edge fog computing	Conceptual framework	Lack of empirical evaluation and specific ML models
Yang et al.	2021	Fog computing: Developments and security concerns	Proposed security solutions plus architecture for fog computing	Various security techniques	Limited focus on practical implementation and real-world testing
S. Mouradian et al.	2017	An inclusive survey on fog State-of-the-art research challenges	A comprehensive appraisal of safety tasks in fog computing	Survey also analysis	No implementation or evaluation of proposed solutions
Atlam et al.	2018	Fogs computing and its part in the Internet of Things	concept of fog computing and discussed its security	Conceptual discussion	Lacked focus on machine learning models for security

2.4 Literature Gap

Despite the significant advancements in applying machine learning techniques for cybersecurity in fog computing, several gaps remain. A majority of the existing studies conceptually lack empirical validation and practical implementation [5]. Many of the frameworks and models proposed have not been tested in real-world environments, which raises questions about their effectiveness and scalability. It is a limited focus on comprehensive security keys that speech a wide variety of threats beyond specific attacks like DDoS. Further use of various machine learning algorithms has remained less explored, with comparative analyses toward determining the greatest actual models for different types of cyber threats being scarce. There is also a need for research that integrates emerging AI methods, such as reinforcement learning and advanced neural networks, to enhance the adaptability and resilience of security systems in dynamic fog-computing environments [7]. Addressing these gaps is crucial for successfully developing the best and most practical cybersecurity solutions to protect fog computing infrastructures alongside evolving threats and developing new policies in the era of AI and ML.

3. Methodology

This section will explore our analysis and methods as techniques that were used to analyze cyber threat detection with the help of AI Models and ML techniques [4]. There are a few key steps to be used in fog computing for cyber-detecting systems. From the Figure 2 proposed framework, the first step in designing the research framework is data collection from the online data repository site Kaggle because it uses secondary data for analysis. The next is data preprocessing, which removes the data complexity and transforms the data features to remove missing values. The 3rd step features engineering for preparing variables to set targets, and normalization is helpful for machine models. Next, implement the ML models and predict the cyber-attacks with fog computing environments. to evaluate their results with different metrics and produce the normal and attack distribution plots soon.

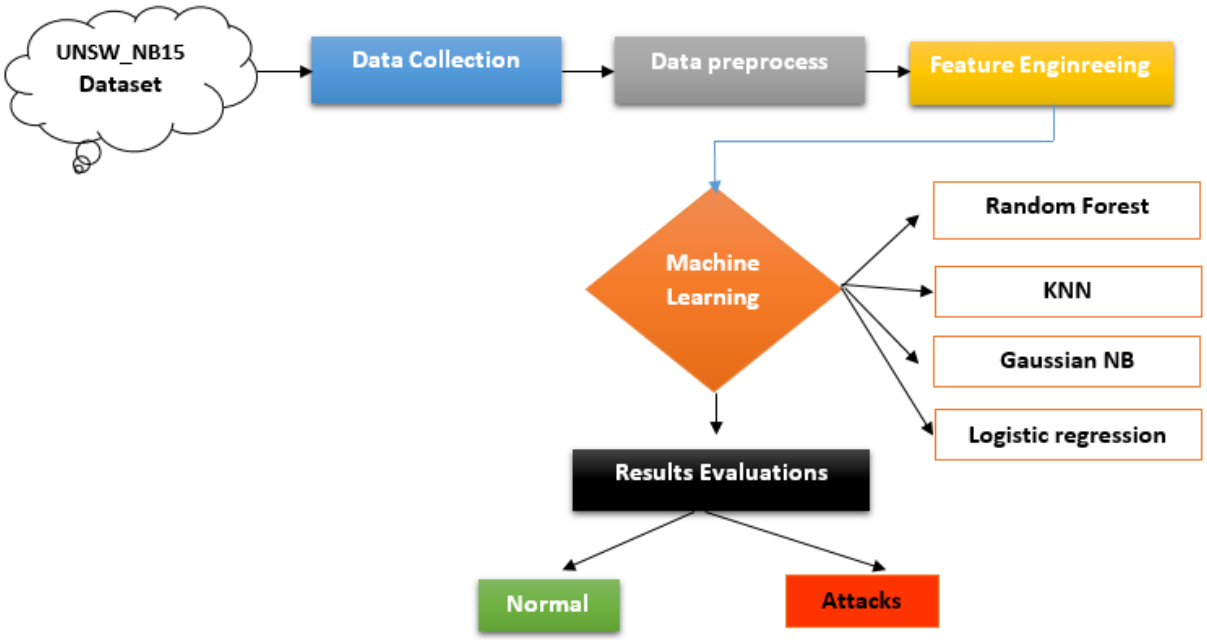

Figure 2: Proposed Framework

3.1 Data Collection

The dataset used for analysis remains the UNSW-NB15 dataset, which stands as a well-known standard dataset for network intrusion detection studies. This data is divided into two categories: training and testing sets. For the testing set in UNSW_NB15_testing-set.csv to be loaded and previewed, the dataset contains detailed records of network traffic, including various features that describe the characteristics of the network packets [25]. Samples of 1000 rows from the full dataset were laden using the panda's library in Python. The complete testing set consists of 175,341 rows and 45 columns representing a comprehensive array of

