International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 02 Feb. 2025, pp: 873-892 www.ijaem.net ISSN: 2395-5252

Ai-Powered Cyber Threat Intelligence System for Predicting and Preventing Cyber Attacks

Aashish Mishra

Date of Submission: 25-02-2025 Date of Acceptance: 05-03-2025

ABSTRACT
This paper aims at discussing and analyzing ways in which artificial intelligence revolutionizes the approach to cybersecurity by focusing on data. This work indicates the incorporating of AI in cybersecurity strategies not only improves security but also minimizes expenditures and errors, all needed in modern-world cybersecurity. The expansion of various fields and industries, along with the integration of numerous smart devices that are connected to the internet, has resulted in a highly secured threat level. Cybersecurity is mainly about identifying threats and responding to them, but that is not possible today with traditional methods. Modern threats and their constant evolution are partially beyond the capacity of traditional security instruments to protect an organization or company
Combining anomaly detection and machine learning (ML) techniques enables the system to adapt to changing security threats. The first phases involve gathering and analyzing data from numerous cloud sources to improve the system's capacity to spot problems. Supervised learning with Random Forest classifies known hazards, while unsupervised learning with Isolation Forest detects new abnormalities. Real-time monitoring and response considerably improve the system's threat detection rates (95%), anomaly detection (93%), and other performance indicators. The proposed system surpasses the existing system by 95% accuracy, 93% precision, and 96% recall. These findings demonstrate how effectively the framework enables cloud safety and its capacity to enhance overall digital safety and proactively prevent assaults.

I. INTRODUCTION

Cybersecurity draws its roots from the past few decades due to an increase in complexity and frequency of cyber threats in modern society. The patterns of protecting systems and networks are mainly constrictive of measures on how best to counter threats once they are detected. The methods using firewalls, anti-virus programs, and intrusion detection systems worked on the principle of pattern matching with the existing known attack patterns. Though these approaches offered a somewhat elementary level of security, they are ineffective in identifying novel or emergent risks. These approaches gradually lost efficiency as cybercriminals increased their levels of work, meaning that companies became more vulnerable (Brown & Patel, 2022). One phenomenon that revolutionized the concept of cybersecurity is the introduction of artificial intelligence. Machine learning and deep learning are kinds of artificial intelligence approaches that have the potential to analyze a huge amount of data, recognize interesting patterns, and detect possible threats more efficiently than manually. Through machine learning algorithms, one can train a system to analyze large quantities of data in the same way the human brain does and be able to perceive potential IT security threats such as a cyberattack. This makes it easier for the cybersecurity systems to identify new threats that are not there and handle new methods easier than rule-based systems, as suggested by Johnson & Smith (2023). This is perhaps one of the biggest strengths of AI in the field of cybersecurity because it means one can move from a reactive approach to the problem. In conventional security models, dealing with threats requires a reaction that might take time, hence no ability to prevent the impact yet. The autonomous systems have the abilities of foreseeing such attacks and even countering them, as the system is constantly analyzing the traffic produced and the behavioral patterns. This means that threats are detected way before they manifest themselves, and in the process of containing the threats, the impact is greatly reduced on the organization. Artificial intelligence can filter the compromised systems or prevent such activities and give a prompt to the

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 02 Feb. 2025, pp: 873-892 www.ijaem.net ISSN: 2395-5252

threats (Miller & Wright, 2021). Artificial intelligence seeks to analyze and make decisions based on large volumes of data in the field of cybersecurity, and the quality of the data is central to it. Big data is useful in training the AI models. It contains the information that is required in pattern and anomaly detection. AI systems are capable of amassing large amounts of data from various sources; hence, they have an added advantage in scoring a more general understanding of threats. The capability to indeed analyze large data sets and process this data in real-time makes AI-empowered cybersecurity systems adaptive to new threats and enhances their detection methods iteratively. The information of citizens has to be collected and used, which is a violation of their privacy rights (Zhang & Liu, 2022). Artificial intelligence can bring many advantages to cybersecurity; it can also produce new issues. One of them is adversarial attacks, the case when malicious individuals or groups change the data that the AI models learn from with the specific intent to deceive the AI. Artificial intelligence has some issues that provide discussion in the field of ethical standards. The use of artificial intelligence in decision-making for privacy or surveillance. This involves enhancing the stability of the AI algorithms to enhance their function, erecting proper measures for the ethical practice of artificial intelligence, and proper guarding measures against such adversarial attacks. That is why prospects of advancing artificial intelligence to the field of cybersecurity may have a significant influence on the protection of digital assets in the future (Williams & Dawson, 2023). The advancement of global industries to the digital platform has highly contributed to the generation and storage of large amounts of data, which has raised some concerns about the security of data among organizations. As the digital structures unfold, so does the sort and severity of threats, inclusive of cyber theft and cyber-blackmail, advanced cyber-spying by subordinates, and state-sponsored hackers. Conventional concepts of security have endeavored to eliminate the use of signatures, where threats are detected with similarity we have recorded in the past. These traditional methods are ineffective in identifying emerging or dynamic threats; hence, systems remain exposed to exploitation (Brown & Patel, 2022). The concept of integrating artificial intelligence into cybersecurity measures has become a viable solution. With the help of machine learning and deep learning, artificial intelligence has the function of real-time data analysis and detection of weak signs of emerging threats. Traditional cybersecurity solutions employ a set of fixed rules and processes that get applied in the course of a process, while AI-based systems use data and develop capabilities based on that learning. This change from reactive to proactive is a revolution in this approach to threat detection and helps organizations to be more secure in their virtual assets given the increasingly hostile nature of cyberspace (Johnson & Smith, 2023). The uptake of AI in cybersecurity solves the scalability question as well. With the constant increase of data being generated in digital form, there is increasing pressure on detecting and securing these assets. Artificial intelligence inspires solutions that are capable of analyzing big data sets in a shorter timeframe as compared to security analysts and conventional IT security systems. This scalability is critical in today's organizations, as they have to safeguard extensive and complicated networks against various attacks. AI is capable of performing a set of routine cybersecurity operations, including the detection and response to threats on network traffic, so that cybersecurity specialists can free up their time by eliminating repetitive routine. There is still some discomfort in thinking about AI in cybersecurity. One major issue that can be seen as a threat is the reliability of the data used to train AI and ML algorithms. It expounds the increasing chances of adversarial attacks, which is a process by which AI is fed with wrong data with the aim of having the system make a particular decision. The resolution of these threats is only possible through continuous research and development of the AI algorithms, which makes them incorporate and then sort out the most efficient way of defending an organization's cyberspace than just an AI (Williams & Dawson, 2023). The use of artificial intelligence in the field of cybersecurity is an important shift in organizations' protection against cyber criminals. The use of AI in the aspects of big data processing and analyzing the likely risks that may occur, organizational security can improve on their defenses, reaction rates, and even tame the effects of cyber threats. The authors noted that as information threats and risks develop further, AI is projected to play a critical part in cybersecurity, which, as the field is characterized by rapid advancements, needs further development and innovation. Refocused to the analysis of Zinul et al. (2024), the authors provide information about the use of V2O5 as HTL as well as its effect on the efficiency of the Sb2Se3-based solar cells. This paper demonstrates that V2O5 is capable of integrating with the energy level of Sb2Se3, hence promoting whole transport while also minimizing

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 02 Feb. 2025, pp: 873-892 www.ijaem.net ISSN: 2395-5252

