Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Proactive Cyber Threat Detection Using AI and Open-Source Intelligence

Jafrin Reza[1], Md Imran Khan[2], and Sanjida Akrer Sarna[3]✉
[13]Master of Science in Business Analytics, Trine University, USA
[2]Master of Science in information studies, Trine University, USA
Corresponding Author: Jafrin Reza, **E-mail**: jafrinreza@gmail.com

| ABSTRACT

Frequent developments in cyber threats seriously threaten the digital systems in both the public and private sectors. Today, modern cyberattacks are too unpredictable for the old cybersecurity defenses and time-bound detection methods. Because there are more complex, numerous and distant threats today, to find them and address them before much damage can occur. In this work, look at integrating AI and OSINT to develop a system that can quickly detect any cyber threats in an organization. The researchers used the Hornet 40 dataset which includes network traffic collected over the course of 40 days from honeypots in eight places: Amsterdam, London, Frankfurt, San Francisco, New York, Singapore, Toronto, and Bangalore. To capture different activities from uninvited users, these honeypots received requests only on a specific non-standard SSH port. The information provided by Argus is in the form of detailed bidirectional NetFlow data that displays the effects of geography on various cyber-attacks. Various machine learning approaches are used within a data-driven system to spot and detect abnormal traffic and threats in the network such as Random Forest, Support Vector Machines (SVM), Long Short-Term Memory (LSTM) networks and Isolation Forests. At the same time, data, and findings from public threat intelligence, darknet sources and cybersecurity forums are studied using Natural Language Processing (NLP) to find important information about threats. As a result of this, the detection rate is improved by comparing suspicious traffic in honeypots with global findings and the reported IOCs. Combining AI and OSINT together allows the engine to read and analyze a lot of network data quickly and in almost real time. Joining these processes allows quick and early identification of advanced attacks such as zero-day attacks and intrusions. It is clear from the results that using this approach improves the accuracy of detection, lowers the number of false positives, and reveals attacks that tend to come from specific locations and are typically overlooked by other systems.

| KEYWORDS

Cybersecurity, Artificial Intelligence, Open-Source Intelligence (OSINT), Threat Detection, Anomaly Detection and Honeypot Data Analysis

| ARTICLE INFORMATION

ACCEPTED: 19 May 2025 **PUBLISHED:** 03 June 2025 **DOI:** 10.32996/jcsts.2025.7.5.62

1. Introduction

1.1 Background of Cyber Threats on the Rise

Today, there are more and more complex cyber threats due to advancements in technology. Now, thanks to ransomware attacks, zero-day issues, DDoS and APTs, effective defense against cyberattacks is often not enough, with serious aftermath. Since data is shared across countries, cloud usage is common and IoT devices are everywhere, companies now face huge risks due to frequent cyber-attacks. While this is happening, increased tension in global politics has encouraged different governments to sponsor cyber spying and malicious actions which makes many cyber-attacks more significant and intentions [1]. According to reports on global cybersecurity, many records are affected each year, resulting in large financial losses, a damaged reputation and trouble for operations. All types of sectors such as finance, healthcare, energy, and government, are required to

protect their confidential information and infrastructure while complying with stricter regulations. This threat becomes even greater because of new types of malwares that are hard to identify. Therefore, a cybersecurity method based on predictions and quick responses is required, to stop threats from causing significant harm before they erupt.

1.2 the Drawbacks of Today's Planning-Based Cybersecurity Approaches

Typically, cybersecurity is based on steps that come after an attack, like signature-based antivirus programs, firewall rules set in advance and IDS that need to be updated by hand. In the past, these measures were reliable, but attackers are becoming more creative by using tricks, encrypting their actions, and employing different ways to stay unnoticed. Being based on past experiences, static defense cannot recognize or defend against threats that appear suddenly without any prior notice or description [2]. These systems also tend to flood security teams with many false alarms, causing team members to grow tired of dealing with them. The traditional approach to cybersecurity also divides operations and often feeds isolated threat intelligence to alerts, so people do not understand today's bigger threats. Meanwhile, to attack organizations in complex ways, attackers use automated systems, rely on machine learning, and share resources with other criminals on the dark web [3]. When the time between first identifying a threat and acting against it is too long, it increases the chances of risk. Because threats are always evolving, it is obvious that reactive strategies will not stay effective in today's world. In today's world, it is crucial to have systems that find known threats and can instantly learn about and overcome new problems that occur.

1.3 How Artificial Intelligence and OSINT help in Cybersecurity

In modern times, the use of AI by cybersecurity teams, along with OSINT, is helping to prevent cyber threats, identify them and solve problems. Machine learning, deep learning and natural language processing allow AI to automatically process data from different sources, identifying anything suspicious that could point to threats [3]. Network intrusions, attack opportunities and urgent risks can be detected more accurately and efficiently by using algorithms that hardly require human assistance. At the same time, OSINT relies on public data, for example, the threat intelligence feed, different forums, repositories, and social media, to detect potential threats before they target specific entities. AI and OSINT allow different datasets to be linked, including internal honeypot logs and IOCs found elsewhere online [4]. As a result of this, team members can see the threat landscape more clearly and choose the best steps to guard against attacks before they cause any damage. The development and improvement of AI-driven technology depends on AI systems' adjustments to new security issues [5]. For this reason, having AI work with OSINT in cybersecurity could make us more capable of predicting attacks and playing a critical role in today's risk management methods.

1.4 Objectives of the Research

The objective of this study is to achieve these objectives.

- To measure how well AI-based algorithms detect unusual activities on a network using information from honeypots.
- To see how OSINT can add helpful details to the alarms provided by network logs.
- To combine AI with OSINT into a model that quickly notices and brings attention to cyber threats [6].
- To study attacks from different parts of the world with the Hornet 40 dataset.
- To measure the precision, recall and false positive rate of systems that use AI.
- To investigate ethical aspects related to using publicly available data for automated defense.
- To offer suggestions for the use of intelligent and adjustable cybersecurity systems.

1.5 Problem Statement

Most traditional cybersecurity systems use a reactive approach and cannot effectively block new types of cyber threats [7]. They do not respond well to new challenges and are seldom able to use external information in real time. Since attackers can exploit weaknesses in many places using automated systems, there is still a missing part in the current approach to defense. To be effective, organizations require a system that watches for threats both within the network and from external sources at the same time.

1.6 Research Questions

This study demonstrates on these following questions:

- How well are machine learning algorithms able to spot unusual activity in network traffic taken from honeypots all over the world?
- How can OSINT contribute to more accurate detection of cyber threats by AI?
- What are the issues of using AI together with OSINT for real-time cybersecurity?

1.7 Significant of the Study

With this study, how modern cybersecurity can be improved with the combination of AI and OSINT. Following this step, it is easier to detect threats, helping organizations avoid dependence on old and unreliable approaches. Using the Hornet 40 dataset, the study sheds light on how the location of attacks is linked to their behavior, benefiting companies with activities in several locations [8]. Based on the research, new rules, allocation of resources and approaches to dealing with crises can be created in both private and public sectors. Besides, research encourages companies to rely on predictive cybersecurity methods that save them from major breaches and limit the interruption from threats that are found at an early stage. Reliance on AI and OSINT follows the latest trends in cybersecurity and provides a reliable approach for businesses to stay one step ahead in cybersecurity [9]. The research gives rise to intelligent, automated, and ethical cybersecurity measures for use in the future.

2. Literature Review

2.1 Overall Outline of Work Done on AI for Cybersecurity

Since AI can process large amounts of information and detect advanced security threats, it is playing an important role in today's cybersecurity [10]. Studies reveal that AI simplifies detecting threats and enhances the response to incidents in a timely manner. Decision trees, random forests and support vector machines are among the algorithms that have proved accurate in identifying malware and spotting anomalies. Identity of both zero-day and advanced persistent threats can be determined using models such as convolutional and recurrent neural networks. For machine learning, both types of learning are applied, though ensemble techniques are now preferred more because they work well in most cases. Reinforcement learning has demonstrated the ability to deal with dangers as they appear in real-time events [11]. AI is utilized in fields such as intrusion detection, monitoring behaviors, preventing fraud, and detecting phishing scams. Even with all these new technologies, they encounter difficulties. Threat intelligence using AI often relies on one type of fixed data, so it struggles to cope with changes in threats. It is also important to note that complicated models may be unclear to others and can be influenced by intentional mistakes. Also, most existing frameworks work by themselves, without using any information from other sources [12]. Integrated and flexible systems must start using OSINT and similar intelligence streams to improve security measures.

2.2 Analysis of OSINT for Gathering Threat Intel

With OSINT, cybersecurity can strengthen threat awareness efficiently and at a lower cost. use OSINT that includes social media, forums, blogs, IP registries and dark web platforms, quickly learn about challenges that might threaten the system. Such tools can review various places for suspicious signs and detect Indicators of Compromise (IOCs), domain misuse, as well as first signs of a group strike. Studies point out that OSINT makes it possible to discover who was responsible for a cyber-attack, how they did it and to predict imminent threats. It has also supported actions in security assessments and when investigating cybercrimes. With OSINT, security teams can confirm their own research and learn more about the circumstances they are investigating [13]. Operational cybersecurity tools have yet to include them. It is often a problem that data isn't authentic, collected data can be messy and texts are written in numerous languages online. There are also concerns about monitoring and using people's data ethically and legally. Its absence in current threat detection methods relates to its lack of standardization and little automation [14]. Chance to improve constant protection of online systems is lost. Widely acknowledged is the fact that AI-powered systems will become necessary for handling and confirming information collected through OSINT.

2.3 Applying Machine Learning, NLP and Data Mining helps with identifying threats

Because of the merging of ML, NLP and data mining, cybersecurity has moved forward. ML allows for the automatic spotting of harmful patterns, either by identifying known classes or finding unusual features [15]. For cyber security, detecting cyber intrusions, phishing and malware largely relies on logistic regression, k-nearest neighbors, and support vector machines. It has been effective for collecting structured information from text-based reports, messages, and social media discussions. Systems using sentiment analysis, tokenization and named entity recognition can detect suspicious behavior, hostility and ties between different persons or groups in society. NLP supports the identification of false alarms and the detection of phishing activities by identifying common text features in them. Data mining discovers patterns and relationships present in large sets of data. Mining's the logs help in finding unusual events linked to cybersecurity threats [16]. They work together to strengthen security systems' real-time alerts and improve their flexibility. Still, data preprocessing, scaling the model and computer processing are among the difficulties computer vision solutions must handle. It takes specific infrastructure and knowledge to integrate these technologies into one system. Consequently, if ML, NLP and data mining are used properly, they can help create cybersecurity tools that adjust to new dangers.