ISSN: 3246-544X

network traffic data. Each row in the dataset corresponds to a single network connection. Multiple attributes used are protocol-type (proto), service-type (service) plus state and packet counts (spkts, dpkts) also, byte counts (sbytes, dbytes), and several other statistical metrics related to the network traffic.

3.2 Data Preprocess

Data preparation is vital in getting the data set ready for examination. They deal with missing data and substantially influence the accuracy of predictive algorithms. This might entail deleting rows containing incomplete data or assigning absent values utilizing appropriate methods [26]. Category variables are further encoded by employing one-hot or label-encoded data and converting their value to numbers suited for data mining algorithms. The feature scale is also used to normalize the input information, where characteristics contribute proportionately towards the model.

3.3 Feature Engineering

Feature engineering for the UNSW-NB15 dataset involves several key processes to transform rare net traffic information bent on an arrangement appropriate for machine-learning models. This includes encoding categorical variables like protocol types of service and connection state into arithmetical symbols finished systems. Continuous features in duration and byte are counted to be scaled to uniformity in their range and impact [27]. Feature interaction drives their new metrics and statistical measures that capture the insight patterns in network traffic. Missing values are handled using imputation methods, and dimensionality reduction techniques are PCA to be applied to streamline the dataset. Their domain-specific features, including attack indicators, must be carefully encoded to develop their model's capacity to detect and classify network anomalies excellently.

3.4 Machine Learning Models

Several machine learning models are employed to predict cybersecurity threats with the help of fog computing environments.

- **Random Forest:** a collective learning method that builds numerous decision trees throughout learning and then combines their results to increase forecast accuracy. It builds randomly, experimenting with portions of the initial data and characteristics, resulting in greater variety and less excessive fitting. Every tree individually votes on the last forecast, and their combined vote determines the conclusion [28]. This approach helps capture multifaceted outlines and interactions in the facts in Random Forest, is robust, and is accurate in handling large datasets with many features.

- **K-Nearest Neighbors (KNN):** This instance-based learning technique identifies a data item using almost every category among its k-nearest neighbors. The sum of time among information points is commonly quantified using measurements like the distance of Euclidean geometry. During their categorization, select the k nearest

ISSN: 3246-544X

training instances to the query location, then give the class labeling, which is the most commonly shared neighbor. Although obvious and simple to construct in KNN, it can become computationally demanding for large data sets to react to both the selection of k and the length of the measure [29].

- **Naive Bayes:** serves as a Bayesian classifier that assumes feature isolation. In its "naive" premise that all features influence the probability of a specific result separately, it frequently outperforms expectations in reality. Naive Bayes's theory computes the aftereffects probability of every group based on the given input characteristics and selects the category with the greatest subsequent certainty [30]. It is especially useful for classifying text issues alongside other situations wherein independent requirements are roughly true. It can be highly computationally effective in struggle handling correlated characteristics.

- **Logistic Regression:** has become a model based on statistics for issues with a binary classification representing the chance that the input being considered belongs to a specific class. It uses a logistic function to assess the link between the input characteristics and the probabilities of a binary result. The framework generates cutting-edge odds, which are thresholded to get class labels. The logistic regression method is simple to understand and efficient in training, and it works nicely with linearly separate data. Their ability to perform might decrease the increasing complexity in non-linear relationships unless combined with forward-thinking techniques like polynomial features or interaction terms [31].

3.5 Evaluation Metrics

Everyone the model's results are evaluated using a variety of measures to provide a full picture of its efficacy. Important metrics include their percentage of properly categorized occurrences in every case. The ratio of real positive predictions to the total positive predictions reflects the model's ability to prevent false positives. The fraction of actual positive prediction across all genuine positives of the model's capacity to detect all relevant cases. The harmonious average of accuracy and recall provides a fair assessment of the algorithm's performance. The area under the receiver operating characteristics curves measures the model's ability to differentiate different classes [32]. The algorithms used give us signals and knowledge of their advantages and drawbacks to pick the most successful model for the assignment under work.