recombination losses at the back interface. It leads to improved VOC, JSC, and PCE values, showing that V2O5 could be used as a promising component for highperformance photovoltaic interface designs. According to Hajjiah&Gorji (2024), to maximize the efficiency of Sb2Se3-based solar cells, it is imperative to control the device structure at every layer. The study also employs SCAPS-1D simulations to optimize each layer of the photocathode: Al/FTO/SnS2/Sb2Se3/V2O5/Ni, each of which serves a unique purpose in charge carrier dynamics. It is equally clear from the results of this work that considerable effort has to be paid to the design and optimization of such layers. The SCAPS-1D simulation tool, which Park et al. (2024), among other scholars, have used as the main tool for analyzing the performance of photovoltaic devices, is quite vital. In this work, the prerequisite knowledge is given through SCAPS-1D, by which the electrical performance of the Sb2Se3-based solar cells is simulated, where several effective parameters, including layer thickness, defect density, and carrier concentration, are considered. Justifying the use of the tool, the author notes that it is most useful in the tuning of a device when real-world conditions are closely approximated. In their recent work, Singh et al. (2023) discuss the effect of absorber layer thickness on Sb2Se3based solar cell efficiency. Here, they demonstrate that an absorber thickness of about 800 nm is optimal in terms of both light absorption and the avoidance of recombination losses. Thinner layers may receive inadequate light, while thicker layers may result in increased recombination losses, thereby making it an optimum parameter for device performance enhancement. In this regard, Duan et al. (2022) make use of Sb2Se3 as an absorber material since it has a bandgap of about 1. 2 eV and high absorption coefficient and thus could find its application in the sol (Hassan Nawaz, 2024) are cell devices. High nontoxicity and availability of constituent elements make Sb2Se3 suitable to replace toxic and scarce materials used in photovoltaic devices in favor of global sustainable development. This is due to the fact that the material's properties allow energy-conversion efficiency at levels that place it at the heart of the photovoltaic technologies of the future. The perception of risk associated with drugs during pregnancy indicated the sources of information sought most commonly were the doctors, printed information leaflets, and pharmacists. To the investigators' knowledge, there is limited empirical work that examines the role of pharmacists for providing teratology information to pregnant women and healthcare practitioners (Nayem Uddin Prince, 2024). The protection of information must be realized that it has to be applied in every aspect of any project or program in the collection, analysis, and use of data, starting or during the conceptualization of any program. Many studies already underscoring this criticality were already mentioned (Nayem Uddin Prince, 2024). It was established that the proper usage of antipsychotics indicated by their rational prescription is necessary to manage schizophrenia in the long run. Data shows that the relapse rate among first-episode patients is as high as 80 percent within five years after developing resistance to treatment, so many others have to go back to receiving treatment in the following years (Nayem Uddin, 2024). AI-Powered Data-Driven Cybersecurity… Nayem Uddin Prince et al. 336 Nanotechnology Perceptions Vol. 20 No. S10 (2024) Schizophrenia is among the top ten illnesses causing the disease burden worldwide, according to the WHO, with a prevalence of twenty-six million, and of this, sixty percent of the patients suffer moderate to severe disabilities. (Uddin Prince, 2024). Pharmacists have a vital role in dealing with the issue of drugs for pregnant women (Nayem Uddin Prince, 2024). In this digital world, they use a number of techniques to lure their prey, and the most common but everevolving and dangerous are the phishing attacks. There are different views on what phish is because its nature and manifestation constantly change due to context, and experts have given numerous definitions based on current and past research of Nayem Uddin Prince (2024). Cybercrime is a threat to the world economy, every country's security, social order, and interests (Nayem Uddin Prince, 2024). According to the 2020 Official Annual Cybercrime Report, the global cybercrime rate has been identified as one of the most engaging activities that humanity will face in the next two decades by Nayem Uddin Prince (2024). The inconsistencies in prescriptive practices and in employing non-potentially useful drugs make a positive change concerning misuse, overuse, and underuse of drugs that are helpful in reducing the disease consequences and the costs involved in disease impacts, higher in the patients. Below is the summary of the portfolio, including the work of the candidate (Nayem Uddin Prince, 2024). The banking sector is at one of the critical moments in its development as the experience in the field of digitalization is gradually deepening. quickly today. There is an urgent need for banks to transform to new generation methods of operation. Offer smooth, effective, and secure financial

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 02 Feb. 2025, pp: 873-892 www.ijaem.net ISSN: 2395-5252

services since they are facing two difficulties at once: they are able to meet the growing needs of their customers, as well as ensure well implemented securities for data (Hassan Nawaz, 2024). The cloud solutions that it offers are quite unique and can be considered revolutionary in the near future. Revolutionarily changing the entire banking sector, Huawei Pakistan has stepped to a front-runner in this rapidly changing market. Today's banking industry is characterized by a dependence on it, and at the same time, traditional physical banks are increasingly becoming outcompeted. Embracing the digital sphere. (Hassan Nawaz, 2024). Personalized and live, customized financial services for the clients as well as the sustenance of pillars data security. It is higher than ever in Pakistan, as the requirement for new ideas is normally higher than the need for routines. The banking industry is expanding and is experiencing greater competition year after year (Hassan Nawaz, 2024). Huawei Pakistan, which is a Huawei-affiliated multinational technology company. Has greatly contributed to this process by offering the most advanced cloud solutions meeting the client's needs to fulfill the specific needs that are characteristic for the banking industry. Huawei Pakistan is at the forefront of assisting the banks to deal with the challenges and respond to the opportunities that the technological advancement brings. Challenges brought about by digitization areas due to its existence, Cloud Base will tackle till computing, artificial intelligence, and telecom sector with (%) as 42 for computing, 31 for artificial intelligence, and 27 for computing (Hassan Nawaz, 2024).

Overview of Legacy IT Systems:

Disadvantages and Negative Aspects Related to an Outdated IT Environment Old hardware and pre-installed software is still present in proprietary IT systems, which are characterized by a high degree of tool commoditization. applied in organizations, and they bring with them a number of issues and concerns. Indeed, such systems, most of the time, are said not to possess the capabilities required to counter present-day security threats, hence making their networks more vulnerable to cyber threats (Hassan Nawaz, 2024). This leads to service incompatibilities, in which companies thus find themselves adopting and owning many systems, which only serves to increase the cost of. They also increase the skill required for proper maintenance and, at the same time, reduce efficiency (Hassan Nawaz, 2024). Further, the conventional systems are likely to have relatively higher failure rates, and they result in technological flaws that interfere with organizational operations and revenue (Hassan Nawaz, 2024).

Problem Statement:

The introduction of AI in cybersecurity presents several solutions to these challenges, given that the threats are detected and addressed more actively and efficiently. There are certain challenges and problems associated with the reinforcement of cybersecurity through the use of artificial intelligence. Traditional approaches to security, which mostly incorporate rulebased and signature-based approaches, are ill-equipped to handle the continuously evolving threat landscape. This inadequacy has contributed to an increase in successful breaches with severe losses in billions of dollars, harm to reputation, and leakage of sensitive information. The increased amount of data available in today's digital world is another big problem for cybersecurity personnel, who have to be constantly vigilant and protect massive networks. These are involved in aspects such as accuracy of data used to train the AI models, security of the AI systems from adversarial tampering, and dealing with the ethical issues of privacy and AI decision-making authority. The purpose of this research is to determine whether the use of AI in protecting systems enhances threat detection and prevention. It looks at the limitations of incorporating artificial intelligence into practices in cybersecurity and possible ways of overcoming the problems. The Dual-Edged Sword of AI and ML in Cybersecurity Leveraging AI for Enhanced Threat Detection and Response. Artificial intelligence and machine learning have greatly impacted the cybersecurity field, adding more threat detection and mitigation. It allows organizations to process large amounts of data in real time and describe risks more effectively than conventional methods. Still, the same features ensure added security, while cyber attackers leverage artificial intelligence and machine learning in formulating better and more advanced tactics (Bhatia & Sharma, 2021; Zuech et al., 2019). For example, adversarial machine learning can manipulate AI systems to produce wrong threat evaluations and, in effect, open security structures to exploit (Chio& Freeman, 2018). Even though artificial intelligence and machine learning bring about compelling opportunities for monitoring cyber threats', they equally present novel risks that cybersecurity professionals have to learn to solve constantly (Sarker et al., 2021).

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 02 Feb. 2025, pp: 873-892 www.ijaem.net ISSN: 2395-5252

Research Objectives:

- Evaluate the current capabilities of AI technologies in cybersecurity, focusing on their ability to detect and respond to threats in real time.
- Identify and analyze various data sources (e.g., network traffic, user behavior, threat intelligence feeds) that can be leveraged to enhance AI-driven cybersecurity strategies. • Investigate different AI algorithms (e.g., machine learning, deep learning) and their effectiveness in identifying vulnerabilities, detecting anomalies, and predicting potential threats.
- Examine how AI can improve automated response mechanisms, reducing the time between threat detection and response to mitigate potential damage.
- Study methods for integrating AI-driven strategies into existing cybersecurity frameworks and systems to enhance overall security posture.
- Identify challenges and limitations associated with implementing AI in cybersecurity, including data privacy concerns, false positives, and the need for human oversight.
- Establish metrics to measure the effectiveness and efficiency of AI-driven cybersecurity strategies in enhancing threat detection and response.
- Analyze case studies that demonstrate successful implementations of AI in cybersecurity and identify best practices for organizations looking to adopt these strategies.
- Explore future trends in AI and cybersecurity, including emerging technologies, evolving threat landscapes, and the potential impact of regulatory changes.
- Provide actionable recommendations for organizations on how to effectively leverage AI for data-driven cybersecurity strategies, focusing on practical implementation and continuous improvement.