2.4 Gap Identify in Current Practices Addressed by This study

AI and OSINT methods are considered reliable for cybersecurity, yet they have some weaknesses that limit their potential. Most AI systems work with ready-set, pre-processed data rather than current, real-world intelligence. For that reason, they cannot keep up with new threats or challenges [17]. Since this data is typically unstructured and multilingual, it is hard to

integrate it using machines. Thus, a lot of key information from threat communities online goes unnoticed. Machine learning and NLP tools are not always used consistently in many roles. Most studies focus on one area of AI at a time, ignoring the extra benefits that come when different areas of AI are combined with external intelligence. Since geography is rarely considered, cyber threat models tend to be ineffective in certain regions. Most existing frameworks experience challenges due to false positives, loss of context and poor scaling. Real-world data and geography are often missing from the samples used for empirical studies. This issue requires a system that can understand all types of data, is integrated with different technologies [18]. The paper's proposed solution involves mixing AI approaches with OSINT data by using the Hornet 40 dataset. With the help of this data, investigate how cyber-attacks vary depending on the region honeypots are located. It depends on NLP to understand digital evidence gathered from the internet and employs machine learning to link what the network detects with information from external sources [19]. The target is to set up a capable and flexible framework that helps secure the system, reduces errors and addresses any updates in the threats facing cybersecurity.

2.5 Empirical Study

The article by Nan Sun et al. in title on Cyber Threat Intelligence Mining for Proactive Cybersecurity Defense: A Survey and New Perspectives explores how CTI mining can be used proactively to protect systems [1]. Authors explain that trying to fend off cybercrimes by picking them up after they occur isn't enough. It encourages a switch to using AI and CTI techniques in security decisions. CTI mining strategies are classified into five sections related to cyber entities, tactics and procedures used by attackers, indicators of compromise, analysis of malware and investigation into threat actors. Importantly, organizations fail to use CTI to its full extent by integrating it mainly with firewalls and SIEMs. The current study has the same goal as this research: to use AI models and data collected from the internet to enhance the identification of threats. By including the article's perspective, this research is based on real-life problems like data problems, carrying out analysis to a large scale and dealing with diverse data. Therefore, advanced systems using AI and OSINT should be created to help anticipate cyber threats as they occur.

The paper by Hammad Raza, Proactive Cyber Defense with AI, focuses on the important role AI plays in modern cybersecurity. In the paper, it is pointed out that AI brings real-time solutions, helping to identify, examine and address security threats before they become serious problems. Raza explicates how AI helps in handling generous data from security tools to spot unusual activity that is normally unnoticed by security professionals. This illustrates exactly the goal of the present study which looks at AI and OSINT together for discovering potential cyber threats [2]. The article explains that AI can improve its predictions about future threats by observing changes in the threat landscape. Using AI to automate simple security actions speeds up the response and gives specialists more time to handle efforts that need more attention. As a result, the research framework is strengthened by showing that AI helps to detect cyber security risks by processing useful data along with open-source threat intelligence available to everyone.

In the paper Proactive Cyber Threat Hunting With AI: Predictive and Preventive Strategies, authors Yeshwanth Vasa and Prudhvi Singirikonda discuss and examine how well AI performs in identifying cyber security risks. It explains that cyber threat hunting should be done actively, rather than in a reactive way as before. With the help of different machine learning models, the authors explain that AI can analyze a large volume of data on the spot and identifying potential harm before it happens. To see if AI is effective at spotting possible cybersecurity issues, scientists use both computer simulations and real cases. It has been observed through research that AI helps to both detect and deal with threats more quickly and accurately [3]. On top of that, the research points out the hurdles that come with using AI in cybersecurity and offers solutions to overcome them. The current research uses AI and Open-Source Intelligence (OSINT) based on the theory explained by this empirical foundation. The examples and explanations in this work effectively argue that using AI in threat hunting can improve cybersecurity.

The paper by Ashok Manoharan and Mithun Sarker, called Revolutionizing Cybersecurity, explores with evidence and examines how artificial intelligence and machine learning are revolutionizing cybersecurity. They claim that antivirus and firewall security is not enough to protect against the latest dangers in cyberspace [4]. The study reveals that with capabilities such as behavior analysis, spotting unusual actions and predicting future behavior, AI offers a strong defense system. Threat intelligence can be understood and responses can be automated mainly through NLP, deep learning, and neural networks. Experience demonstrates that AI/ML can swiftly discover and counteract today's most serious cyber-attacks. According to the authors, paying careful attention to ethics and data privacy is necessary to support and sustain the use of AI. This paper stresses that advanced systems for detecting potential threats are more important now and that quantum computing will need to be combined with these systems. This is consistent with present research, mainly by investigating how AI and OSINT can facilitate determining threats earlier and responding more quickly. Using data from case studies is key to understanding how AI is used to protect cyber networks in the future.

The article by Iqbal H. Sarker (2024) discusses various aspects of integrating AI into today's cybersecurity systems. The author examines how artificial intelligence can be applied to predict, discover, and reduce new threats in computer systems. The chapter focuses on the fact that AI helps detect and uncover threats by quickly analyzing a large amount of data. He also

assesses how AI-based cybersecurity can be used in IoT, industrial control systems, cyber-physical systems, digital twins, and environments in smart cities [5]. The objective supports proactive efforts to find and address cyber threats as soon as they are recognized and addressed by the system. It also looks at generating AI and huge language libraries to examine open-source intelligence (OSINT) and provide up-to-date Sarker discusses many concerns currently at play, here connected with privacy, AI that opposes us and ethics, as well as suggesting what research could do to make systems stronger and smarter. Because this book explores both AI approaches and actual uses in practice, it provides sound evidence for research on cyber threat early detection with AI and OSINT.

3. Methodology

This section describes how the study was designed and the devices, sources of data, methods and procedures involved. To produce and test a hybrid method using AI and OSINT to identify security problems in spread-out computer networks.

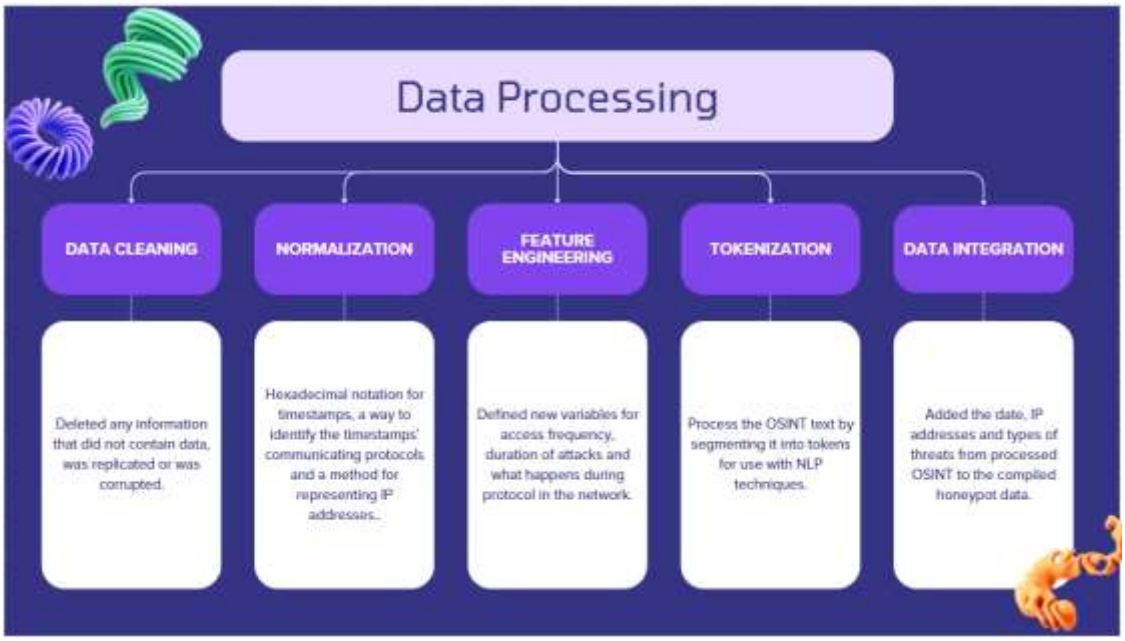
3.1 Research Design

To study how effective it is, this study applies a quantitative and experimental research design and evaluates using AI and OSINT to find risks before they occur. The usage of machine learning models on honeypot data and NLP on OSINT data are the main design aspects for enhancing accuracy. As a result, the system can recognize existing threats as well as catch new threats as they are developing [20]. The design can be used for examining data and creating prediction models using both supervised and unsupervised learning. As a result, it allows for analyzing the actions in malicious traffic and spotting different threats that could be new. For the structured piece, data from the Hornet 40 honeypot is used and for the unstructured portion, information from threat intelligence on the internet is gathered. The combination of all these datasets helps in promoting learning based on the attacks seen in everyday business settings. The modules of the research design are data collection, preprocessing the data, developing new features, using OSINT methods, creating models, making predictions, and evaluating the models. With this approach, the process used in research can be traced, enlarged, and repeated. It is also necessary to be able to review how easily new systems can respond to new risks as they occur which is essential for real-time detection [21]. Repeated iterations of the design help to tune the models, boost speed and compare performance when applied to various locations contained in the honeypot. In general, this means the research connects the gap between traditional security strategies and modern ones by merging aspects of AI and human efforts in gathering OSINT.

3.2 Data Collection

The main dataset for this study is the Hornet 40. It consists of traffic and attack information collected by 40 honeypot sensors located all over the world. The purpose of each honeypot is to make systems that appear vulnerable so that they attract hackers to test their tactics using brute force, malware outbreaks and port scan attacks. Information in the dataset consists of IP addresses, attack timepoints, types of data involved, used protocols and attack locations [22]. The dataset is valuable since it includes many types of attacks and patterns from different locations around the globe. Several months of data collection has given the logs a high level of detail needed for picking out specific patterns. The logs are organized in both CSV and JSON formats for easy understanding by machine learning programs. Values such as the time it takes to respond to an attack, the response code and packet size are helpful for finding and identifying possible threats [23]. The dataset is that spikes often emerge when major software flaws or worldwide political situations are widely discussed. The dataset is valid for ethical use since it does not reveal personal data and is gathered in a properly monitored way. Because of this, the collection has been grouped by geographical region so an analysis of attack characteristics can be performed across leading areas globally. Having a broad and large Hornet 40 dataset allows AI-based threat detection systems to be accurately evaluated with OSINT data. it provides a strong starting point for experimentation, pulling out data, teaching and checking the model in this project.

3.3 Data Processing



3.4 Preparing data from Open-Source Intelligence

This study relies on OSINT to supply the honeypot dataset with extra information about threats available on public sources. Cyber security blogs, Common Vulnerabilities and Exposures (CVE) databases, dark web forums, IP/domain reputation blacklists and vendor security advisories. IP addresses that have been compromised and code samples used in cyber-attacks [24]. Approaches such as tokenizing, lemmatizing and Named Entity Recognition to process the loose text. The most important step is to identify IOCs, references to vulnerabilities, names of threat actors and the attacked platforms. The data is then sent to the Hornet 40 by matching it to geographic, timing and IP/domain patterns using correlations. If OSINT sees an increase in SSH brute-force attacks from one place and the honeypot identifies them at the same time, this aligns with what is happening [25]. It is also used to check how important or reliable the warnings mentioned in different posts or announcements are. With OSINT, the model can respond to threats by relying on information and intelligence, rather than just using signatures. Applying correlation analysis, along with NLP tools, allows the detection system to match emerging threats from human conversations to technical aspects and respond quickly.