4. Results

The evaluation of our analysis involved a detailed investigation of the performance of numerous machine learning models. Figure 3 shows that while some models excelled in certain metrics, others offered a balanced performance across the board. These comprehensive evaluations facilitated an informed selection of the optimal model, for the

ISSN: 3246-544X

chosen algorithm performed well on the test set and maintained generalizability and reliability in real-world applications.

	id	dur	proto	service	state	spkts	dpkts	sbytes	dbytes	rate	sttl	dttl	sload	dload	sloss	dloss	sinpkt	dinpkt
0	1	0.121478	tcp	-	FIN	6	4	258	172	74.087490	252	254	14158.942380	8495.365234	0	0	24.295600	8.375000
1	2	0.649902	tcp	-	FIN	14	38	734	42014	78.473372	62	252	8395.112305	503571.312500	2	17	49.915000	15.432865
2	3	1.623129	tcp	-	FIN	8	16	364	13186	14.170161	62	252	1572.271851	60929.230470	1	6	231.875571	102.737203
3	4	1.681642	tcp	ftp	FIN	12	12	628	770	13.677108	62	252	2740.178955	3358.622070	1	3	152.876547	90.235726
4	5	0.449454	tcp	-	FIN	10	6	534	268	33.373826	254	252	8561.499023	3987.059814	2	1	47.750333	75.659602

Figure 3: Dataset Overview

This dataset is divided into 2 parts: training and testing, in which 1000 rows and 45 columns contain different types of protocols and attack services, cyber threats singles and ids, and ip addresses for detection. Figure 4 shows the description of the data set's stats.

	id	dur	spkts	dpkts	sbytes	dbytes	rate	sttl	dttl	sload	
count	82332.000000	82332.000000	82332.000000	82332.000000	8.233200e+04	8.233200e+04	8.233200e+04	82332.000000	82332.000000	8.233200e+04	8.233
mean	41166.500000	1.006756	18.666472	17.545936	7.993908e+03	1.323379e+04	8.241089e+04	180.967667	95.713003	6.454902e+07	6.305
std	23767.345519	4.710444	133.916353	115.574086	1.716423e+05	1.514715e+05	1.486204e+05	101.513358	116.667722	1.798618e+08	2.393
min	1.000000	0.000000	1.000000	0.000000	2.400000e+01	0.000000e+00	0.000000e+00	0.000000	0.000000	0.000000e+00	0.000
25%	20583.750000	0.000008	2.000000	0.000000	1.140000e+02	0.000000e+00	2.860611e+01	62.000000	0.000000	1.120247e+04	0.000
50%	41166.500000	0.014138	6.000000	2.000000	5.340000e+02	1.780000e+02	2.650177e+03	254.000000	29.000000	5.770032e+05	2.112
75%	61749.250000	0.719360	12.000000	10.000000	1.280000e+03	9.560000e+02	1.111111e+05	254.000000	252.000000	6.514286e+07	1.585
max	82332.000000	59.999989	10646.000000	11018.000000	1.435577e+07	1.465753e+07	1.000000e+06	255.000000	253.000000	5.268000e+09	2.082

Figure 4: Dataset stats descriptions

INTERNATIONAL JOURNAL OF SUSTAINABLE DEVELOPMENT IN COMPUTING SCIENCE
OPEN ACCESS, PEER REVIEWED, REFEREED JOURNAL
ISSN: 3246-544X