II. LITERATURE REVIEW

The cybersecurity domain has changed quite dramatically as to the volume and specifics of threats. In the Internet Security Threat Report by Symantec published in 2019, organizations are reported to undergo increased ransomware and, at the same time, phishing attacks, thus calling for more developed and flexible cybersecurity measures. The threats change constantly, and the methods used for countering threats change as well. There is an increasing use of artificial intelligence for threat detection and response. Machine learning and deep learning are becoming popular in the cybersecurity domain and are labeled as game-changers. These technologies allow systems to gather large amounts of data in real-time, which makes it easier for the systems to detect signs that indicate that the security has been compromised. It is important to indicate that Kahn and Steinberg (2021) note that through machine learning algorithms, it is possible to train several ML algorithms using the experience that has been accrued in order to identify patterns of behavior that are likely indicators of would-be security breaches. Aside from improving threat identification, this capability also shortens the time organizations take to respond to threats, thus minimizing their impact on the organization. Artificial intelligence technologies can actually help to automate a number of tasks. For example, when analyzing a flow of traffic in the computer network, with the help of AI systems, security issues can be addressed with no intervention of people on a regular basis, and efficiently occurring dangers can be detected and counteracted. of processes, which is important given today's amount of data generated daily. According to Dey et al. (2020), artificial intelligence can help secure operations more efficiently and allow the security personnel to pay attention to other, more critical issues instead of just observing the trends. For example, when analyzing a flow of traffic in the computer network, with the help of AI systems, security issues can be addressed without the intervention of people on a regular basis, and emerging dangers can be detected and counteracted. This is not only improving operational productivity but also helping cybersecurity personnel to focus on more important and unique tasks rather than spending a lot of their time on repetitive restoration work. There are various issues and drawbacks to implementing and utilizing AI in cybersecurity. Concerning false positives, which can be defined as the fact that benign activities might be classified as threats, their restoration is explained by Shafique et al. (2021). This challenge can put pressure on a security team with an enormous number of messages and thus overlook actual threats since almost everything looks like a threat. Moreover, the implementation of AI solutions into the curricula is questionable because an organization must pay attention to legal frameworks while applying AI-dand applications. These challenges accentuate the fact that artificial intelligence requires the integration of the use of

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 02 Feb. 2025, pp: 873-892 www.ijaem.net ISSN: 2395-5252

artificial intelligence in company operations while at the same time harnessing the capabilities of human personnel to monitor and, in some cases, correct some of the decisions made by the artificial intelligence systems. Artificial intelligence in cybersecurity as probable and eventual innovations in technology and methodology are possible. Xu et al. (2022) have considered future trends in artificial intelligence that are expected to define the future of cyberspace, including advancements in algorithms and the integration of artificial intelligence with others, namely blockchain technologies. Organizations in such areas will not only complement the use of AI in threat detection and response but will also spur more ideas for coming up with proactive actions in security. Given the growing awareness of the need for using AI in organizations' cybersecurity systems, further studies are crucial to tackle the existing problems and to achieve the full effectiveness of AIbased systems.

III. PROPOSED NETWORK THREATS DETECTION SYSTEM- AI@NTDS

An intelligent threats detection system, called the AI@NTDS system, is designed to investigate network threats and analyze them using specific features and algorithms, primarily for SSH sessions. This section describes the automatic data acquisition process and defines the features. A multilabel classification model will also be introduced.

A. System Architecture

The proposed AI@NTDS system architecture that is shown in Figure 1 has five parts, which perform data collection, data preprocessing, feature-based analysis, model training, and model output.

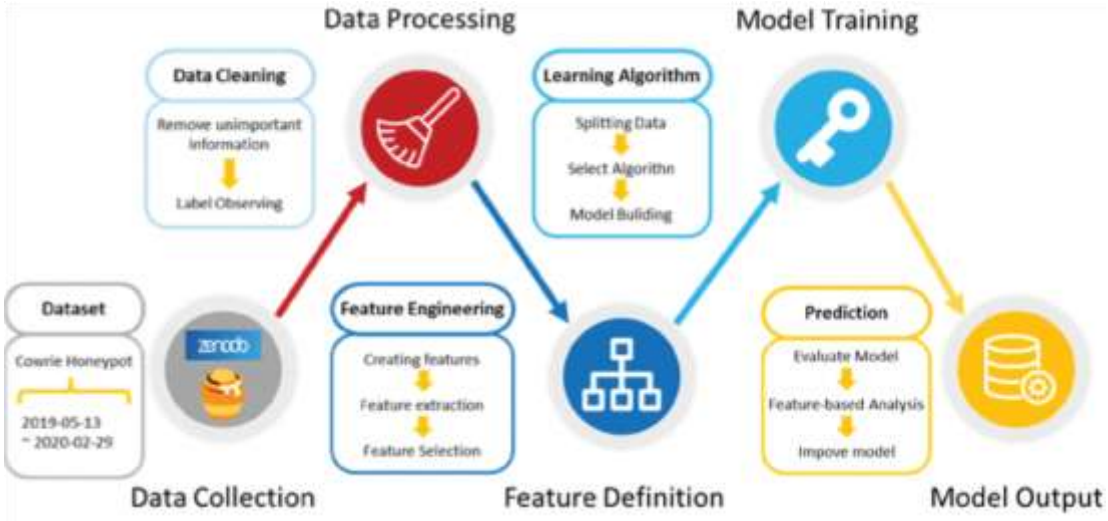

FIGURE 1. Proposed AI@NTDS system architecture.

The dataset of Cowrie Honeypots was obtained from the CyberLab Honeypot-Zenodo. The various attack features are identified from variations among attacks. Fifty-two features were extracted from the Cowrie Honeypot dataset and then grouped into message-based, host-based, and geographicbased features. The data preprocessing part ensures that the labels and contents in the samples are correct. The algorithm is used to evaluate the importance of various feature combinations. To ensure the stability of an AI model and prevent overfitting, the model training process should not learn too closely with the result of the training dataset. The model is validated during the training process. The performance of the presented model is determined at various times, and the results thus obtained are presented in the following section.

B. Data Collection

The Cyberlab Honeypot collected attackers' data from June 2019 to February 2020 for use in this study. Cowrie Honeypots, with approximately 50 nodes mostly at universities and companies in the European Union and the United States were used. Each file in the dataset is based on reports of daily intrusions. Sessions are grouped according to the attacker invades and leaves. Each group of sessions contains various events and explicit intentions. This goal of this work is to

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 02 Feb. 2025, pp: 873-892 www.ijaem.net ISSN: 2395-5252

reduce the complexity and automatically to collect results daily. This system automates the process by applying the concept of a crawler. The data collection program automatically decompresses and converts the extracted JSON file into a CSV file. Figure 2 presents the flow chart.

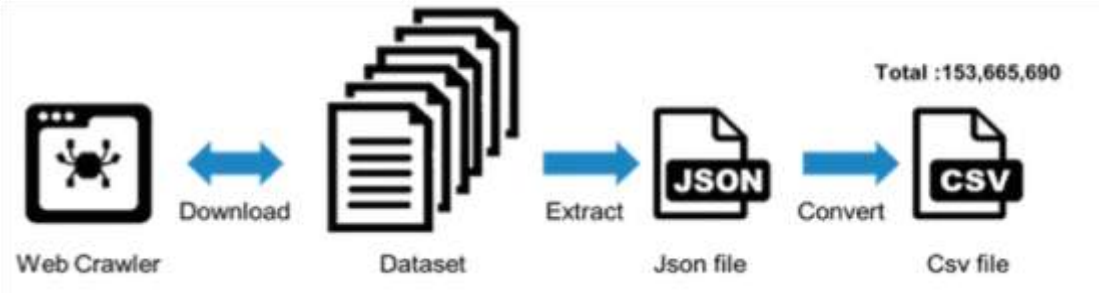

FIGURE 2. Flow chart of data collection.

C. Data Processing

Data processing firstly removes irrelevant information from the dataset to ensure data quality. The first step in this process is the removal of data associated with failed intrusions. Empty fields are deleted to save storage space and increase computing efficiency. Then, the cleaned data are labeled in a manner consistent with Enterprise Tactics in MITRE ATT&CK Enterprise Tactics comprise 14 groups of Tactics, of which those used herein are indicated below.

Table 1 presents the results obtained using the indicated Tactics. Nine tactics in the dataset are labeled with "no intention". They include No Action, Execution, Persistence, Privilege Escalation, Defense, Credential Access, Discovery, Command and Control, Impact. These tactics are associated with three malicious levels based on severity. Level 1 refers to actions that may damage the system, such as the execution of malicious files that stop the system. It is the most dangerous and malicious intention for a system, such as when a hacker inputs the command "kill", or "rm", or executes some unknown executable binaries. Level 2 refers to setting file permissions for personal accounts. For example, a hacker may input the "chmod" command or "chattr" command to change the file permission. Level 3 refers to the absence of action or scouting actions. For example, if a hacker inputs a command like "cat/etc/passwd" and "lscpu" to obtain system information, the command will be assigned to level 3.

ID	Name	Description
TA0043	Reconnaissance	The adversary is trying to gather information they can use to plan future operations.
TA0042	Resource Development	The adversary is trying to establish resources they can use to support operations.
TA0001	Initial Access	The adversary is trying to get into your network.
TA0002	Execution	The adversary is trying to run malicious code.
TA0003	Persistence	The adversary is trying to maintain their foothold.
TA0004	Privilege Escalation	The adversary is trying to gain higher-level permissions.
TA0005	Defense Evasion	The adversary is trying to avoid being detected.
TA0006	Credential Access	The adversary is trying to steal account names and passwords.
TA0007	Discovery	The adversary is trying to figure out your environment.
TA0008	Lateral Movement	The adversary is trying to move through your environment.
TA0009	Collection	The adversary is trying to gather data of interest to their goal.
TA0011	Command and Control	The adversary is trying to communicate with compromised systems to control them.
TA0010	Exfiltration	The adversary is trying to steal data.
TA0040	Impact	The adversary is trying to manipulate, interrupt, or destroy your systems and data.