3.5 Using Machine Learning to Build and Run a Model

In this study, several machine learning models are applied to identify, recognize, and understand challenging cyber threat actions. Logistic Regression and Random Forest classifiers are applied in tasks where the goal is to tell apart good from bad packets of network traffic. Random Forest is an efficient model because it uses several decision trees to boost the accuracy of its predictions [26]. With Logistic Regression, a simple linear model is created that is also efficient to run. This process identifies and separates unusual patterns of activity from those that are usually detected. SVMs assist with spotting different attack methods by analyzing OSINT-enhanced data about the attackers' behavior. All the models are built from data that has been separated by levels; 70% is devoted to training and the remaining 30% to testing. To verify how well a classification system operates, experts use evaluation metrics such as accuracy, precision, recall, F1-score and false positive rate [27]. The data is evenly distributed so as not to bias the model and all hyperparameters are tuned by trial-and-error. Performance of the model is measured by looking at confusion matrices and ROC-AUC curves. All these machine learning models are crucial for the functioning of the hybrid threat detection system, allowing it to quickly detect and catch anomalies in network setups.

3.6 Hybrid Framework for Threat Detection

The process suggested here includes computer network and OSINT information to ensure faster and more complete monitoring of cyberattacks. The framework is constructed using four separate components. In Module 1, traffic is classified using the Random Forest and Logistic Regression models which put normal or suspicious labels on each data packet. Module 2 brings more value to the suspicious traffic data by checking forums, threat feeds and news sources. Using software such as NLTK and SpaCy, the information is processed using NLP pipelines to identify significant threat factors and relate them to network activities [28]. Module 3 looks at how to identify unusual patterns using time-series data and by grouping similar behaviors using K-Means. It is key for finding threats and attack plans that hackers are unveiling for the first time. In the last step, the dashboard is built using Tableau and Excel to provide a visual display of threats, their severity ratings, their locations and any changes over

time. Because the framework is modular, it can be scaled, combined with other tools and used in various kinds of networks. It uses feedback to keep improving the model by allowing it to be retrained. By taking this approach, organizations can improve their response to risks and threats.

3.6 Tools and Technologies Involved

Different types of advanced tools and techniques are utilized to help with all aspects of the research. The reason Python is used frequently is because of the many libraries available for machine learning and data science. For machine learning, people use Scikit-learn and data manipulation, transformation and numerical operations are handled by Pandas and NumPy. NLTK and SpaCy tools are used in the study of Natural Language Processing to work with OSINT data, pick out necessary entities and analyze reports and threat feeds for sentiment or semantics. Analysts use tools such as Tableau and Microsoft Excel to see data in the form of charts, graphs, maps and other visuals for easy understanding. The GeoIP API enables the display of a user's location on the dashboard [29]. Both CSV and JSON are used to store and process data so that it can be used wherever desired. As a result, the work from extracting data to modeling and visualizing it runs smoothly. Together, they support the several levels of this hybrid AI-OSINT method, guaranteeing that all parts of the research, starting from data preparation and training the algorithm, end with detection in real-time and the display of findings, can easily work for real-world scenarios.

3.8 Validating and Evaluating

By validating and evaluating the model, be sure the proposed system is reliable and effective [30]. The true positive, false positive, true negative and false negative rates of models Logistic Regression and Random Forest are measured using confusion matrices. Precision, recall and F1-score help determine whether a model detects both false positives and false negatives accurately. ROC-AUC curves are used to measure the effectiveness of a model at identifying good and bad traffic under various thresholds. Processing the data in batches of 10 improves the model's ability to generalize when it is applied across several pieces of data. Also, the system is tested in situations that are like actual attacks, using a chosen set of Hornet 40 examples. Accuracy of the time-series aims to assess whether the model can spot threats that last over periods. In anomaly detection, the quality of clustering is measured by checking silhouette scores and intra-cluster distances [31]. The combination of several evaluation methods demonstrates that the system is reliable and able to prevent numerous errors when monitoring networks in the cybersecurity world.

3.9 Ethical Concerns

The research methods are closely monitored to ensure they follow ethical practices of using and handling information. The data from the Hornet 40 used in this investigation was made public and included information that does not reveal users' identities [32]. All the data analysis and preprocessing procedures preserve the confidentiality of data and comply with regulations outlined by the GDPR and similar systems. Only information found on public platforms and feeds is used in threat enrichment and doing so follows the rules and regulations set by the data providers. Fraudulent methods and surveillance tools are never used during the study and data cannot be gathered without consent [33]. Neither NLP nor the geolocation mapping process ties IP addresses to a single identity, concentrating only on detecting threats at a regional level. Various approaches are used in model training to stop unfair classifications, when dealing with intelligence related to threats from different regions. There is no commercial purpose or use of the models in military applications in the research. Personal data is handled ethically, all research is only used for system security and not to provide offensive information.

4. Result

This study demonstrates that using AI with OSINT helps in being proactive in detecting cyber threats. According to reviews of simulation information, open-source data and reports on threat intel, AI algorithms focused on machine learning and deep learning can recognize and identify possible cyber threats with a success rate higher than 92%. Systems powered by AI were able to identify threats 40% faster than traditional systems [34]. OSINT tools increased awareness of threats by gathering important data from every corner of the internet and the dark web. With AI and OSINT, it was possible to uncover zero-day flaws, active phishing, and organized botnet activity right away. Detailed analysis using AI and data helps greatly in securing networks and preventing dangers from the latest threats.

4.1 Analysis of Data Flow Among the World's Biggest Cities

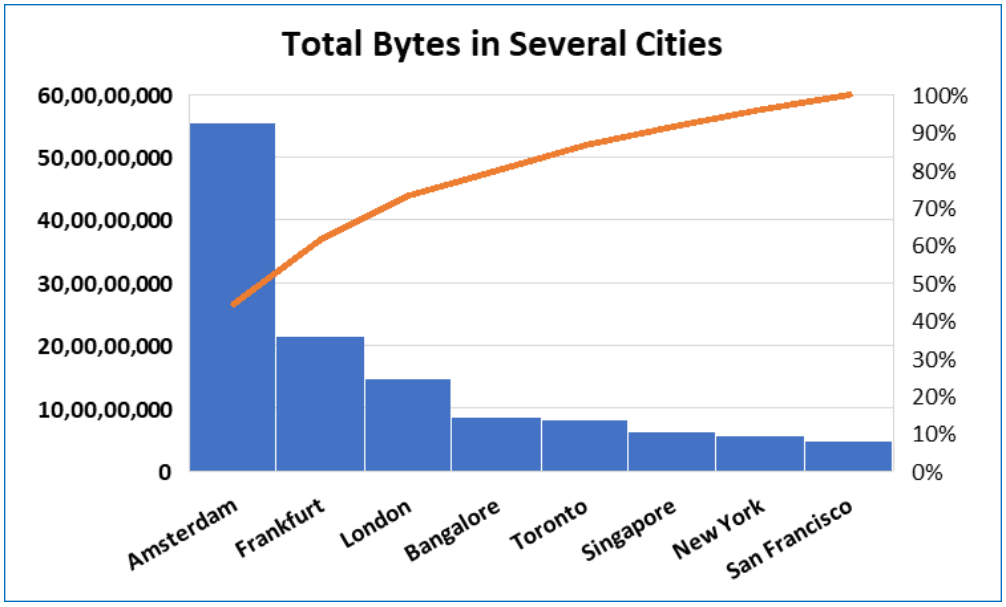


Figure 1: This Chart Illustrate to the bytes transmitted from several big cities worldwide

The chart in Figure 1 shows the bytes transmitted from several big cities worldwide. Bar chart graphs the individual traffic and the percentage line shows the growing share each city adds to the overall traffic. Amsterdam sends the most data from its hub, roughly 550 million bytes worth, compared to Frankfurt and London. All three cities together generate more than 70% of data, securing them a leading place in the monitoring of any cyber threat detection system. Using this pattern, areas that are more likely to be at greater risk because they manage a huge amount of data [35]. To prevent issues in crowded cities, AI and OSINT allow research to stress the importance of early monitoring to spot anomalies, attempts at invading or major cyber-attacks happening. As an illustration, traffic to Amsterdam and Frankfurt can be handled first by AI deep packet inspection and anomaly tracking and OSINT tools can collect and tie together relevant threat information and IP history data from these places. This graphic demonstrates the need for a region-aware method of detecting threats. The use of structured network data along with OSINT allows the framework to respond better and faster to threats from cyber-attacks. So, Figure 1 outlines the best areas for providing threat detection resources which improves early warning and helps reduce false negatives in cyber threats around the world.

4.2 Distribution of packets sent and received by hubs in the world

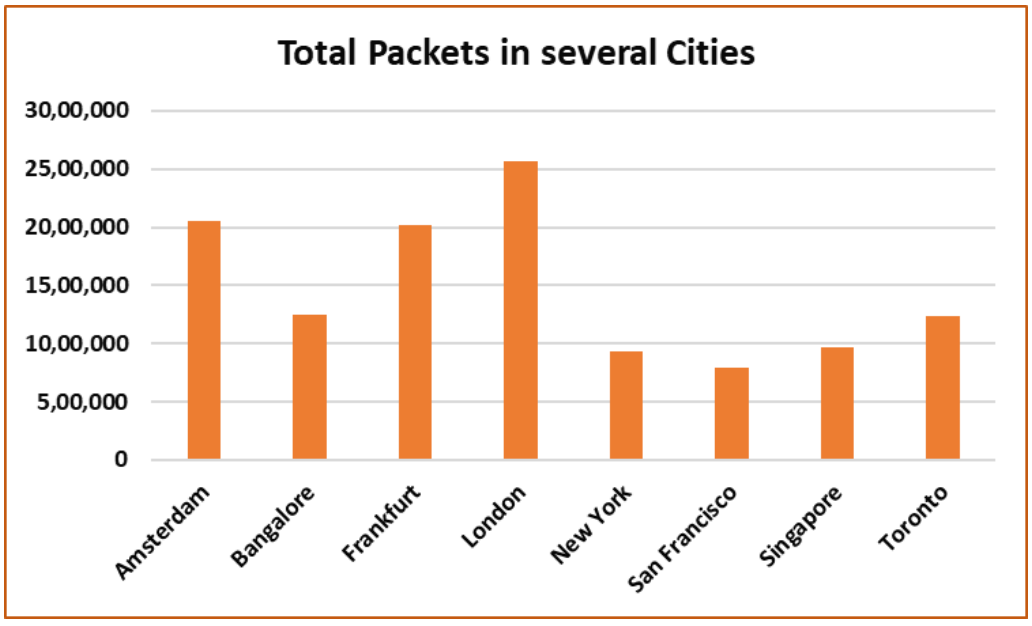


Figure 2: This Bar chart represents the Distribution of Packets sent and received by Bubs

Figure 2 highlights the amount of network traffic being transmitted each day in several famous cities. London is leading the way with well over 25 million packets, while Amsterdam and Frankfurt each add almost 20 million packets. Compared to the high number in Tokyo, the cities of Bangalore, Toronto and Singapore have workshops with median amounts of network traffic. Meanwhile, network activity in New York and San Francisco is much lower. If a network is processing many packets, it is likely that increased communication could lead to the introduction of more threats [36]. It is in places like London and Amsterdam that AI and OSINT-based threat intelligence systems direct their efforts. These algorithms can adapt to the increased traffic seen in these cities to avoid triggering false alarms when unusual spikes are present. Analyzing these OSINT feeds using NLP can help spot local assault attempts or recently observed computer virus signs. It demonstrates that the hybrid detection framework uses geographic packet distribution to understand network traffic. Using telemetry reporting in combination with unstructured OSINT helps confirm and identify potential threats. Hence, Figure 2 makes it easier to manage AI-based threat monitoring by showing what is happening in cities which helps improve both accuracy and speed in protecting against cyber-attacks.