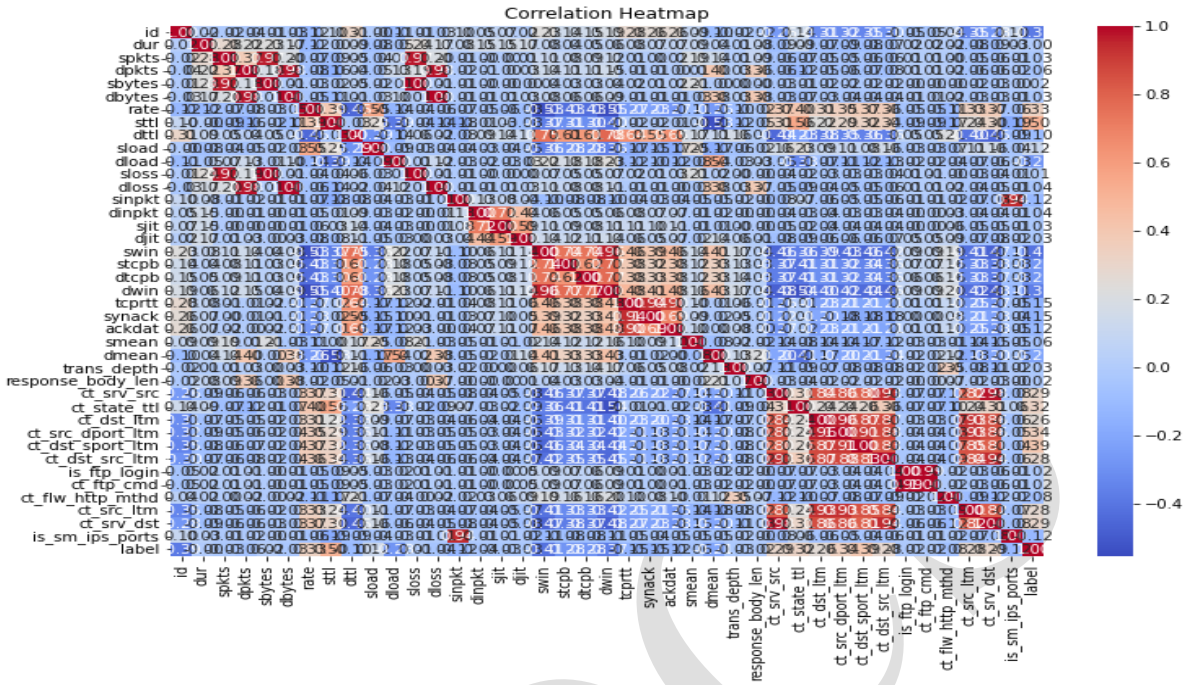

Figure 5: Correlation matrix of dataset

Figure 5 shows the correlation heat map and displays the pairwise correlation coefficients between different characteristics in the dataset, which have numbers that vary from minus one to one. A positive correlation with (closer near one) suggests that one attribute grows the additional one. Negative relationships (nearer to -1) indicate that if one attributes an alternative, it decreases over time. The color scale on the privileged half of this heat map shows these associations have blue colors, suggesting negative correlations and reddish shades show associations that are neutral colors around 0, showing a slight to the no-linear association. For strong positive correlations (red) between and strong negative correlations (blue) between "rate" and "id." This heat map helps identify which features are likely redundant or highly related and can inform feature selection and engineering decisions in the data analysis process.

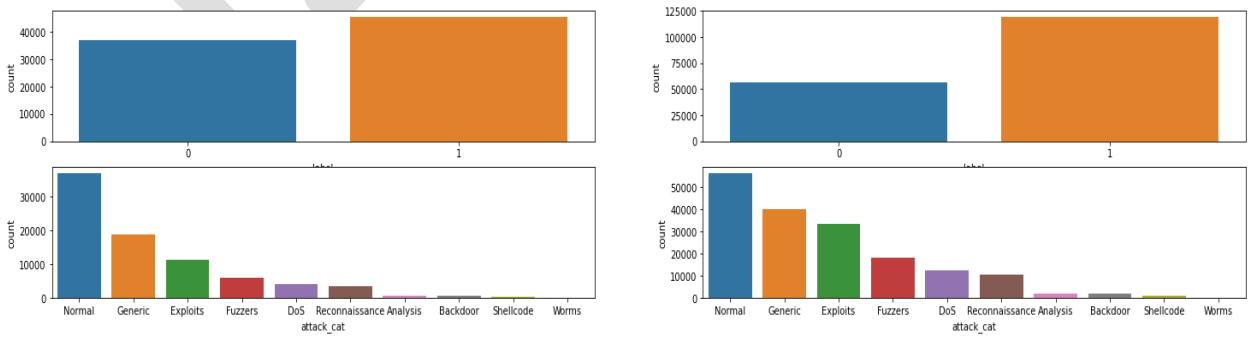

Figure 6: Cyber attack distributions

ISSN: 3246-544X

The top left graph shows the distribution of the binary labels, in which 0 signifies normal traffic, plus 1 denotes malicious traffic, and Figure 6 shows the Cyber attack distributions. It indicates that the dataset has a fairly stable delivery of normal besides malicious traffic samples. Their top right graph presents a similar distribution of the composed nature of the data set between normal (label 0) and malicious (label 1) traffic. The bottom left graph illustrates the count of different attacks in that 'Normal' traffic has the highest counts of 'Generic' attacks. Other levels of attack categories similar to 'Exploits', 'Fuzzers,' and 'DoS' have significantly fewer occurrences. The bottom right zoomed-in to detailed views of mirrors, and the bottom-left graphs distributions show that 'Normal' traffic and 'Generic' attacks dominate their categories like 'Back-door,' 'Shell-code,' and 'Worms' are relatively rare. Table 2 shows the data preparation for AI Modeling, and Table 3 shows the AI model comparison performance.