TABLE 1 Enterprise Tactics in MITRE ATT&CK

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 02 Feb. 2025, pp: 873-892 www.ijaem.net ISSN: 2395-5252

Figure 3 presents the tactics distribution of labeled data. Defense Evasion is the most common tactics at Level 1. The attacker's purpose is not to leave records of the removal of downloaded programs, to destroy files, and to obfuscate the system. Malicious programs are commonly used to run scripts to set permissions and perform other actions that are typical of attackers in Honeypots. The most common tactic in Level 2 is Persistence. The attacker's purpose is to escalate privilege to an account. When a connection break occurs, an attacker maintains access to the system to support the malicious operations, or change the system configuration. The figure below shows that the most common attack following login at Level 3 is Discovery because when an attacker accesses the system, the first task is always to perform reconnaissance. The distribution of the tactics of an attacker when he enters a Honeypot can be identified from statistical data.

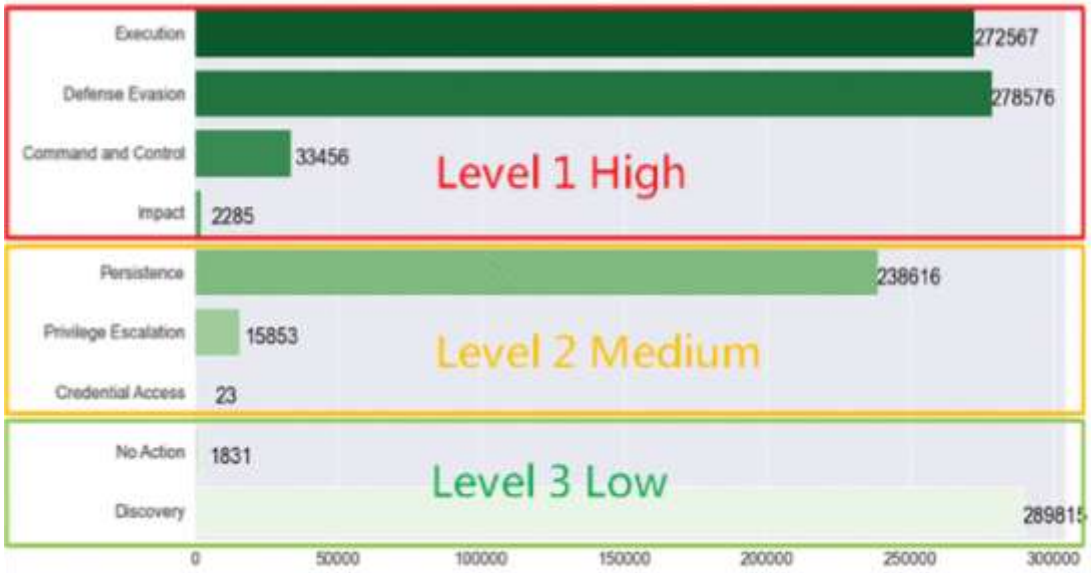

FIGURE 3. Data distribution of aggressive behavior.

The data in our work were collected from the 4th of June, 2019, to the 29th of February, 2020. The data were automatically collected using a web crawler, which grabbed 153,665,690 samples. After the invalid samples were removed, 298,667 valid sample entries remained to undergo the following process. The data from 4th June 2019 to 31th December 2019 formed the training dataset, while those from the 1st January 2020 to the 28th February 2020 formed the test dataset. Of the training data, 15% were allocated to the validation dataset that was used to evaluate the model. Table 2 presents the applications of the split dataset.

Dataset Purpose	Quantity	Collection Time	Proportion
Training	184463 samples	2019-06-04 ~ 2019-12-31	66%
Validation	46116 samples		11%
Test	68088 samples	2020-01-01~ 2020-02-29	23%

TABLE 2 Training, Validation and Test Datasets

D. Feature Definition

This subsection introduces and describes in detail 52 groups of features.

These groups were divided into message-based, host-based, and geography-based types groups, as shown in Table 3. The algorithms that are used for machine learning focuses on the weights of feature data, so feature extraction is crucial. In addition, the features proposed by the authors are the use of red font marks.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 02 Feb. 2025, pp: 873-892 www.ijaem.net ISSN: 2395-5252

Features Set : 52 Features					
F1	Keyword_bash	F18	Keyword_cat /etc (Get_sys_info)	F35	Count_Hex
F2	Keyword_shell	F19	Keyword_uname (Get_sys_info)	F36	Count_set
F3	Keyword_ssh	F20	Keyword_wc (Get_sys_info)	F37	Message_length
				F38	Messages / sec
F4	Keyword_help	F21	Keyword_crontab (Get_sys_info)	F39	Protocol
F5	Keyword_passwd [user] (Set_account)	F22	Keyword_w (Get_sys_info)	F40	Src_Port
F6	Keyword_chpasswd (Set_account)	F23	Keyword_ps (Get_sys_info)	F41	SSH_Client_Version
F7	Keyword_useradd -[suffix] (Set_account)	F24	Keyword_free (Get_sys_info)	F42	Username
F8	Keyword_. [file] (Execution_file)	F25	Keyword_lscpu (Get_sys_info)	F43	Password
F9	Keyword_sh [file] (Execution_file)	F26	Keyword_nproc (Get_sys_info)	F44	Duration
F10	Keyword_./ [file] (Execution_file)	F27	Keyword_uptime (Get_sys_info)	F45	Received_Size (AVG.)
F11	Keyword_perl [file] (Execution_file)	F28	Keyword_wget (Network_connect)	F46	File
F12	Keyword_python [file] (Execution_file)	F29	Keyword_tftp (Network_connect)	F47	Continent_Code
F13	Keyword_/bin/ [file] (Execution_file)	F30	Keyword_scp (Network_connect)	F48	Country_Name
F14	Keyword_chmod [file] (Set_permissions)	F31	Keyword_ping (Network_connect)	F49	Region_Name
F15	Keyword_sudo su (Set_permissions)	F32	Keyword_kill (Shutdown_action)	F50	City_Name
F16	Keyword_rm [file] (Delect_record)	F33	Keyword_reboot (Shutdown_action)	F51	Longitude
F17	Keyword_history -[suffix] (Delect_record)	F34	Count_base64	F52	Latitude

TABLE 3 Fifty-Two Features

1) Message-Based Features

Features of F1 to F4 are observed an invalid instruction is generated. Features F5 to F7 are accountrelated instructions. Features F8 to F13 are instructions concerning executing files. Features F14 and F15 are directives concerning to promote permissions. Features F16 and F17 are instructions to delete the attack history. Features F18 to F27 are information about the attacker observing the system. Features F28 to F31 are the network-related instructions to download or transmit files. Features F32 and F33 are instructions that affect the state of the system. Features F34 and F35 are used to satisfy the requirements of an attack by obfuscation. Features F36 and F37 calculate the number of URLs in messages and the length of each messages, respectively. F38 is the number of characters entered per second. This study propose eight features. Features F11 to F13 are used to calculate the number of keywords perl [file], python [file], and /bin/ [file], respectively. The functions of Features F35 to F36 were given above. Features 34, 37 and 38 are described below.

F34 Count_base64: An attacker will always use base64 encoding to obfuscate malicious behavior. One of the most common features is that attack scripts are encoded and decoded at execution time. Therefore, this feature is used to determine whether the command Base64 is present in messages. Figure 4 indicates the results of comparison between Base64 encoding and decoding.

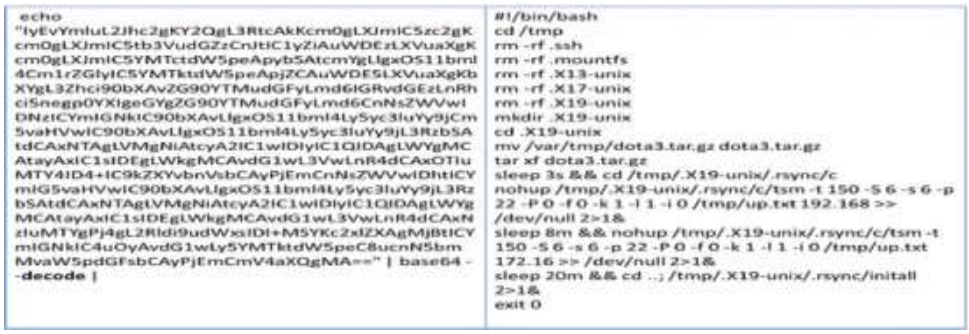

FIGURE 4. Comparison of base64 encoding and decoding.

F37 (Message_length) and F38 (Messages/sec): These two features are used to calculate the total length of a message. A message without any intention is shorter in total than one with a particular intention. The essential purpose of these kinds of commands is to evade detection or to achieve multiple goals. The number of characters entered per second is used to determine whether the attacker is a robot or script execution.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 02 Feb. 2025, pp: 873-892 www.ijaem.net ISSN: 2395-5252

Table 4 presents the details of the message-based features.