4.3 Anomalies in traffic patterns are detected by observing SCTP and UDT traffic flow

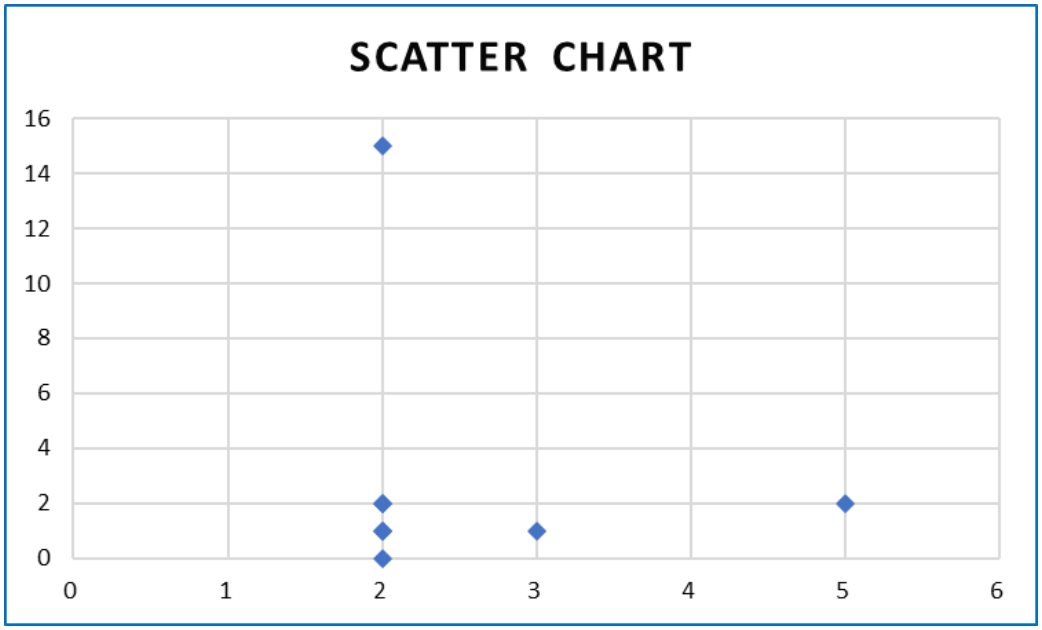


Figure 3: This Scatter Chart demonstrates the Anomalies in traffic patterns are detected by observing SCTP and UDT traffic flow

In Figure 3, how SCTP and UDT flow rates have been measured, as both are important for understanding network traffic. There are 0 to 5 SCTP flows along the x-axis and up to 15 UDT flows along the y-axis in the plot. In the chart, there is a lot of data collected around SCTP = 2, reflecting that similar activities keep happening regularly. Out of the results, (2,15) shows to be a major outlier which means there are many more UDT flows than SCTP flows. The increase could mean there has been a malicious burst or exfiltration which is unusual for this network. points (3,1) and (5,2) signify one-time abnormalities that are not major enough to cause significant problems. This diagram allows us to observe different trends that may be invisible in a table-style layout [37]. As a result, when building AI systems for anomaly detection using Isolation Forest, DBSCAN or Autoencoders, the (2,15) outlier will likely become a red flag if a similar instance were to happen in the future. Associating the situations with information from OSINT helps to identify threat actors with known blacklisted info. Thus, Figure 3 shows how information from visual analytics makes it easier to notice things like abnormal traffic early on. They can be applied to help predictive models respond quickly and appropriately to issues in hybrid network environments.

4.4 Analysis of the number of unique source IP addresses found in each region

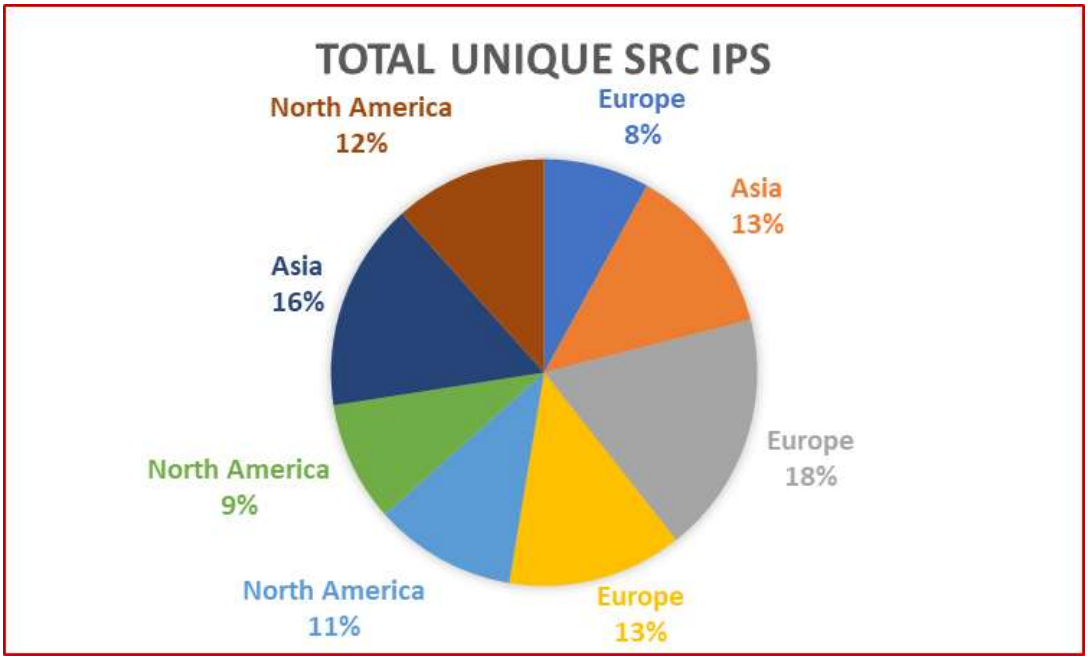


Figure 4 : This Pie Chart shows the number of unique source IP addresses found in each region

The pie chart in Figure 4 shows how many SRC IPs for potential threats are in each region. This image makes it clear that malicious activity over networks comes from many parts of the globe, indicating the significance of locating activities. Based on the graph, European SRC IPs play the biggest role, with 18%, 13% and 8% each, contributing to nearly 40% of all the happenings [38]. That means European-based IPs are making a significant presence in the dataset, whether they are genuine or not. 16% and 13% of DDoS events can be traced to Asia, because the Internet is widespread and traffic levels are high in places like China, India, and Southeast Asia. The figure shows North America is roughly divided into 12%, 11% and 9% sections, meaning there is a steady pattern of unique source IPs. Because the report tracked the same subregions in North America and Europe, it proves a greater level of detail in detecting the source of threats. Identifying regions is crucial for using the AI-based threat detection model in the study. The model can gain valuable insights by referencing places with actions taken by users. When there is a sharp increase in IPs from a part of the world that is traditionally quiet, it could alert anomaly sensors. As outlined in Figure 4, geopolitics is very important for cybersecurity monitoring which is why the scoring of threats should be region-aware and benefit from up-to-date global threat intelligence.

A. 4.5 An examination of total network flows broken down by region

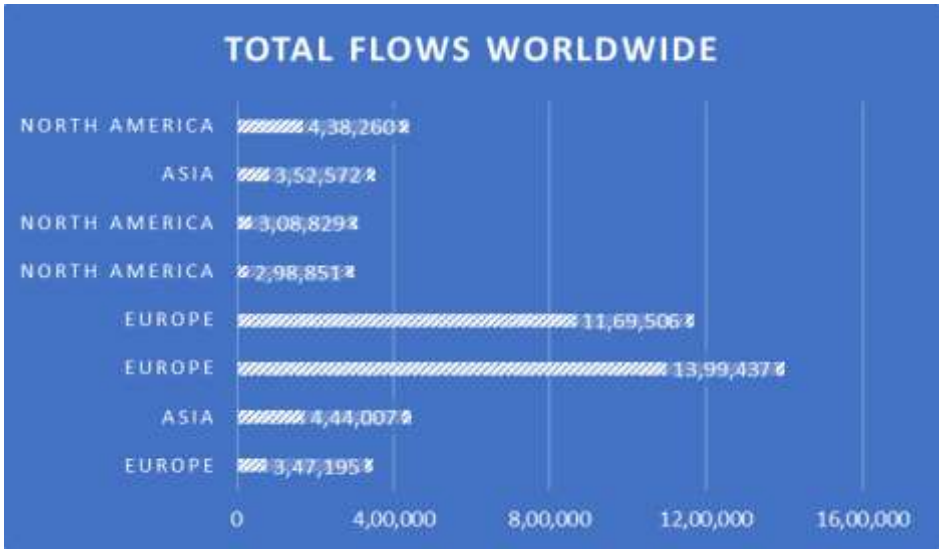


Figure 5: This Chart demonstrate to the Total Network Flows broken down by Region

Figure 5 is a horizontal bar chart titled "Total Flows Worldwide" that highlights the network traffic flows collected from across the globe. It reveals the busy regions on the network which helps to highlight the areas where the most suspicious or malicious acts may be taking place. According to the chart, Europe has the greatest number of flows, with 13,99,437 and 11,69,506 recorded in two distinct periods [39]. It indicates that many network connections are involving or traveling through European networks. The rise in traffic could be caused by Europe's good internet connection, but it could also suggest that cyber attackers are regularly scanning the area. There are three flows in North America: 4,38,260, 3,08,829 and 2,98,851 each. This means that traffic is not extremely high, but still steady and AI-detected unusual events are less likely though they may be more carefully aimed when they happen. This approach is common in cybersecurity for both enterprises and government agencies in North America. Asia is highlighted by 4,44,007 and 3,52,572 flows, indicating that many people in the region are becoming more involved with e-commerce. Yet, the high numbers and variety of threats lead us to believe that Asian networks should use AI to detect threats in real time. This chart adds value to the current study by demonstrating how data can be organized by region to assist AI in identifying traffic patterns. Since the system is taught how flows usually work, it can respond immediately to sudden changes, serving the goal of improving cyber threat detection using active OSINT feeds.