Table 2: Data Preparation for AI Modeling

Dataset	**Features Shape**	**Labels Shape**
Train	(175,341, 196)	(175,341)
Test	(82,332, 196)	(82,332)

Table 3: AI Models Comparison Performance

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Random Forest	0.94	0.97	0.94	0.95
KNN	0.91	0.96	0.91	0.93
Gaussian Naive Bayes	0.20	0.95	0.20	0.27
Logistic Regression	0.9473	0.9540	0.9903	0.9718

ISSN: 3246-544X

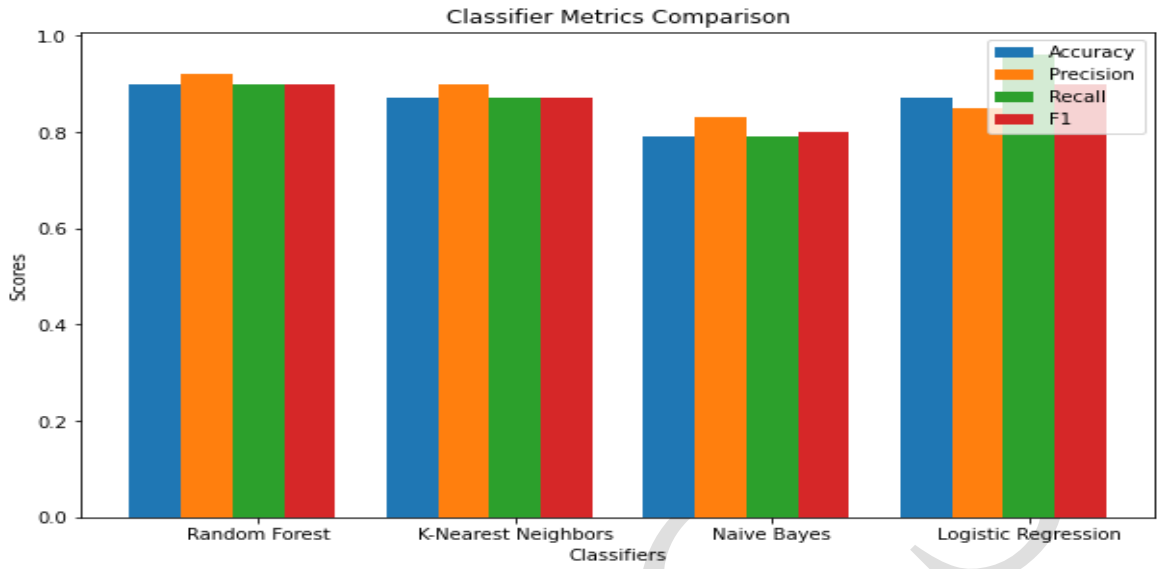

Figure 7: Models Classifier comparison graph

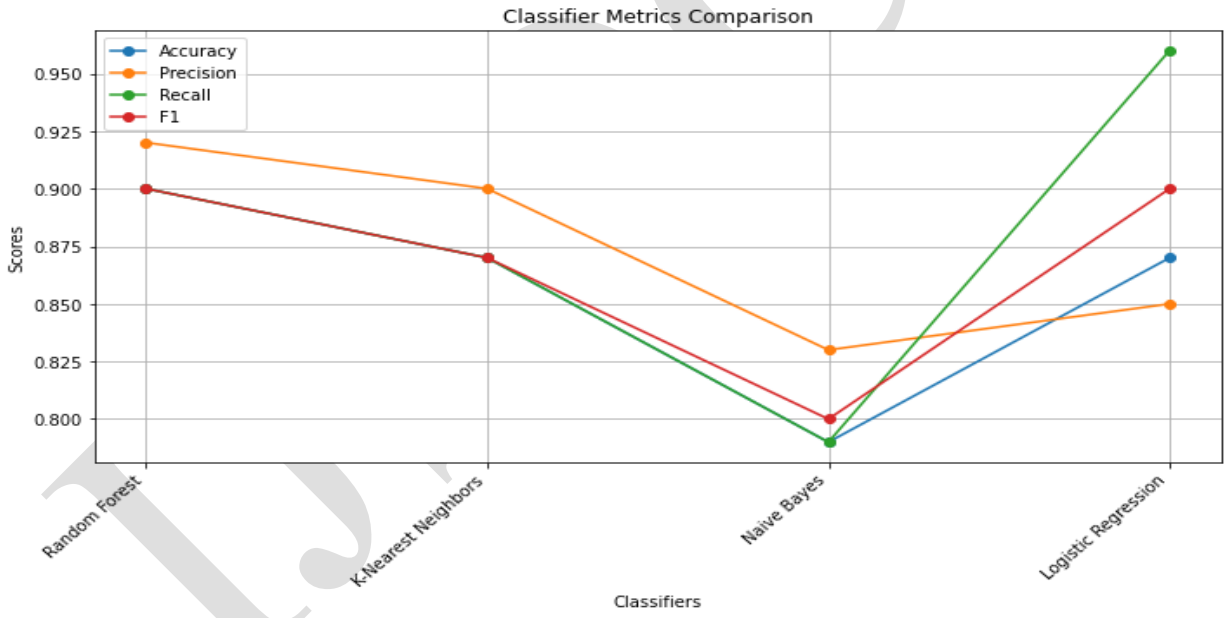

Figure 8: Line plot of Model results

Figure 7 shows the Models Classifier comparison graph, and Figure 8 shows the Line plot of Model results. The comparison of the performance metrics for the models reveals distinct differences in their effectiveness. The Logistic Regression model shows the highest accuracy (94.73%), and F1-Scores (0.9718) that indicate to performs in distinguishing between classes is too high precision (0.9540) and recall (0.9903) also is highly reliable for the task. The Random Forest model also performs strongly with an accuracy of 94% in a high F1-Score of 0.95 with the balance between precision (0.97) and recall (0.94). The K-Nearest

ISSN: 3246-544X

Neighbors (KNN) model- with an accuracy of 91% and F1-Score of 0.93, performs slightly shoddier, showing their strong precision (0.96) and reasonable recall (0.91). In Gaussian, the Naive Bayes model performs poorly with an accuracy of just 20% and an F1-Score of 0.27, despite having high precision (0.95) with recall (0.20). The most effective models are logistic Regression and Random Forest, but the Gaussian Naive Bayes are less reliable.

5. Proposed Research Comparison with Traditional Methods

In predictive analytics, machine learning also delivers effectiveness in various algorithms that can significantly impact the outcomes of analysis tasks [33]. Table 4 compares traditional methods and illustrates how traditional methods fare against contemporary approaches using machine learning models to predict income levels. Traditional methods often characterize their models as having less computational complexity in a scrap when handling complex datasets or imbalanced classes. In modern machine learning, logistic Regression, random forests, and KNN have demonstrated superior performance metrics [34]. This comparison will explore differences and insights into how advanced methods enhance predictive capabilities and robust analysis for improved decision-making and data-driven insights.

Table 4: Comparison of Traditional Methods

Aspect	**Traditional Methods**	**Proposed Research Methods**