ID	Name	Description
F1	*Keyword_bash*	Calculate how many " bash " keywords are in the message.
F2	*Keyword_shell*	Calculate how many " shell " keywords are in the message.
F3	*Keyword_exit*	Calculate how many " exit " keywords are in the message.
F4	*Keyword_help*	Calculate how many " help " keywords are in the message.
F5	*Keyword_passwd [user] (Set_account)*	Calculate how many " passwd [user]" keywords are in the message.
F6	*Keyword_chpasswd (Set_account)*	Calculate how many " chpasswd " keywords are in the message.
F7	*Keyword_useradd -[suffix] (Set_account)*	Calculate how many " useradd -[suffix] " keywords are in the message.
F8	*Keyword_\ [file] (Execution_file)*	Calculate how many "\ [file]" keywords are in the message.
F9	*Keyword_sh [file] (Execution_file)*	Calculate how many " sh [file] " keywords are in the message.
F10	*Keyword_./ [file] (Execution_file)*	Calculate how many "./ [file] " keywords are in the message.
F11	*Keyword_perl [file] (Execution_file)*	Calculate how many " perl [file] " keywords are in the message.
F12	*Keyword_python [file] (Execution_file)*	Calculate how many " python [file] " keywords are in the message.
F13	*Keyword_/bin/ [file] (Execution_file)*	Calculate how many " /bin/ [file] " keywords are in the message.
F14	*Keyword_chmod .[file] (Set_permissions)*	Calculate how many " chmod . [file] " keywords are in the message.
F15	*Keyword_sudo su (Set _permissions)*	Calculate how many " sudo su " keywords are in the message.
F16	*Keyword_rm [file] (Delect_record)*	Calculate how many " rm [file] " keywords are in the message.
F17	*Keyword_history -[suffix] (Delect_record)*	Calculate how many " history -[suffix] " keywords are in the message.
F18	*Keyword_cat /etc (Get_sys_info)*	Calculate how many " cat /etc " keywords are in the message.
F19	*Keyword_uname (Get_sys_info)*	Calculate how many " uname " keywords are in the message.
F20	*Keyword_wc (Get_sys_info)*	Calculate how many " wc " keywords are in the message.
F21	*Keyword_crontab (Get_sys_info)*	Calculate how many " crontab " keywords are in the message.
F22	*Keyword_w (Get_sys_info)*	Calculate how many " w " keywords are in the message.
F23	*Keyword_ps (Get_sys_info)*	Calculate how many " ps " keywords are in the message.
F24	*Keyword_free (Get_sys_info)*	Calculate how many " free " keywords are in the message.
F25	*Keyword_lscpu (Get_sys_info)*	Calculate how many " lscpu " keywords are in the message.
F26	*Keyword_nproc (Get_sys_info)*	Calculate how many " nproc " keywords are in the message.
F27	*Keyword_uptime (Get_sys_info)*	Calculate how many " uptime " keywords are in the message.
F28	*Keyword_wget (Network_connect)*	Calculate how many " wget " keywords are in the message.
F29	*Keyword_tftp (Network_connect)*	Calculate how many " tftp " keywords are in the message.
F30	*Keyword_scp (Network_connect)*	Calculate how many " scp " keywords are in the message.
F31	*Keyword_ping (Network_connect)*	Calculate how many " ping " keywords are in the message.
F32	*Keyword_kill (Shutdown_action)*	Calculate how many " kill " keywords are in the message.
F33	*Keyword_reboot (Shutdown_action)*	Calculate how many " reboot " keywords are in the message.
F34	*Count_base64*	Calculate the number of times the message has been in base64.
F35	*Count_Hex*	Calculate the number of times the message has been in hexadecimal.
F36	*Count_url*	Calculate how many URLs are in the message.
F37	*Message_length*	Calculate the total length of the message.
F38	*Messages / sec*	Calculate the length of Message input per second.

TABLE 4 Message-Based Features **2) Host-Based Features**

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 02 Feb. 2025, pp: 873-892 www.ijaem.net ISSN: 2395-5252

Features F39 to F41 are the communication protocol of the connection, information about the connection, and the version of the connection client, respectively. Features F42 and F43 are related to login information. Features F44 to F46 are the duration, average string length of the response, and the presence or absence of a file during the connection. The authors proposed one feature, F45, in the file type.

F45 Received_Size (AVG): The attacker will always query the content through the Linux command. For example, the user types "uname" returning the string "Linux". The size result is six. The act of stealing information is determined by the length of the string returned.

Table 5 presents an example of the command calculation. Table 6 presents the details of the hostbased features.

Action	Size
CMD: top	35
CMD: uname	6
CMD: crontab -l	33
Total:	74
Received_Size (AVG):	**24.66**

TABLE 5 An Example of Feature F45

ID	Name	Description
F39	*Protocol*	Which communication protocol is it? If it is Telnet, it is 0, and SSH is 1.
F40	*Src_Port*	The port of the source device is connected with the Honeypot
F41	*SSH_Client_Version*	The time from the connection until the user leaves the Honeypot
F42	*Username*	The user's name in the Honeypot.
F43	*Password*	The user's password in the Honeypot.
F44	*Duration*	Time the user stays in the Honeypot.
F45	*Received_Size (AVG)*	The length of the data returned after entering the command and number of discover commands in the Honeypot.
F46	*File*	All files of the Honeypot collect.

TABLE 6 Host-Based Features

3) Geography-Based Features

Features F47 to F50 are, city names that were analyzed globally. Latitude and longitude are used to determine the location of an attack. The corresponding geographic location can be used to determine whether an abnormal attack has occurred. Table 7 presents the details of the geographybased features.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 02 Feb. 2025, pp: 873-892 www.ijaem.net ISSN: 2395-5252

ID	Name	Description
F47	*Continent_Code*	Codes for global continents.
F48	*Country_Name*	The name of a unique territorial subject or political entity.
F49	*Region_Name*	The name of a part of a country.
F50	*City_Name*	The name of a more densely populated and developed area.
F51	*Longitude*	Longitude is a geographic representation of the east-west position of a point on the Earth's surface.
F52	*Latitude*	Latitude is a geographic representation of the north-south position of a point on the Earth's surface.
F53	*Received_Size (AVG)*	The length of the data returned after entering the command and number of discover commands in the Honeypot.
F54	*File*	All files of the Honeypot collect.

TABLE 7 Geography-Based Features

E. Model Training

The proposed AI@NTDS system is designed using the LightGBM algorithm. XGBoost and LightGBM are based on the Tree Boosting mechanism. The LightGBM algorithm is well-known for its better training efficiency and lower memory usage than the XGBoost algorithm. The LightGBM algorithm differs from the traditional Gradient Boosting Decision Tree (GBDT) algorithm and is optimized using various strategies. It has the following four main characteristics; a histogram algorithm, Gradient-based One-Side Sampling (GOSS), Exclusive Feature Bundling (EFB), and Leaf-Wise Tree Growth. The GOSS algorithm is described below Algorithm 1.

Algorithm 1 Gradient-Based One-Side Sampling

Input:
I : training data, d : iterations Input:
a : sampling ratio of large gradient data Input:
b : sampling ratio of small gradient data Input:
loss: loss function, L : weak learner
models ← {}, fact ←1−ab
topN ← a × len(I), randN ← b × len(I)
for i =1 **to** d **do**
preds ← models.predict(I)
g ← loss (I ,preds), w ← {1,1,...}
sorted ← GetSortedIndices(abs(g))
topSet ← sorted[1:topN]
randSet ← RandPick(sorted[topN:len(I)], randN)
usedSet ← topSet + randSet
w[randSet] ×= fact
newModel ← L(I[usedSet], - g[usedSet], w[usedSet])
models.append(newModel)

When the feature data have many dimensions, other GBDT algorithms have the greatest drawbacks like poor efficiency and scalability. In this study, LightGBM is used with GOSS algorithms to solve this problem. GOSS maintains the random sampling on the small gradient and introduces a small gradient constant weight. The asymptotic approximation ratio of GOSS is O(1nji(d)+1njr(d)+1n√) ,and the generalization performance in GOSS is ϵGOSSgen(d)=|Vj(d)+V∗(d)| . For fearure j, Vj(d) is the variance gain for a given decision tree algorithm.

A sorting strategy can optimize the performance without a graph in many features. EFB algorithms reduce the number of dimensions consists two algorithms, which are Greedy Bunding and Merge Exclusive Features (MEF) algorithms. The MEF algorithms can merge multiple features in a bundle by adding the offset parameter. The EFB algorithm reduces the time complexly of original algorithms. It can reduce from O(#data×#feature) to O(#data×#bundle) if #bundle ≪ #feature It accelerates the speed without reducing the accuracy.

In LightGBM processing, feature analysis is performed based on the parameter settings that are shown in Table 8.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 02 Feb. 2025, pp: 873-892 www.ijaem.net ISSN: 2395-5252

Learning Parameter	Value
Boosting	Gbdt
Learning Rate	0.1
Number of estimators	100
Max depth	-1
Number Leaves	31

TABLE 8 The Learning Parameters of LightGBM Algorithm

The concept of GBDT is used to calculate the residuals as a generation decision tree. The most effective learning rate in this work is 0.1. The iterative process revealed that the best number of LightGBM estimators was 100. Since LightGBM grows leaf by leaf based on the tree model, the number of leaves here is set to 31.

IV. PERFORMANCE ANALYSIS

This section concerns the performances of the machine learning (ML) mechanism, feature-based analysis, and AI@NTDS system analysis. Features are analyzed and discussed. The most effective detection model algorithm is identified. The following section provides experimental proof of the result.