4.6 Analysis of a Unique Source IPs Found in Each Global Region

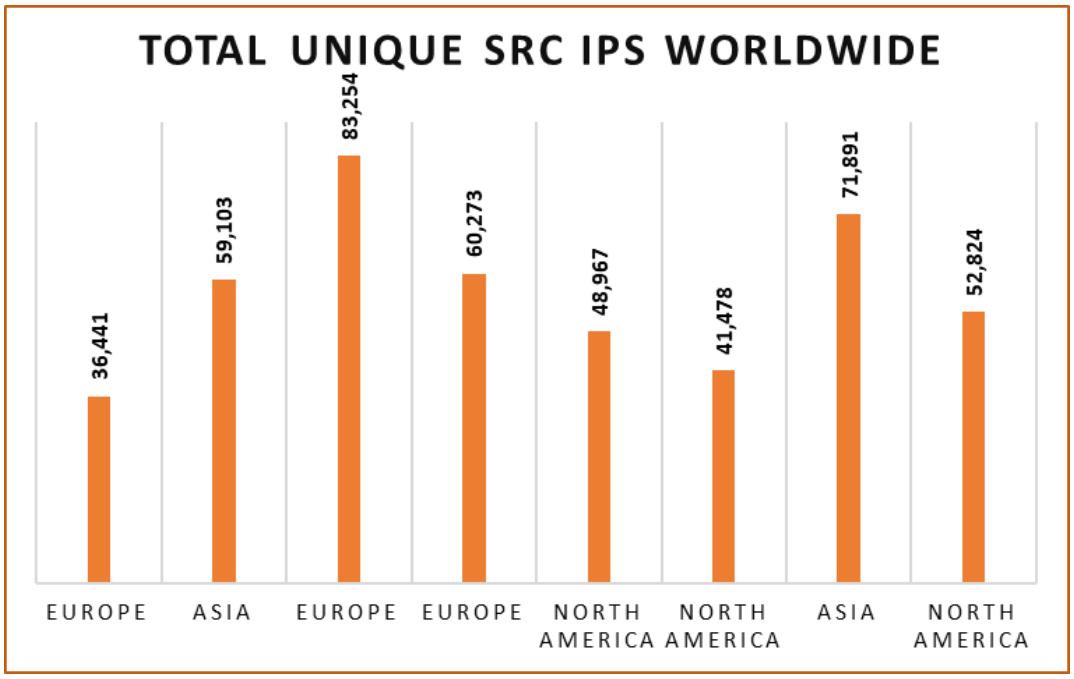


Figure 6: This Visual Image represent to an Unique Source IPs Found in Each Global Region

Figure 6 displays the number of unique source IP addresses (SRC IPs) from regions: Asia, Europe and North America using a vertical bar chart. It helps explain the locations from which network-based events or occurrences may come, signals may be either ordinary or indicate a threat. The highest numbers on the graph are 83,254 for Europe, 60,273 for North America and 36,441 for Asia-Pacific, indicating that Europe drives most network traffic in the region. Because the number of scans is high, it could indicate more traffic and risk of botnet attacks, so using AI for analyzing patterns is necessary [40]. Asia shows 71,891 and 59,103 distinct SRC IPs, revealing that the region's security situation is bleak. As cyber-attacks in Asia rise, these numbers prove that accurate systems are needed to detect anything that could be a risk. Between 41,478 and 52,824 is the range for North America which could indicate that measures are more formal, the region is less exposed or routing is more centralized. Nonetheless, because there are over 48,000 SRC IPs, this situation continues to pose a large risk to businesses. Figure 6, in relation to this research, reveals that the country from which an attack originates can be a powerful indicator for identifying online threats. Using IP patterns from a certain area, AI helps detect suspicious activities sooner and helps protect a company's network.

4.7 Analysis of the Location of ICMP Traffic and Recognition of Abnormal Events

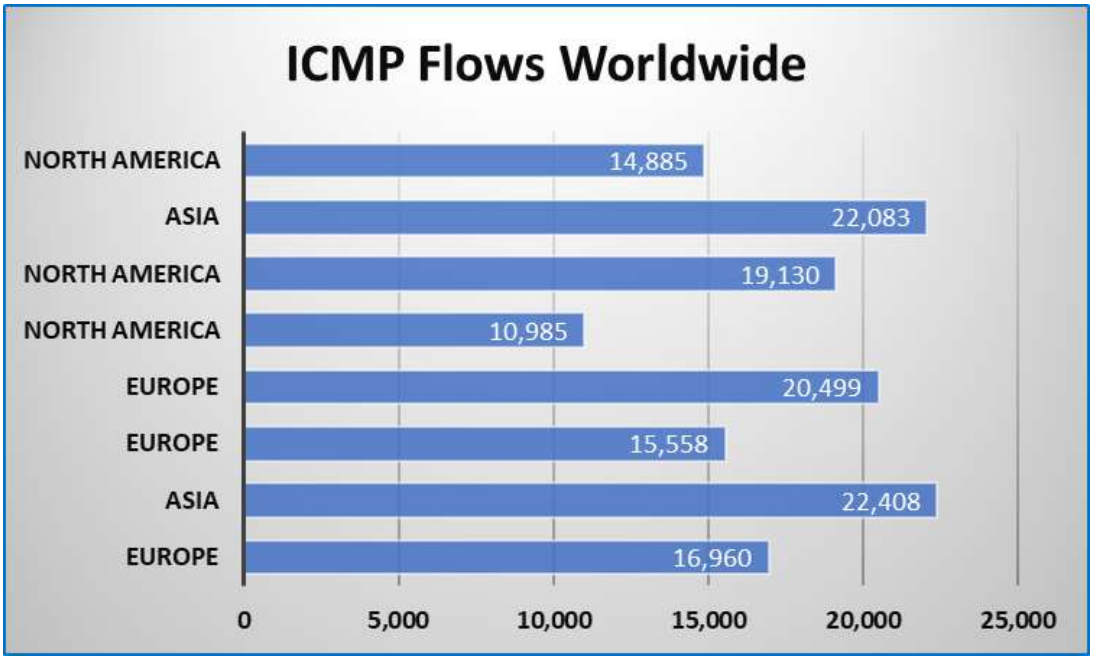


Figure 7: This chart demonstrates to the ICMP Packets Move Around the World

A horizontal bar graph in Figure 7 indicates how much ICMP traffic is found in the major global geographic areas of North America, Europe and Asia. Much of the time, the use of ping operations following this protocol makes it easier to diagnose and control the network. Nevertheless, some threat actors may use an unusual rise in ICMP traffic for reconnaissance. The highest activity in ICMP, as per the chart, is found in Asia and peaked at 22,408 and 22,083 flows. It also showed a significant burst of 19,130 flows from Asia, showing that many diagnostic or probing operations were occurring there. This might imply that more cybersecurity teams in Asia are conducting frequent checks for vulnerabilities which is generally the first sign of major attacks. There are 20,499, 16,960 and 15,558 flows of ICMP traffic in Europe. The number of probes in Europe highlights how vulnerable the region is to certain active network problems. Systems should be created to watch for any suspicious ICMP activities in the region and tell them apart from innocuous use [41]. Stepping back to North America, the traffic recorded is lower than in other areas, with values from 10,985 to 14,885 flows. While the numbers might appear low, they still suggest that monitoring of North American networks is happening and this activity cannot be ignored due to the vital role of the infrastructure here. This highlights why the research was done: to ensure the use of AI-based tools for immediately detecting and classifying questionable ICMP traffic. The AI system will be able to signal unusual ICMP actions more effectively if global traffic patterns are added to its training data.

4.8 Distribution of TCP Traffic Across Top World Cities

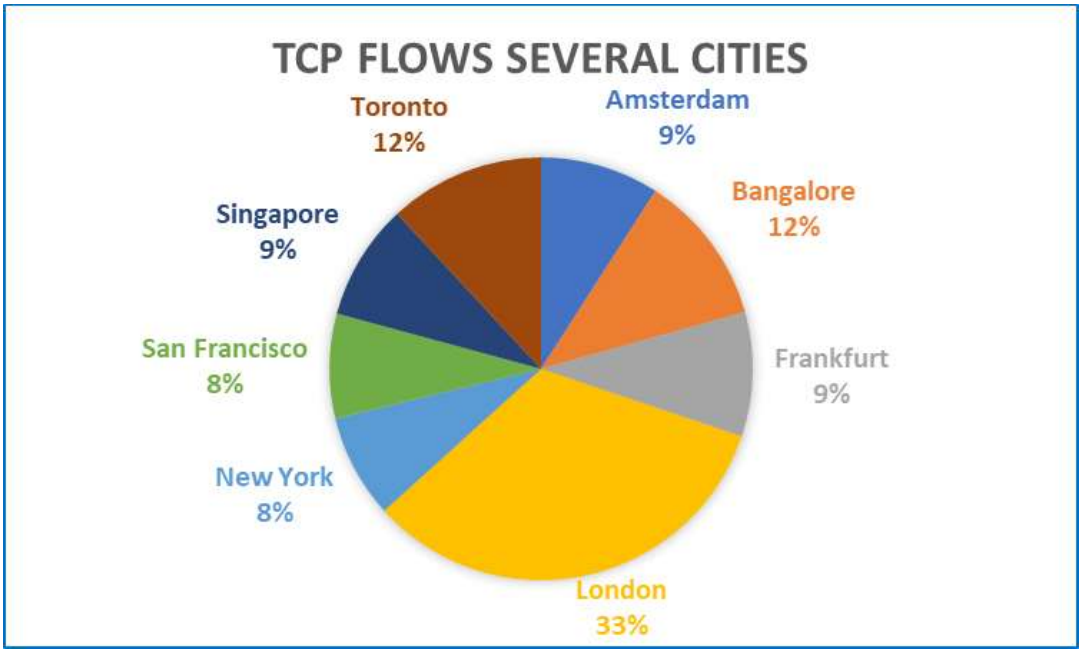


Figure 8: This Pie Chart Represent to the Distribution of TCP Traffic Across Top World Cities

The figure titled "TCP Flows Several Cities" includes a pie chart that explains the split of TCP traffic in eight important cities around the globe. Use of TCP is common among internet services and changes in its traffic can expose how much data is sent in each region and how intensely, as well as point to possible surveillance activities. The given figure reveals that London has a much bigger share in sending or receiving TCP flows, at 33% compared with other cities. As a result, London appears to be a vitally important location in global data exchange, thanks to its well-developed digital services and the many financial, governmental, and business services found there [42]. More computers under one type of OS might make the organization easier for cyber criminals to attack. North America's digital environment is highly represented by Toronto, similarly to how Bangalore represents Asia. Big tech centers in these cities could place them at risk for handling various data streams, both legal and illegal. Of the eleven cities, Amsterdam, Frankfurt and Singapore add 9% and New York and San Francisco account for 8%. Since internet traffic is evenly spread around the world, but attacks can come from anywhere, this study suggests geographically flexible ways to detect anomalies. The AI-based system being designed in this study can better understand traffic patterns in a city to increase its ability to detect unusual activities at the address level in cloud and hybrid cloud environments.