Accuracy	Typically lower, depending on the algorithm (e.g., traditional statistical methods might have lower accuracy)	Higher accuracy was achieved with advanced machine learning techniques and fine-tuned models (e.g., Logistic Regression, Random Forest) [5].
Model Complexity	Simpler models with less computational complexity	More complex models can capture intricate patterns and interactions [9].
Computational Resources	Typically less demanding	Higher computational resources are required due to complex algorithms and large datasets [15]
Interpretability	Often higher, with more straightforward models	Potentially lower interpretability with complex models like Neural Networks but improved with

ISSN: 3246-544X

		models like Logistic Regression [20].
Adaptability to Data Changes	Less adaptable, often requiring manual updates and adjustments	More adaptable, with the ability to update models and incorporate new data efficiently [22].

6. Conclusion and Discussion

This research demonstrates the acute roles of cutting-edge AI-driven techniques in the cybersecurity challenges inherent in fog computing environments. The study highlights that some of the keys in Foggy computing introduce new cybersecurity vulnerabilities that are outstanding to their decentralized environment of edge devices and fog nodes. Out-of-date consolidated cybersecurity solutions are often scarce for these dynamic and mixt environments. Challenges are classified as malware injections and unauthorized access, with data breaches and service disruptions prevalent in specialized approaches to protect fog computing infrastructures. This study uses the data set to examine the practicality of machine learning models in predicting cybersecurity concerns in fog computing settings. The Kaggle gathering of data and subsequent preparation techniques are incomplete value management, plus categorical parameter encoding and scaling of features are acute in ready the dataset for assessment. The feature engineering technique improved the sets worth training machine learning models to convert network activity information to a more organized format. The logarithmic regression model with Random Forest outperformed the other models with excellent precision recollection and accuracy, and scores for F1 emphasize their robustness in distinguishing between normal and malicious traffic. Due to its simplistic assumptions, k-Nearest Neighbors also performed slightly less effectively than Gaussian Naive Bayes. This study emphasizes the importance of advanced machine learning techniques over traditional methods to showcase their ability to handle complex data and provide reliable predictions while significantly improving cybersecurity in fog computing environments.