A. Analysis of ML Mechanism

Tables 2, 3, and 8 presents the used dataset, features, and learning parameters, respectively. The authors evaluate the AI prediction model with different multi-classification algorithms to assign malicious payloads to three levels.

Table 9 provides various evaluation indexes and the operation time of each machine learning mechanism.

Algorithms	Accuracy	Precision	Recall	F1-Score	Computation Time
Naive Bayes	11.83%	99.11%	67.72%	71.15%	2.07s
SVM	26.67%	93.53%	67.60%	76.65%	8.86s
Decision Tree	92.51%	98.10%	97.91%	97.99%	1.42s
Random Forest	98.62%	99.76%	99.67%	99.72%	57.83s
XGBoost	98.72%	99.72%	99.72%	99.72%	44.25s
LightGBM	98.72%	98.78%	98.68%	99.73%	5.51s

TABLE 9 Performance of Different ML Mechanisms

Many machine learning classification algorithms in the security domain use SVM and Random Forest. In previous works, a Decision Tree and naive Bayes algorithm are used to detect this issue .XGBoost and LightGBM are also well-known classification algorithms. In this work, each algorithm with default parameters are compared in terms of accuracy, precision, recall, F1-score, and computation time metrics.

Following a comprehensive evaluation, the LightGBM was used as the proposed AI@NTDS learning model because the values of any indicator in this study measured were better than average. It required the least computation time, making it easier to deploy in various devices for real-time detection.

Naive Bayes and SVM perform poorly. Although these two algorithms are widely used, they are not suitable for the detection of malicious shell commands in this study.

B. Analysis of Feature-Based

The features of the AI@NTDS system are divided into host-based, message-based, and geographybased groups. One of the purposes of this group is to find the classification model and identify the essential features.

Four case studies were performed, and the results of the relevant analysis are provided in Table 10. In Case 1, message-based features alone resulted in good performance of accuracy and

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 02 Feb. 2025, pp: 873-892 www.ijaem.net ISSN: 2395-5252

precision. An attacker may attack a target by such means as causing confusion, recon, and deletion.

These actions cause the attacker's input character to be more than a low-level threat. In Case 2, only host-based features are used. This study contributed significantly to the returned strings and SSH-related information. Case 3 identified the essential features by observing the distribution of attackers based on geographical features. Latitude and longitude were the critical features, but the accuracy of the model using these features and other indicators combined with all of the features were not as high as in the preceding two case studies The results show that risks and hazards cannot be assessed using only the features that are associated with geographical location. In Case 4, all of the features were used, and the best values of all indicators were obtained.

Testing Features	Accuracy	Precision	Recall	F1_score	Log loss	AUC
Message-based Features	98.19%	99.69%	99.69%	98.18%	3.3474	98.20%
Host-based Features	90.07%	97.07%	97.74%	97.39%	4.8581	83.50%
Geography-based Features	83.82%	91.44%	99.89%	95.35%	3.4473	50.57%
All Features	99.20%	99.75%	99.85%	99.80%	2.9773	98.53%

TABLE 10 Feature Analysis With 4 Studies

Finally, 52 dimensions are used to analyze the three types of features to obtain the best model of the detecting system, based on the feature engineering with gradient boosting machine, as shown in Figure 5. The message-based features account for about 50% of the ten features; these are Message_length (F37) and \ . \w*(F5) features. The host-based features account for 40% of the top ten features; these are Received_size (F45) and duration (F41) features. A 99.75% precision, 99.85% recall, and F1-score of 99.80% are achieved. Therefore, the evaluation of AI@NTDS model is based mainly on host-based and message-based features. The proposed features are proved to be very effective.

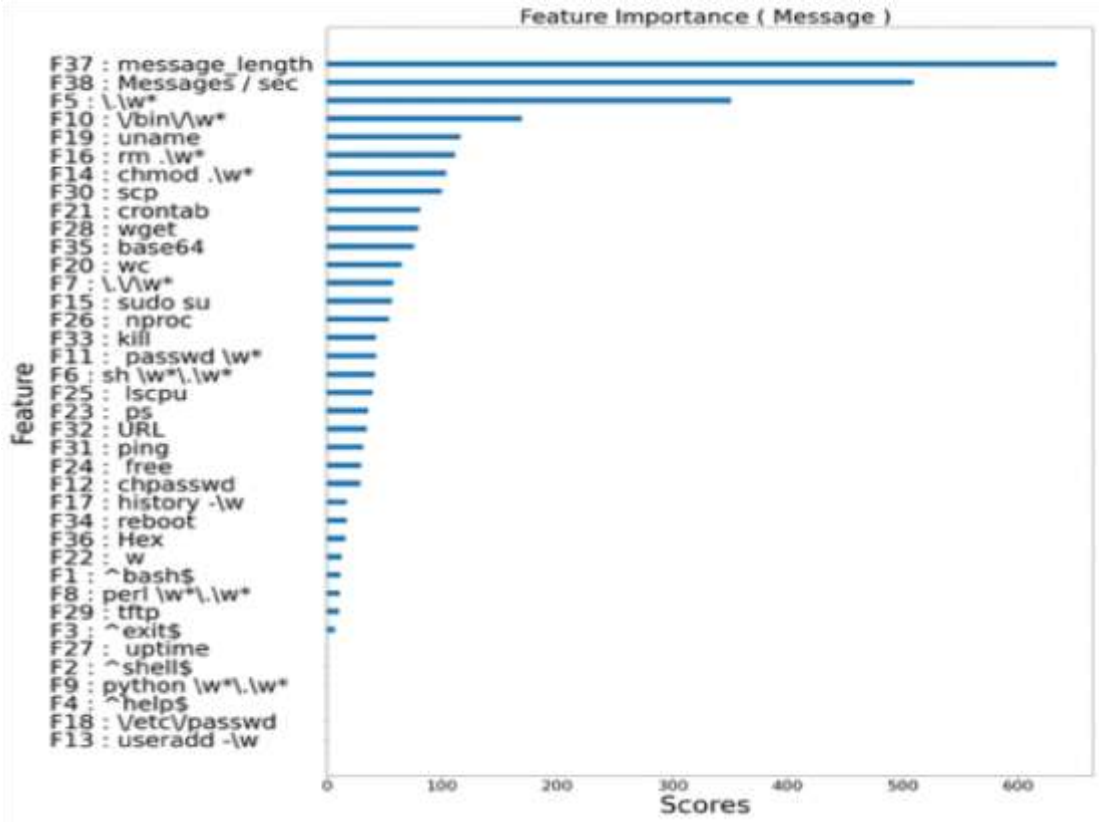

FIGURE 5 Feature importance analysis.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 02 Feb. 2025, pp: 873-892 www.ijaem.net ISSN: 2395-5252

C. Analysis of AI@NTDS System

The test dataset comprises 23% of the data in all experiment datasets. The training set comprises data from 2019, and the test set consists of data from 2020.

The AI@NTDS classifier predicts the classification of each threat in the test dataset, yielding the results in Table 11. From the confusion matrix, the total misclassification ratio of the classifier for threat level 1 is 0.17%; that for threat level 2 is 0.37%, and that for threat level 3 is 0.86%. The F1-score reaches 99.80%, indicating that the AI@NTDS effectively detected samples of various threats. The AUC (Area Under the Curve) reaches 98.53%; the precision rate can reaches 99.75%, and the recall rate reaches 99.85%. Therefore, the detection model that is trained using the LightGBM algorithm can detect malicious sample changes in various periods of attack and has excellent efficiency and performance.

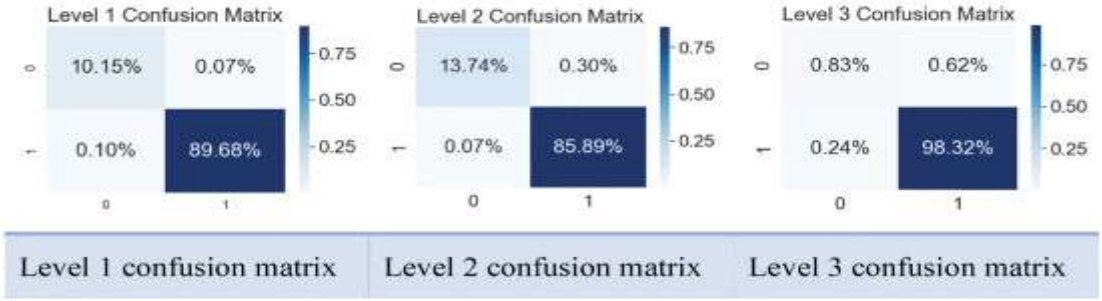

TABLE 11 Performance of AI@NTDS System

D. Comparison With Other Studies

Table 12 compares the performance of the proposed AI@NTDS classifier with those of related methods and algorithms in previous studies. Since previous studies have not provided detailed parameter settings and features of each mechanism, the parameter settings of Random Forest and K-NN were those used in their closest methods when they were originally developed. These mechanisms were compared using the same dataset. The accuracy of the AI@NTDS model is 4% higher than those Random Forest and K-NN, and the F1-score is 1% better. Therefore, the AI@NTDS classifier with the LightGBM algorithm is the most effective in classifying threat levels and identifying the attacker's intent. The difficulties of implementation and the number of data dimensions must be addressed are also compared. Although the model herein yielded better results than both of the other, it requires more features to be extracted from the dataset. Therefore, the proposed mechanism requires more time to spend in preprocessing data. All of the experiment datasets used in Table 12 use the Zenodo dataset, based on the Cowrie Honeypots. The proposed AI@NTDS in this study can be applied to any real-world scenario involves IoT devices and Linux server shell command analysis.