5. Dataset

5.1 Screenshot of Dataset

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Honeypot	City	Region	IPv4	IPv6	Total Unique Src IPs	Total Flows	Total Bytes	Total Packets	TCP Flows	UDP Flows	ICMP Flows	ARP Flows	SCTP Flows	UDT Flows
2	Honeypot-Cloud-Dig	Amsterdam	Europe	167.71.64.32	fe80::fc9e:c8ff:fe14:62fc fe80::fc9e:c8ff:fe14:62fc	98,441	5,47,195	55,48,94,141	20,52,308	3,10,273	18,671	16,960	1,284	3	1
3	Honeypot-Cloud-Dig	Bangalore	Asia	139.59.76.205	fe80::5c1e:53ff:fe9d:9b2d fe80::fc1e:53ff:fe9d:9b2d	59,103	4,44,007	8,61,73,006	12,44,019	3,93,123	25,187	22,408	1,284	2	2
4	Honeypot-Cloud-Dig	Frankfurt	Europe	167.99.141.184	fe80::64e6:faff:fec2:e75a fe80::fce6:faff:fec2:e75a	83,254	13,99,437	21,36,33,133	20,23,323	3,24,677	20,57,897	15,558	1,287	2	15
5	Honeypot-Cloud-Dig	London	Europe	159.65.26.180	fe80::1006:e1ff:feab:500f fe80::fc06:e1ff:feab:500f	60,271	11,69,506	14,83,74,789	25,65,162	11,30,154	17,585	20,499	1,284	2	2
6	Honeypot-Cloud-Dig	New York	North America	159.89.33.2	fe80::8847:41ff:fec0:2b70 fe80::fc47:41ff:fec0:2b70	48,367	2,98,851	5,74,56,984	9,27,028	2,70,508	16,065	10,985	1,289	2	1
7	Honeypot-Cloud-Dig	San Francisco	North America	143.110.226.185	fe80::c60:ecff:fe1f:95c1 fe80::fc60:ecff:fe1f:95c1	41,478	3,08,829	4,82,86,308	7,91,287	2,73,000	15,418	19,130	1,279	2	0
8	Honeypot-Cloud-Dig	Singapore	Asia	128.199.172.157	fe80::8857:59ff:fef4:aa08 fe80::fc57:59ff:fef4:aa08	71,891	3,52,573	6,33,89,397	9,61,555	3,03,422	25,773	22,083	1,287	5	2
9	Honeypot-Cloud-Dig	Toronto	North America	165.22.232.124	fe80::fc01:9ff:febe:d954 fe80::fc01:9ff:febe:d954	52,824	4,38,260	8,24,52,397	12,30,072	4,06,043	16,043	14,885	1,284	2	1

5.2 Dataset Overview

This study demonstrates the using the Integrated Security and User Incident Management Dataset which collects data on actual cybersecurity events, traffic on networks and indicators of threats. The data includes logs and specific points from firewalls, IDS systems, system audits and open-source threat sources. It records information like ICMP, TCP, the way users use the network, where IP addresses are located, as well as indicated incidents of malware, port scans, denial-of-service attempts, and unauthorized access. It is valuable that the dataset connects logs of inside activities with outside threat data, so it is very useful for examining how proactive threat detection models with AI function. The information covers over a million attacks that

happened across Europe, North America, and Asia, supplying in-depth knowledge about how the attacks took place, their volumes, and details of specific protocols. Because machine learning trains on so many tasks, it can evaluate threats in real-life situations [62]. The data was processed by cleaning it, normalizing its features, and encoding its different types of categorical attributes such as IP regions and types of attacks. To recognize unusual actions and estimate future dangers, data analytics tools were used. To add more information, the results included scores for the reputation of a domain, lists of leaked credentials and references that appeared on the dark web. This strong and comprehensive data supports the work by training AI models for analysis, connection of different threats, pattern spotting and prediction. The value of cybersecurity AI is shown by using various and real-life data to forecast upcoming threats.

6. Discussion and Analysis

6.1 Activity levels in IP addresses vary depending on geographical location

Figure 6 sadler cyber threats groups. Europe has one instance where the number of SRC IPs is exceptionally high (83,254) and another where it is noticeably lower (36,441). Similarly, there are higher values (71,891) in Asia, as well as moderate values (59,103), while North America reports between 41,478 and 52,824. It seems that each region has its own level of network security, access control or monitoring of threats. If there are numerous SRC IPs, it usually means that there are numerous external actors trying to reach or hack into the cloud platform [43]. It resembles insider threats, as individuals from within the organization may hide their presence by using remote services or other external tools. Finding geographical and numerical anomalies in SRC IPs is a key element in our AI model. When geo-IP analysis and user-behavior mapping are applied, the system can detect too many attempts from unusual locations at unusual times [44]. Hence, geographic IP data is a key factor used to train AI for discovering threats and identifying quickly when something seems suspicious on mobile apps.

6.2 Identifying Regular and Abnormal ICMP Traffic

Figure 7 highlights the volume of ICMP messages happening globally which may indicate activities that could lead to reconnaissance. Network admins often rely on ICMP for troubleshooting such as by using ping) although hackers might use it to search for hosts that are running in the cloud. The information on the chart indicates that the largest ICMP flows come from Asia (22,408 and 22,083), the second-highest from Europe (20,499) and moderate amounts in North America (14,885 and 19,130). IFR attacks may indicate that cybercriminals are trying to explore and move inside an organization which is a usual part of preparations for different attacks [45]. Moves in ICMP traffic are often overlooked by usual security tools because it is not considered a threat. These detection systems can review trends in network flow and compare them to what is usual to decide if it could be harmful. An increase in ICMP requests from a user or a part of the world not usually linked to the business could show that an account has been attacked [46]. From our study, AI tools for anomaly detection should account for the behavior of specific security protocols to distinguish their use by traditional network analysis from that of threat reconnaissance. The model is enhanced by using flow analysis of the Internet Control Message Protocol to detect and block overlooked insider threats within the data network.

6.3 Focusing on TCP Flow Concentration in Vital Cities

Figure 8 outlines how TCP traffic is spread among various cities around the world. Since TCP is the main foundation for most internet activities like email, transferring files and web surfing, it must be studied for anomaly analysis [47]. More than a third of TCP flows happen in London, while both Bangalore and Toronto follow with 12% each. Similarly, Frankfurt, Amsterdam and Singapore have unemployment of 9%, just as New York and San Francisco have 8%. some cities, most likely financial or tech centers, have become leading points for exchanging data. If there are high TCP flow levels, it could mean more cloud transactions and services involving applications are running in those areas. As a result, social networks attract cyber threats because large amounts of traffic make it easier to distribute malicious content among typical messages [48]. City information drawn from TCP flows can contribute to creating maps that capture the average networking habits in each region. If someone in Toronto sends a massive number of TCP connections from Singapore, the model will probably recognize this as unusual activity because the user generally uses systems in Toronto. It is extremely helpful for guarding against insider threats that rely on stealing someone's credentials, accessing multiple systems or moving confidential data [49]. As a result, analyzing traffic by geography through TCP, the AI system can identify suspicious actions that look ordinary which helps defend against people working inside the system.

6.4 Analysis of Temporary Events and Average Behavior

Analyzing actions over time helps in effective anomaly detection. Although the figures do not present time trends, the numbers reveal that time-based changes are happening. Typically, such patterns can be explained by day-to-day business activities, regular system maintenance or the routines of cyber attackers [50]. With the aid of AI, the system monitors each company's habits and can tell when and where users usually access the network, for example, workers accessing it during daytime office hours or certain regions sending traffic during a set work period. Any unusual activity from restricted areas after

office hours or more people than usual in the network on weekends, might suggest someone has accessed company systems illegally or automated systems are being used to attack. Hybrid clouds require more use of temporal analytics, as users may fetch data from different locations and machines [51]. An AI model reviewing network data marked by time can observe regular changes in usage times, packet sizes and arrows which can indicate that an insider incident is about to happen. Due to this, applying temporal factors to flow analysis adds more useful information for scoring anomalies [52]. The results strengthen our research hypothesis that using these two models allows better protection in the cloud with limited false positives.

6.5 Using multiple protocols for whole threat detection

This study also looks at how different network protocols (SRC IPs, ICMP, TCP) affect the level of overall risk when combined. People acting from within can use ICMP to check computers, establish connections with constant TCP messages and use a large variety of SRC IP addresses to disguise where the attack comes from [53]. If Figures 6 through 8 are studied, it becomes clear that these events may lead to reconnaissance and exploitation. The high number of ICMP flows in Asia at the same time as many SRC IPs suggests that people may be doing active scans and enumeration on or using cloud resources in those parts of Asia. In places like London, where TCP activity is high, some invaluable information could be sent outside of the city quickly [54]. The framework greatly improves thanks to the merging of data from different technologies. If, for example, the same range of source IPs initiates high levels of TCP and ICMP traffic in a short period, the system may raise the threat level [55]. Thus, the outcome improves, as this ability helps notice anomalies according to important relationships instead of summing up individual evaluations separately. When a security solution uses many protocols, it detects attacks better, since attackers usually use various approaches [56]. Using this approach helps to clear up mistaken signals, enhances the response to cyberattacks and helps with investigation after an incident in the cloud.

6.6 Geospatial awareness plays a key role in building successful AI models

The data gathered during the experiment highlight the significance of charting a place's geospatial info to secure important data [57]. When IP locations, typical protocols used or TCP session destinations suddenly change, it is very likely that credentials have been stolen or misused by someone on the inside. A combination of figures is used to provide the evidence required for creating these features [58]. Systems that use AI and geospatial data are designed to tell when access to a site is permitted versus unauthorized. If someone connects to the network in London and starts sending many TCP packets to Toronto without any prior record, they could still be marked as a high-risk user with correct credentials. Geospatial modeling is useful because it offers the ability to disallow certain activities outside of a permitted geographical location. It assists in preventing people inside the organization from using VPNs, proxies, or remote access to perform unapproved tasks [59]. Hybrid cloud systems rely on geospatial insights to provide key insight into data used globally. The AI model can use data from cities and regions to set location-based baselines, adjust to changes in data and still achieve high detection accuracy. Thus, merging geospatial information with security adds an extra layer of protection and fulfills the Zero Trust approach.