References

[1] Bécue, A., Praça, I., & Gama, J. (2021). Artificial intelligence, cyber-threats and Industry 4.0: Challenges and opportunities. *Artificial Intelligence Review*, *54*(5), 3849-3886. Arif, H., Kumar, A., Fahad, M., & Hussain, H. K. (2024). Future Horizons: AI-Enhanced Threat Detection in Cloud Environments: Unveiling Opportunities for Research. *International Journal of Multidisciplinary Sciences and Arts*, *3*(1), 242-251.

[2] Kaloudi, N., & Li, J. (2020). The ai-based cyber threat landscape: A survey. *ACM Computing Surveys (CSUR)*, *53*(1), 1-34.

[3] Gudala, L., Shaik, M., Venkataramanan, S., & Sadhu, A. K. R. (2019). Leveraging Artificial Intelligence for Enhanced Threat Detection, Response, and Anomaly Identification in Resource-Constrained IoT Networks. *Distributed Learning and Broad Applications in Scientific Research*, *5*, 23-54.

[4] Sarker, I. H., Furhad, M. H., & Nowrozy, R. (2021). Ai-driven cybersecurity: an overview, security intelligence modeling and research directions. *SN Computer Science*, *2*(3), 173.

[5] Sohal, A.S., Sandhu, R., Sood, S.K. and Chang, V., 2018. A cybersecurity framework to identify malicious edge device in fog computing and cloud-of-things environments. *Computers & security*, *74*, pp.340-354.

[6] Sharma, A., & Singh, U. K. (2022). Modelling of smart risk assessment approach for cloud computing environment using AI & supervised machine learning algorithms. *Global Transitions Proceedings*, *3*(1), 243-250.

[7] Ahanger, T. A., Tariq, U., Ibrahim, A., Ullah, I., Bouteraa, Y., & Gebali, F. (2022). Securing iot-empowered fog computing systems: machine learning perspective. *Mathematics*, *10*(8), 1298.

[8] Chang, V., Golightly, L., Modesti, P., Xu, Q. A., Doan, L. M. T., Hall, K., ... & Kobusińska, A. (2022). A survey on intrusion detection systems for fog and cloud computing. *Future Internet*, *14*(3), 89.

[9] Lawal, M. A., Shaikh, R. A., & Hassan, S. R. (2020). An anomaly mitigation framework for iot using fog computing. *Electronics*, *9*(10), 1565.

[10] Qiu, M., Kung, S. Y., & Gai, K. (2020). Intelligent security and optimization in Edge/Fog Computing. *Future generation computer systems*, *107*, 1140-1142.

[11] Krishnaraj, N., Daniel, A., Saini, K., & Bellam, K. (2022). EDGE/FOG computing paradigm: Concept, platforms and toolchains. In *Advances in Computers* (Vol. 127, pp. 413-436). Elsevier.

[12] Sari, A. (2018). Context-aware intelligent systems for fog computing environments for cyber-threat intelligence. *Fog computing: Concepts, frameworks and technologies*, 205-225.

[13] Alwakeel, A. M. (2021). An overview of fog computing and edge computing security and privacy issues. *Sensors*, *21*(24), 8226.

[14] Abdullahi, M., Baashar, Y., Alhussian, H., Alwadain, A., Aziz, N., Capretz, L. F., & Abdulkadir, S. J. (2022). Detecting cybersecurity attacks in Internet of things using artificial intelligence methods: A systematic literature review. *Electronics*, *11*(2), 198.

[15] Sandhu, R., Sohal, A. S., & Sood, S. K. (2017). Identification of malicious edge devices in fog computing environments. *Information Security Journal: A Global Perspective*, *26*(5), 213-228.

[16] Zhou, L., Guo, H., & Deng, G. (2019). A fog computing based approach to DDoS mitigation in IIoT systems. *Computers & Security*, *85*, 51-62.

[17] Moura, J., & Hutchison, D. (2020). Fog computing systems: State of the art, research issues and future trends, with a focus on resilience. *Journal of Network and Computer Applications*, *169*, 102784.

[18] Yakubu, J., Abdulhamid, S. I. M., Christopher, H. A., Chiroma, H., & Abdullahi, M. (2019). Security challenges in fog-computing environment: a systematic appraisal of current developments. *Journal of Reliable Intelligent Environments*, *5*(4), 209-233.