Title	Our proposed system	Identification and classification of cyber threats through ssh honeypot systems (2020)	Detection of Malicious Remote Shell Sessions (2019)
Dataset	**Zenodo Dataset**		
Purpose	The study is classified according to different threat levels.		Only benign and malicious.
Algorithm	**LightGBM**	Random Forest	K-NN
Accuracy	**99.20%**	95.664%	94.08%
F1 Score	**99.80%**	98.76%	98.72%
Implementation	Difficult	Medium	**Simple**
Dimension	Medium	**Small**	Large

TABLE 12 Comparison With Different Studies

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 02 Feb. 2025, pp: 873-892 www.ijaem.net ISSN: 2395-5252

Many studies of related issues have bee performed. Table 13 presents problem-solving mission with various features. The mechanisms that are listed in the first, fourth, and fifth columns have similar purposes and are analyzed in detail Table 12 presents.

TABLE 13 Comparison With Related Works

V. CONCLUSION

This study proposed an AI@NTDS detection system that incorporates the LightGBM machine learning algorithm for identifying and classifying threats. Attackers' intentions are analyzed using collected data, and the degree of harm that is caused by malicious instructions is determined. Three types of attack are identified by threat levels of attack are identified using Enterprise Tactics of MITRE ATT&CK. A total of 52 features of three types - message-based, host-based, and geography-based features - are ultimately identified. The results of an analysis demonstrate that our model performed best when all features were used. Message-based features and host-based features accuracy for the model are largest.

REFERENCE:

[1]. G. K. Sadasivam, C. Hota and B. Anand, "Classification of SSH attacks using machine learning algorithms", Proc. 6th Int. Conf. IT Converg. Secur. (ICITCS), pp. 1-6, Sep. 2016.

[2]. Miller, S., & Wright, E. (2021). "Data-Driven Security: The Future of Cyber Defense." International Journal of Information Security, 20(1), 55-73.

[3]. Adebayo VI, Ige AB, Idemudia C, Eyieyien OG. Ensuring compliance with regulatory and legal requirements through robust data governance structures. Open Access Res J Multidiscip Stud. 2024;8(1):36–44. doi:10.53022/oarjms.2024.8.1.0043.

[4]. Khan, Noman&Aslam, Saleem. (2025). A Psychometric Approach to Occupational Health and Safety: Developing a Validated Scale for Construction Firms. 10.13140/RG.2.2.34080.08968.

[5]. Fawzy, H. A., ALakkad, A., &Sarwar, M. S. (2022). Ascarislumbricoides infestation of bile ducts: case report. Asian Journal of Research in Medical and Pharmaceutical Sciences, 11(4), 56-61.

[6]. Brijesh Pandya. (2025). Using cad systems for automating the design and prototyping process. The American Journal of Engineering and Technology, 7(02), 6–11. https://doi.org/10.37547/tajet/Volume07Issue02-02

[7]. Afolabi AI, Hussain NY, Austin-Gabriel B, Adepoju PA, Ige AB. Geospatial AI and data analytics for satellitebased disaster prediction and risk assessment. Open Access Res J Eng Technol. 2023. doi:10.53022/oarjet.2023.4.2.0058.

[8]. Madathala, H., Barmavat, B., &Thumala, S. (2023). Performance Optimization of SAP HANA using AI-based Workload Predictions. International Journal of Innovative Research in Science,

Engineering and Technology, 12, 15315-15326.

[9]. Vel, D. V. T., Durgaraju, S., &Madathala, H. INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING.

[10]. Thumala, Srinivasarao. (2021). Running Sustainable Virtual Desktop Infrastructure (VDI) Solutions in the Cloud. International Journal on Recent and Innovation Trends in Computing and Communication.

[11]. Vel, T. (2021). AI-Driven Adaptive Authentication for Multi-Modal Biometric Systems. J. Electrical Systems, 17(1), 75-88.

[12]. Ahmed S, Khan M. Securing the Internet of Things (IoT): A comprehensive study on the intersection of cybersecurity, privacy, and connectivity in the IoT ecosystem. AI, IoT and the Fourth Industrial Revolution Review. 2023;13(9):1–17.

[13]. Akinade AO, Adepoju PA, Ige AB, Afolabi AI, Amoo OO. Advancing segment routing technology: A new model for scalable and low-latency IP/MPLS backbone optimization.

[14]. Alao OB, Dudu OF, Alonge EO, Eze CE. Automation in financial reporting: A conceptual framework for efficiency and accuracy in US corporations. Global J Adv Res Rev. 2024;2(2):40–50.

[15]. Angelopoulos A, Michailidis ET, Nomikos N, Trakadas P, Hatziefremidis A, Voliotis S, et al. Tackling faults in the industry 4.0 era—a survey of machine-learning solutions and key aspects. Sensors. 2019;20(1):109.

[16]. Austin-Gabriel B, Hussain NY, Ige AB, Adepoju PA, Amoo OO, Afolabi AI. Advancing zero trust architecture with AI and data science for enterprise cybersecurity frameworks. Open Access Res J Eng Technol. 2021;1(1):107. doi:10.53022/oarjet.2021.1.1.0107.

[17]. Bhattacharjee S. Practical industrial internet of things security: A practitioner's guide to securing connected industries. Packt Publishing Ltd; 2018.

[18]. Boyes H, Hallaq B, Cunningham J, Watson T. The industrial internet of things (IIoT): An analysis framework. Comput Ind. 2018;101:1–12.

[19]. Chukwurah N, Ige AB, Idemudia C, Eyieyien OG. Integrating agile methodologies into data governance: Achieving flexibility and control simultaneously. Open Access Res J Multidiscip Stud. 2024;8(1):45–56. doi:10.53022/oarjms.2024.8.1.0044.

[20]. George EP-E, Idemudia C, Ige AB. Blockchain technology in financial services: Enhancing security, transparency, and efficiency in transactions and services. Open Access Res J Multidiscip Stud. 2024;8(1):26–35. doi:10.53022/oarjms.2024.8.1.0042.

[21]. George EP-E, Idemudia C, Ige AB. Recent advances in implementing machine learning algorithms to detect and prevent financial fraud in real-time. Int J Eng Res Dev. 2024;20(07).

[22]. Hassan SK, Ibrahim A. The role of artificial intelligence in cyber security and incident response. Int J Electron Crime Investig. 2023;7(2).

[23]. Homaei M, Mogollón-Gutiérrez Ó, Sancho JC, Ávila M, Caro A. A review of digital twins and their application in cybersecurity based on artificial intelligence. ArtifIntell Rev. 2024;57(8):201.

[24]. Hussain NY, Austin-Gabriel B, Ige AB, Adepoju PA, Amoo OO, Afolabi AI. AI-driven predictive analytics for proactive security and optimization in critical infrastructure systems. Open Access Res J Sci Technol. 2021;2(2):59. doi:10.53022/oarjst.2021.2.2.0059.

[25]. Dey, R., Roy, A., Akter, J., Mishra, A., & Sarkar, M. (2025). AI-Driven Machine Learning for Fraud Detection and Risk Management in US Healthcare Billing and Insurance. Journal of Computer Science and Technology Studies, 7(1), 188-198.

[26]. Ige AB, Austin-Gabriel B, Hussain NY, Adepoju PA, Amoo OO, Afolabi AI. Developing multimodal AI systems for comprehensive threat detection and geospatial risk mitigation. Open Access Res J Sci Technol. 2022;6(1):63. doi:10.53022/oarjst.2022.6.1.0063.

[27]. Ige AB, Chukwurah N, Idemudia C, Adebayo VI. Managing data lifecycle effectively: Best practices for data retention and archival processes. Int J Eng Res Dev. 2024;20(8):199– 207.

[28]. Ike CC, Ige AB, Oladosu SA, Adepoju PA, Amoo OO, Afolabi AI. Advancing

machine learning frameworks for customer retention and propensity modeling in ecommerce platforms. GSC Adv Res Rev. 2023;14(2):17. doi:10.30574/gscarr.2023.14.2.0017.

[29]. Munirathinam S. Industry 4.0: Industrial internet of things (IIoT). In: Advances in Computers, Vol. 117. Elsevier; 2020. p. 129–164.

[30]. Ofoegbu KDO, Osundare OS, Ike CS, Fakeyede OG, Ige AB. Data-driven cyber threat intelligence: Leveraging behavioral analytics for proactive defense mechanisms.