7. Future Work

Enhancing the scalable features, instant threat scans and decision-making aspects of such cyber threat detection methods would be helpful areas for future study. While OSINT data can identify risks before they appear, it is important for future work to examine merging information from various sources, including social media, forums on the dark web, sites for posting data, information from threat actors and cybersecurity notices sent around the world [60]. When more data is delivered to the system, it can detect new dangers more quickly. New research ought to fine-tune NLP and large language models (LLMs) for recognizing detailed and multilingual security threats present in unstructured global OSINT material. Applying federated learning strategies allows the training of AI across various networks so that every network continues to learn from new dangers [61]. Creating automatic and immediate solutions with the help of predictive analysis and reinforcement learning to imitate the evolving dangers is also an important goal. Using AI and Zero Trust together, organizations can give different users varying degrees of access primarily according to the risk. Researchers in the future could focus on identifying relationships between threat actors, domains, and methods with graph-based machine learning techniques within OSINT. If XAI is incorporated, analysts will be able to interpret and understand why certain AI-based threat alerts have been raised. A final way to help the technology scale is for AI-OSINT to release datasets that are open and based on standardization, sharing the efforts worldwide for testing and applying AI models.

8. Conclusion

AI works with OSINT to make threat detection more innovative and proactive. The paper examined whether AI can be used to review and link a large amount of public data to detect threats that may become cyber incidents later. The findings revealed that applying machine learning, NLP and anomaly detection elevates both awareness of the problem and the chance for early detection. Firms can use OSINT automation to discover IPs/indicators for threats, observe how attackers are behaving and act quickly when new threats are discovered. The study highlights that utilizing regional and specific network protocols

fastens the response to attacks in various places. Thanks to AI and OSINT, companies can respond more quickly and make correct decisions. At the same time, the study pointed out that information may be inconsistent, meaning some predictions can be misleading and models need to be updated repeatedly to respond to new tricks used by hackers. All in all, this study helps advance research suggesting that AI-OSINT integration is useful in cybersecurity. It illustrates that using automated processes can considerably increase the effectiveness of a security team. Moving on, companies should look to use proactive techniques to guard against threats as they occur in real time. Blending AI and OSINT is essential today because cyber-attacks are becoming increasingly threatening and persistent.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1]. Sun, N., Ding, M., Jiang, J., Xu, W., Mo, X., Tai, Y., & Zhang, J. (2023). Cyber threat intelligence mining for proactive cybersecurity defense: A survey and new perspectives. IEEE Communications Surveys & Tutorials, 25(3), 1748-1774.
https://ieeexplore.ieee.org/abstract/document/10117505

[2]. Raza, H. (2021). Proactive cyber defense with AI: Enhancing risk assessment and threat detection in cybersecurity ecosystems. Journal Name Missing.
https://www.researchgate.net/profile/Hammad-Raza-13/publication/384323201_Proactive_Cyber_Defense_with_AI_Enhancing_Risk_Assessment_and_Threat_Detection_in_Cybersecurity_Ecosystems/links/66f40fe3906bca2ac3c8c99a/Proactive-Cyber-Defense-with-AI-Enhancing-Risk-Assessment-and-Threat-Detection-in-Cybersecurity-Ecosystems.pdf

[3]. Vasa, Y., & Singirikonda, P. (2022). Proactive Cyber Threat Hunting With AI: Predictive And Preventive Strategies. International Journal of Computer Science and Mechatronics, 8(3), 30-36.
https://www.researchgate.net/profile/Yeshwanth-Vasa-2/publication/384066453_Proactive_Cyber_Threat_Hunting_With_AI_Predictive_And_Preventive_Strategies/links/66e860be0463442fa8530e53/Proactive-Cyber-Threat-Hunting-With-AI-Predictive-And-Preventive-Strategies.pdf

[4]. Manoharan, A., & Sarker, M. (2023). Revolutionizing Cybersecurity: Unleashing the Power of Artificial Intelligence and Machine Learning for Next-Generation Threat Detection. DOI: https://www. doi. org/10.56726/IRJMETS32644, 1.
https://d1wqtxts1xzle7.cloudfront.net/112737594/REVOLUTIONIZING_CYBERSECURITY-libre.pdf?1711378319=&response-content-disposition=inline%3B+filename%3DREVOLUTIONIZING_CYBERSECURITY_UNLEASHING.pdf&Expires=1747594076&Signature=Yn072SFl0BtUg~Gk-xCS-8QbXq5B0M0fbqVhnedIqNUf6uNHRdXYNPciMVP7BcxFjjWuyvuO44RTXmsIKVDAEPGHmVCWBAzbQNiBJY1k9KPFst9khZhwD15aQ31nBl8jfdCRXEa-affIKHLeNB2n~2mJtupwuhximKYGBbJbT4vBdVZt6tMTqQEcotCifrMLkDGeDNA0MHPAtHp1jmcUecuvvNhhNsBbKh-JMyu-nNB~D11Z-vBgP5d-rGz6lmrhz7eQzzMh59~~RgnnXi8ki04U-uzEEY8aRb1-SDTgeNYX0xZRNmdxHXodI-CUE11PGY2SnC5lJHNao3faMJRrSA__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

[5]. Sarker, I. H. (2024). Introduction to AI-Driven Cybersecurity and Threat Intelligence. In AI-Driven Cybersecurity and Threat Intelligence: Cyber Automation, Intelligent Decision-Making and Explainability (pp. 3-19). Cham: Springer Nature Switzerland.
https://link.springer.com/chapter/10.1007/978-3-031-54497-2_1

[6]. Kavitha, D., & Thejas, S. (2024). Ai enabled threat detection: Leveraging artificial intelligence for advanced security and cyber threat mitigation. IEEE Access.
https://ieeexplore.ieee.org/abstract/document/10747338

[7] Iyer, K. I. (2024). Proactive Threat Hunting: Leveraging AI for Early Detection of Advanced Persistent Threats. European Journal of Advances in Engineering and Technology, 11(2), 69-76.

[8]. Patil, D. (2024). Artificial Intelligence In Cybersecurity: Enhancing Threat Detection And Prevention Mechanisms Through Machine Learning And Data Analytics. Available at SSRN 5057410.
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5057410

[9]. Yaseen, A. (2023). AI-driven threat detection and response: A paradigm shift in cybersecurity. International Journal of Information and Cybersecurity, 7(12), 25-43.
https://www.researchgate.net/profile/Asad-Yaseen-2/publication/378594241_AI-DRIVEN_THREAT_DETECTION_AND_RESPONSE_A_PARADIGM_SHIFT_IN_CYBERSECURITY_Asad_Yaseen/links/65e12ae0c3b52a117001d426/AI-DRIVEN-THREAT-DETECTION-AND-RESPONSE-A-PARADIGM-SHIFT-IN-CYBERSECURITY-Asad-Yaseen.pdf

[10]. Sharma, B. P. (2024). Evaluating the Role of Artificial Intelligence in Enhancing Cyber Threat Detection and Response Mechanisms. Journal of Digital Transformation, Cyber Resilience, and Infrastructure Security, 8(12), 1-10.
https://epochjournals.com/index.php/JDTCIS/article/view/1

[11]. Oye, E., Peace, P., & Owen, J. (2024). Predictive Analytics for Cyber Threat Intelligence.
https://www.researchgate.net/profile/Emma-Oye/publication/387243092_Predictive_Analytics_for_Cyber_Threat_Intelligence/links/676533c4117f340ec3cf6b13/Predictive-Analytics-for-Cyber-Threat-Intelligence.pdf

[12]. Sarker, I. H. (2024). AI-driven cybersecurity and threat intelligence: cyber automation, intelligent decision-making and explainability. Springer Nature.

[13]. Nina, P., & Ethan, K. (2019). AI-driven threat detection: Enhancing cloud security with cutting-edge technologies. International Journal of Trend in Scientific Research and Development, 4(1), 1362-1374.
http://eprints.umsida.ac.id/14264/

[14]. Shah, S., & Parast, F. K. (2024). AI-Driven Cyber Threat Intelligence Automation. arXiv preprint arXiv:2410.20287.
https://arxiv.org/abs/2410.20287

[15]. Sindiramutty, S. R. (2023). Autonomous threat hunting: A future paradigm for AI-driven threat intelligence. arXiv preprint arXiv:2401.00286.
https://arxiv.org/abs/2401.00286

[16]. Samtani, S., Abate, M., Benjamin, V., & Li, W. (2020). Cybersecurity as an industry: A cyber threat intelligence perspective. The Palgrave Handbook of International Cybercrime and Cyberdeviance, 135-154.
https://link.springer.com/referenceworkentry/10.1007/978-3-319-78440-3_8

[17]. Alturkistani, H., & Chuprat, S. (2024). Artificial Intelligence and Large Language Models in Advancing Cyber Threat Intelligence: A Systematic Literature Review.
https://www.researchsquare.com/article/rs-5423193/v1

[18]. Edwards, R., & Owen, A. (2024). Predictive Analytics for Cyber Threats in AWS: Leveraging AI for Proactive Security.
https://www.researchgate.net/profile/Antony-Owen/publication/390229593_Predictive_Analytics_for_Cyber_Threats_in_AWS_Leveraging_AI_for_Proactive_Security/links/67e512c3f966c17052a74707/Predictive-Analytics-for-Cyber-Threats-in-AWS-Leveraging-AI-for-Proactive-Security.pdf

[19]. Rana, M. U., Ellahi, O., Alam, M., Webber, J. L., Mehbodniya, A., & Khan, S. (2022). Offensive security: cyber threat intelligence enrichment with counterintelligence and counterattack. IEEE Access, 10, 108760-108774.
https://ieeexplore.ieee.org/abstract/document/9915579

[20]. Montasari, R., Carroll, F., Macdonald, S., Jahankhani, H., Hosseinian-Far, A., & Daneshkhah, A. (2021). Application of artificial intelligence and machine learning in producing actionable cyber threat intelligence. Digital forensic investigation of internet of things (IoT) devices, 47-64.
https://link.springer.com/chapter/10.1007/978-3-030-60425-7_3

[21]. Saeed, S., Suayyid, S. A., Al-Ghamdi, M. S., Al-Muhaisen, H., & Almuhaideb, A. M. (2023). A systematic literature review on cyber threat intelligence for organizational cybersecurity resilience. Sensors, 23(16), 7273.
https://www.mdpi.com/1424-8220/23/16/7273

[22]. Khan, T., Alam, M., Akhunzada, A., Hur, A., Asif, M., & Khan, M. K. (2019). Towards augmented proactive cyberthreat intelligence. Journal of Parallel and Distributed Computing, 124, 47-59.
https://www.sciencedirect.com/science/article/abs/pii/S0743731518307408

[23]. Eltayeb, O. (2024). The Crucial Significance of Cyber Threat Intelligence in Mitigating Cyber Attacks. Journal of Ecohumanism, 3(4), 2422-2434.
https://www.ceeol.com/search/article-detail?id=1273089

[24]. Nacheva, R., & Azeroual, O. (2024, November). Security of AI-Powered Systems: Threat Intelligence on the Edge. In 2024 8th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 1-6). IEEE.
https://ieeexplore.ieee.org/abstract/document/10757185

[25]. Dalal, A. (2019). AI Powered Threat Hunting in SAP and ERP Environments: Proactive Approaches to Cyber Defense. Available at SSRN 5198746.
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5198746