[19] Abdulkareem, K. H., Mohammed, M. A., Gunasekaran, S. S., Al-Mhiqani, M. N., Mutlag, A. A., Mostafa, S. A., ... & Ibrahim, D. A. (2019). A review of fog computing and machine learning: concepts, applications, challenges, and open issues. *Ieee Access*, *7*, 153123-153140.

[20] de Souza, C. A., Westphall, C. B., Machado, R. B., Loffi, L., Westphall, C. M., & Geronimo, G. A. (2022). Intrusion detection and prevention in fog based IoT environments: A systematic literature review. *Computer Networks*, *214*, 109154.

[21] Zhou, I., Makhdoom, I., Shariati, N., Raza, M. A., Keshavarz, R., Lipman, J., ... & Jamalipour, A. (2021). Internet of things 2.0: Concepts, applications, and future directions. *IEEE Access*, *9*, 70961-71012.

[22] Zaman, S., Alhazmi, K., Aseeri, M. A., Ahmed, M. R., Khan, R. T., Kaiser, M. S., & Mahmud, M. (2021). Security threats and artificial intelligence based countermeasures for Internet of things networks: a comprehensive survey. *Ieee Access*, *9*, 94668-94690.

[23] Alwakeel, A. M. (2021). An overview of fog computing and edge computing security and privacy issues. *Sensors*, *21*(24), 8226.

[24] Wazid, M., Das, A. K., Shetty, S., Rodrigues, J. J., & Guizani, M. (2022). AISCM-FH: AI-enabled secure communication mechanism in fog computing-based healthcare. *IEEE Transactions on Information Forensics and Security*, *18*, 319-334.

[25] Kasongo, S. M., & Sun, Y. (2020). Performance analysis of intrusion detection systems using a feature selection method on the UNSW-NB15 dataset. *Journal of Big Data*, *7*(1), 105.

[26] Kumar, V., Das, A. K., & Sinha, D. (2020). Statistical analysis of the UNSW-NB15 dataset for intrusion detection. In *Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2019* (pp. 279-294). Springer Singapore.

[27] Choudhary, S., & Kesswani, N. (2020). Analysis of KDD-Cup'99, NSL-KDD and UNSW-NB15 datasets using deep learning in IoT. *Procedia Computer Science*, *167*, 1561-1573.

[28] Mosavi, A., Salimi, M., Faizollahzadeh Ardabili, S., Rabczuk, T., Shamshirband, S., & Varkonyi-Koczy, A. R. (2019). State of the art of machine learning models in energy systems, a systematic review. *Energies*, *12*(7), 1301.

[29] Sullivan, E. (2022). Understanding from machine learning models. *The British Journal for the Philosophy of Science*.

[30] Reddy, E. M. K., Gurrala, A., Hasitha, V. B., & Kumar, K. V. R. (2022). Introduction to Naive Bayes and a review on its subtypes with applications. *Bayesian reasoning and gaussian processes for machine learning applications*, 1-14.

[31] Boateng, E. Y., & Abaye, D. A. (2019). A review of the logistic regression model with emphasis on medical research. *Journal of data analysis and information processing*, *7*(04), 190.

[32] Handelman, G. S., Kok, H. K., Chandra, R. V., Razavi, A. H., Huang, S., Brooks, M., ... & Asadi, H. (2019). Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods. *American Journal of Roentgenology*, *212*(1), 38-43.

[33] Bowles, M. (2019). *Machine Learning with Spark and Python: Essential Techniques for Predictive Analytics*. John Wiley & Sons.

[34] Shah, K., Patel, H., Sanghvi, D., & Shah, M. (2020). A comparative analysis of logistic regression, random forest and KNN models for the text classification. *Augmented Human Research*, *5*(1), 12.

[35] W. Yu, T. Dillon, F. Mostafa, W. Rahayu, and Y. Liu, "A global manufacturing big data ecosystem for fault detection in predictive maintenance," IEEE Transactions on Industrial Informatics, vol. 16, no. 1, pp. 183–192, Jan. 2020. doi:10.1109/tii.2019.2915846

[36] Yang, Y., Chen, X., Tan, R., & Xiao, Y. (2021). *Intelligent IoT for the Digital World: Incorporating 5G Communications and Fog/Edge Computing Technologies*. John Wiley & Sons.

ISSN: 3246-544X

[37] Mouradian, C., Naboulsi, D., Yangui, S., Glitho, R. H., Morrow, M. J., & Polakos, P. A. (2017). A comprehensive survey on fog computing: State-of-the-art and research challenges. *IEEE communications surveys & tutorials*, *20*(1), 416-464.

[38] Atlam, H. F., Walters, R. J., & Wills, G. B. (2018). Fog computing and the internet of things: A review. *big data and cognitive computing*, *2*(2), 10.