[31]. Ofoegbu KDO, Osundare OS, Ike CS, Fakeyede OG, Ige AB. Empowering users through AI-driven cybersecurity solutions: Enhancing awareness and response capabilities. Open Access EngSciTechnol J. 2024;4(6):707–727. doi:10.51594/estj.v4i6.1528.

[32]. Ogunbiyi-Badaru O, Alao OB, Dudu OF, Alonge EO. Blockchain-enabled asset management: Opportunities, risks, and global implications.

[33]. Ogunbiyi-Badaru O, Alao OB, Dudu OF, Alonge EO. The impact of FX and fixed income integration on global financial stability: A comprehensive analysis.

[34]. Ojukwu P, Cadet E, Osundare O, Fakeyede O, Ige A, Uzoka A. The crucial role of education in fostering sustainability awareness and promoting cybersecurity measures. Int J Frontline Res Sci Technol. 2024;4(1):18– 34.

[35]. Ojukwu PU, Cadet E, Osundare OS, Fakeyede OG, Ige AB, Uzoka A. Advancing green bonds through fintech innovations: A conceptual insight into opportunities and challenges. Int J Eng Res Dev. 2024;20:565–576.

[36]. Oladosu SA, Ige AB, Ike CC, Adepoju PA, Amoo OO, Afolabi AI. Reimagining multi-cloud interoperability: A conceptual framework for seamless integration and security across cloud platforms. Open Access Res J Sci Technol. 2022;4(1):26. doi:10.53022/oarjst.2022.4.1.0026.

[37]. 27. Oladosu SA, Ige AB, Ike CC, Adepoju PA, Amoo OO, Afolabi AI. AI-driven security for next-generation data centers: Conceptualizing autonomous threat detection and response in cloud-connected environments. GSC Adv Res Rev. 2023;15(2):162–172. doi:10.30574/gscarr.2023.15.2.0136.

[38]. 28. Oladosu SA, Ike CC, Adepoju PA, Afolabi AI, Ige AB, Amoo OO. Advancing cloud networking security models: Conceptualizing a unified framework for hybrid cloud and on-premise integrations.

[39]. 29. Oladosu SA, Ike CC, Adepoju PA, Afolabi AI, Ige AB, Amoo OO. Frameworks for ethical data governance in machine learning: Privacy, fairness, and business optimization.

[40]. 30. Onoja JP, Ajala OA. Innovative telecommunications strategies for bridging digital inequities: A framework International Journal of Multidisciplinary Research and Growth Evaluation www.allmultidisciplinaryjournal.com 1164 | P a g e for empowering underserved communities. GSC Adv Res Rev. 2022;13(1):210–217. doi:10.30574/gscarr.2022.13.1.0286.

[41]. 31. Onoja JP, Ajala OA. AI-driven project optimization: A strategic framework for accelerating sustainable development outcomes. GSC Adv Res Rev. 2023;15(1):158–165. doi:10.30574/gscarr.2023.15.1.0118.

[42]. 32. Osundare OS, Ige AB. Optimizing network performance in large financial enterprises using BGP and VRF-lite. Int J Scholarly Res Sci Technol. 2024;5(1).

[43]. 33. Osundare OS, Ike CS, Fakeyede OG, Ige AB. Active/Active data center strategies for financial services: Balancing high availability with security. Open Access ComputSci IT Res J. 2024;3(2):92–114. doi:10.51594/csitrj.v3i3.1494.

[44]. 34. Osundare OS, Ike CS, Fakeyede OG, Ige AB. Application of machine learning in detecting fraud in telecommunication-based financial transactions. Open Access ComputSci IT Res J. 2024;4(3):458–477. doi:10.51594/csitrj.v4i3.1499.

[45]. 35. Osundare OS, Ike CS, Fakeyede OG, Ige AB. Evaluating core router technology upgrades: Case studies from telecommunications and finance. Open Access ComputSci IT Res J. 2024;4(3):416–435. doi:10.51594/csitrj.v4i3.1497.

[46]. 36. Osundare OS, Ike CS, Fakeyede OG, Ige AB. Secure communication protocols

for realtime interbank settlements. Open Access ComputSci IT Res J. 2024;4(3):436–457. doi:10.51594/csitrj.v4i3.1498.

[47]. 37. Sarker IH. Deep cybersecurity: A comprehensive overview from neural network and deep learning perspective. SN Comput Sci. 2021;2(3):154.

[48]. 38. Shahin M, Maghanaki M, Hosseinzadeh A, Chen FF. Advancing network security in industrial IoT: A deep dive into AI-enabled intrusion detection systems. AdvEng Informatics. 2024;62:102685.

[49]. 39. Vijay AJ, William BNJ, Haruna AA, Prasad DD. Exploring the synergy of IIoT, AI, and data analytics in Industry 6.0. In: Industry 6.0. CRC Press; 2024. p. 1–36.

[50]. O'Neill, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishing Group.

[51]. Kumar, A., Yadav, A., & Singh, D. (2020). Reinforcement Learning in Cybersecurity: A Survey. Future Generation Computer Systems, 112, 556-566.

[52]. Kumar, A., & Singh, D. (2021). User Behavior Analytics: A Data-Driven Approach to Cybersecurity. Journal of Cybersecurity and Privacy, 1(2), 123-135.

[53]. Kumar, A., Kumar, A., & Rani, P. (2018). Comparative Analysis of Traditional and AIBasedCybersecurity Techniques. International Journal of Computer Applications, 182(7),

[54]. 20-27.Mäntylä, M. V., Määttä, S., &Taalas, P. (2022). Natural Language Processing in Cybersecurity: A Review of Techniques and Applications. Computers & Security, 113, 102530

[55]. T. Bajtoš, P. Sokol and T. Mézešová, "Virtual honeypots and detection of TelNet botnets", Proc. Central Eur.Cybersecur. Conf., pp. 1-6, Nov. 2018.

[56]. R. M. Arifianto, P. Sukarno and E. M. Jadied, "An SSH honeypot architecture using port knocking and intrusion detection system", Proc. 6th Int. Conf. Inf. Commun. Technol. (ICoICT), pp. 409-415, May 2018.

[57]. R. K. Shrivastava, B. Bashir and C. Hota, "Attack detection and forensics using honeypot in IoT environment", Proc. Int. Conf. Distrib. Comput. Internet Technol., pp. 402-409, Jan. 2019.

[58]. P. Dumont, R. Meier, D. Gugelmann and V. Lenders, "Detection of malicious remote shell sessions", Proc. 11th Int. Conf. Cyber Conflict, pp. 1-20, May 2019.

[59]. S. Udhani, A. Withers and M. Bashir, "Human vs bots: Detecting human attacks in a honeypot environment", Proc. 7th Int. Symp. Digit. Forensics Secur. (ISDFS), pp. 1-6, Jun. 2019.

[60]. B. Lingenfelter, I. Vakilinia and S. Sengupta, "Analyzing variation among IoT botnets using medium interaction honeypots", Proc. 10th Annu. Comput. Commun. Workshop Conf. (CCWC), pp. 761-767, Jan. 2020.

[61]. J. T. M. Garre, M. G. Pérez and A. Ruiz-Martínez, "A novel machine learning-based approach for the detection of SSH botnet infection", Future Gener. Comput. Syst., vol. 115, pp. 387-396, Feb. 2021.

[62]. B.-X. Wang, J.-L. Chen and C.-L. Yu, "Cyber security threat intelligence monitoring and classification", Proc. IEEE Int. Conf. Intell. Secur. Informat. (ISI), pp. 1-3, Nov. 2021.

[63]. A. O. Alzahrani and M. J. F. Alenazi, "Designing a network intrusion detection system based on machine learning for software defined networks", Future Internet, vol. 13, no. 5, pp. 111, Apr. 2021.

[64]. S. Iranmanesh, F. S. Abkenar, A. Jamalipour and R. Raad, "A heuristic distributed scheme to detect falsification of mobility patterns in internet of vehicles", IEEE Internet Things J., vol. 9, no. 1, pp. 719-727, Jan. 2022.

[65]. B. Lingenfelter, I. Vakilinia and S. Sengupta, "Analyzing variation among IoT botnets using medium interaction honeypots", Proc. 10th Annu. Comput. Commun. Workshop Conf. (CCWC), pp. 761-767, Jan. 2020.

[66]. J. T. M. Garre, M. G. Pérez and A. Ruiz-Martínez, "A novel machine learning-based approach for the detection of SSH botnet infection", Future Gener. Comput. Syst., vol. 115, pp. 387-396, Feb. 2021.

[67]. B.-X. Wang, J.-L. Chen and C.-L. Yu, "Cyber security threat intelligence monitoring and classification", Proc. IEEE Int. Conf. Intell. Secur. Informat. (ISI), pp. 1-3, Nov. 2021.

[68]. A. O. Alzahrani and M. J. F. Alenazi, "Designing a network intrusion detection system based on machine learning for software defined networks", Future Internet, vol. 13, no. 5, pp. 111, Apr. 2021.

[69]. S. Iranmanesh, F. S. Abkenar, A. Jamalipour and R. Raad, "A heuristic distributed scheme to detect falsification of mobility patterns in internet of vehicles", IEEE Internet Things J., vol. 9, no. 1, pp. 719-727, Jan. 2022.