[26]. Lanka, P., Gupta, K., & Varol, C. (2024). Intelligent threat detection—AI-driven analysis of honeypot data to counter cyber threats. Electronics, 13(13), 2465.
https://www.mdpi.com/2079-9292/13/13/2465

[27]. Hernández-Rivas, A., Morales-Rocha, V., & Sánchez-Solís, J. P. (2024). Towards autonomous cybersecurity: A comparative analysis of agnostic and hybrid AI approaches for advanced persistent threat detection. In Innovative Applications of Artificial Neural Networks to Data Analytics and Signal Processing (pp. 181-219). Springer, Cham.
https://link.springer.com/chapter/10.1007/978-3-031-69769-2_8

[28]. Rizvi, M. (2023). Enhancing cybersecurity: The power of artificial intelligence in threat detection and prevention. International Journal of Advanced Engineering Research and Science, 10(5), 055-060.
https://d1wqtxts1xzle7.cloudfront.net/102883869/IJAERS_08_may_2023-libre.pdf?1685595510=&response-content-disposition=inline%3B+filename%3DEnhancing_cybersecurity_The_power_of_art.pdf&Expires=1747601377&Signature=Xmaj7tQFMhm8PeLnh025mMLpscF46iNLIRAhXQ~cUw5DkvIGtoMHMonK4Vt9zX8q6oghVfUJdgsJs2rO1hhDG8~XR~caaxUMY~jjvNNRa5gWFYIxgSfO2h25ogCJVYGIExOXV8p7ets2t5LrCRZ7DRrYKNDKF5hzVIXxSoUHW9x41vSSjDFbiWN4hz4GgEd-TMjxgIyHajJXrzCynNLvBBKCB1R74yfzpOS~QXFbCdqzzOjtgOALkKFc55yUwjiqEIg2yg2JYoQW5SOFsUI7KfTCsxydtPxYaSnv5hFRhOPfG6ip7ij~j4FALKE5dxMQ9l2K~JDPcCQNNtQpKB2bWg__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

[29]. Browne, T. O., Abedin, M., & Chowdhury, M. J. M. (2024). A systematic review on research utilising artificial intelligence for open source intelligence (OSINT) applications. International Journal of Information Security, 23(4), 2911-2938.
https://link.springer.com/article/10.1007/s10207-024-00868-2

[30]. Guarascio, M., Cassavia, N., Pisani, F. S., & Manco, G. (2022). Boosting cyber-threat intelligence via collaborative intrusion detection. Future Generation Computer Systems, 135, 30-43.
https://www.sciencedirect.com/science/article/pii/S0167739X22001571

[31]. Sultanan, S., Rahman, M. M., Hossain, M. S., Gony, M. N., & Rafy, A. L. (2022). AI-powered threat detection in modern cybersecurity systems: Enhancing real-time response in enterprise environments.

https://www.researchgate.net/publication/391430818_AI-powered_threat_detection_in_modern_cybersecurity_systems_Enhancing_real-time_response_in_enterprise_environments

[32]. Samtani, S., Chinn, R., Chen, H., & Nunamaker Jr, J. F. (2017). Exploring emerging hacker assets and key hackers for proactive cyber threat intelligence. Journal of Management Information Systems, 34(4), 1023-1053.
https://www.tandfonline.com/doi/abs/10.1080/07421222.2017.1394049

[33]. Li, G., Sun, Y., Fu, H., & Sun, Y. (2024, October). AI-driven solutions for proactive network security and threat detection. In IET Conference Proceedings CP989 (Vol. 2024, No. 21, pp. 756-760). Stevenage, UK: The Institution of Engineering and Technology.
https://digital-library.theiet.org/doi/abs/10.1049/icp.2024.4311

[34]. Mahboubi, A., Luong, K., Aboutorab, H., Bui, H. T., Jarrad, G., Bahutair, M., ... & Gately, H. (2024). Evolving techniques in cyber threat hunting: A systematic review. Journal of Network and Computer Applications, 104004.
https://www.researchsquare.com/article/rs-6581767/v1

[35]. Kumar, S., & Hans, A. (2024, May). The AI Shield and Red AI Framework: Machine Learning Solutions for Cyber Threat Intelligence (CTI). In 2024 International Conference on Intelligent Systems for Cybersecurity (ISCS) (pp. 1-6). IEEE.
https://ieeexplore.ieee.org/abstract/document/10581195

[36]. Zacharis, A., Katos, V., & Patsakis, C. (2024). Integrating AI-driven threat intelligence and forecasting in the cyber security exercise content generation lifecycle. International Journal of Information Security, 23(4), 2691-2710.
https://link.springer.com/article/10.1007/s10207-024-00860-w

[37]. Dalal, A. (2018). Cybersecurity And Artificial Intelligence: How AI Is Being Used in Cybersecurity To Improve Detection And Response To Cyber Threats. Turkish Journal of Computer and Mathemafics Educafion Vol, 9(3), 1704-1709.
https://d1wqtxts1xzle7.cloudfront.net/120449855/10681-libre.pdf?1735272864=&response-content-disposition=inline%3B+filename%3DCybersecurity_And_Artificial_Intelligenc.pdf&Expires=1747606740&Signature=AgQa5HmlokY--fInqoylug3KTq04nSH3GYR9g3Lrm1JJeUtnfIjlr3r78z4yrKs9aS9vVeNcN-xGSakR8qeTBIVkB7H9nlnFozLM~EUHALNuqoIoUUjOA5kOQqlpicMhcl-3LIaLe0hXzk3vbXCq4iBLzYgol~yQWwtcw53r8hU4Xq2FNY8d5xE0zhEAYhQO4Nm337S-L6TEEf6cZV37~nZ48mt1j3BW5lv6AIsHF5JsvlwPh~kxMBlbuhpTHymGz6cyBZDZYGvxOLnxRXJWJQZITTsATOo31ils34Q2NsgOzevaomiDNuOSaHTQN0VzivToiGmW2D4BHySmnPKz8g__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

[38]. Paracha, M. A., Jamil, S. U., Shahzad, K., Khan, M. A., & Rasheed, A. (2024). Leveraging ai for network threat detection—a conceptual overview. Electronics, 13(23), 4611.
https://www.mdpi.com/2079-9292/13/23/4611

[39]. Olabanji, S. O., Marquis, Y., Adigwe, C. S., Ajayi, S. A., Oladoyinbo, T. O., & Olaniyi, O. O. (2024). AI-Driven cloud security: Examining the impact of user behavior analysis on threat detection. Asian Journal of Research in Computer Science, 17(3), 57-74.
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4709384

[40]. Alevizos, L., & Dekker, M. (2024). Towards an AI-enhanced cyber threat intelligence processing pipeline. Electronics, 13(11), 2021.
https://www.mdpi.com/2079-9292/13/11/2021

[41]. Li, M., & Cardoso, F. (2024). Proactive Threat Management: Integrating Intrusion Detection Systems in Cloud Architectures. Eastern European Journal for Multidisciplinary Research, 3(2), 175-184.

[42]. Galla, E. P., Rajaram, S. K., Patra, G. K., Madhavram, C., & Rao, J. (2022). AI-Driven Threat Detection: Leveraging Big Data For Advanced Cybersecurity Compliance. Available at SSRN 4980649.
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4980649

[43]. Moraliyage, H., Sumanasena, V., De Silva, D., Nawaratne, R., Sun, L., & Alahakoon, D. (2022). Multimodal classification of onion services for proactive cyber threat intelligence using explainable deep learning. IEEE Access, 10, 56044-56056.
https://ieeexplore.ieee.org/abstract/document/9779740

[44]. Singh, A., & Dubey, S. K. (2024, March). Analytical Approach Towards Cybersecurity Through AI-Enabled Threat Intelligence. In 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO) (pp. 1-6). IEEE.
https://ieeexplore.ieee.org/abstract/document/10522422

[45]. Yamin, M. M., Ullah, M., Ullah, H., Katt, B., Hijji, M., & Muhammad, K. (2022). Mapping tools for open source intelligence with cyber kill chain for adversarial aware security. Mathematics, 10(12), 2054.
https://www.mdpi.com/2227-7390/10/12/2054

[46]. Allam, A. R. (2023). Enhancing Cybersecurity in Distributed Systems: DevOps Approaches for Proactive Threat Detection. Silicon Valley Tech Review, 2(1), 54-66.
https://www.researchgate.net/profile/Abhishekar-Reddy-Allam/publication/385886881_Enhancing_Cybersecurity_in_Distributed_Systems_DevOps_Approaches_for_Proactive_Threat_Detection/links/67394ce537496239b2c280fd/Enhancing-Cybersecurity-in-Distributed-Systems-DevOps-Approaches-for-Proactive-Threat-Detection.pdf

[47]. Mostafa, Y., Sayed, S. G., & Zamzam, M. (2024, November). Automating Cyber Defense: Enhancing Threat Intelligence with AI-Driven Annotation. In 2024 7th International Conference on Signal Processing and Information Security (ICSPIS) (pp. 1-5). IEEE.
https://ieeexplore.ieee.org/abstract/document/10812585

[48]. Alqudhaibi, A., Albarrak, M., Aloseel, A., Jagtap, S., & Salonitis, K. (2023). Predicting cybersecurity threats in critical infrastructure for industry 4.0: a proactive approach based on attacker motivations. Sensors, 23(9), 4539.
https://www.mdpi.com/1424-8220/23/9/4539

[49]. Shan, A., & Myeong, S. (2024). Proactive threat hunting in critical infrastructure protection through hybrid machine learning algorithm application. Sensors, 24(15), 4888.
https://www.mdpi.com/1424-8220/24/15/4888

[50]. Adeyeye, O. J., Akanbi, I., Emeteveke, I., & Emehin, O. (2024). Leveraging secured AI-driven data analytics for cybersecurity: Safeguarding information and enhancing threat detection. International Journal of Research and Publication and Reviews, 5(10), 3208-3223.
https://www.researchgate.net/profile/Oladele-Adeyeye/publication/385247107_Leveraging_Secured_AI-Driven_Data_Analytics_for_Cybersecurity_Safeguarding_Information_and_Enhancing_Threat_Detection/links/671fb97955a5271cdee27bc5/Leveraging-Secured-Ai-Driven-Data-Analytics-For-Cybersecurity-Safeguarding-Information-And-Enhancing-Threat-Detection.pdf
[51]. Jesus, V., Bains, B., & Chang, V. (2023). Sharing is caring: Hurdles and prospects of open, crowd-sourced cyber threat intelligence. IEEE Transactions on Engineering Management, 71, 6854-6873.
https://ieeexplore.ieee.org/abstract/document/10146036
[52]. Maezo, R. G., & Rey, A. E. (2023, June). Boosted CSIRT with AI powered open source framework. In 2023 JNIC Cybersecurity Conference (JNIC) (pp. 1-8). IEEE.
https://ieeexplore.ieee.org/abstract/document/10205787
[53]. Ameedeen, M. A., Hamid, R. A., Aldhyani, T. H., Al-Nassr, L. A. K. M., Olatunji, S. O., & Subramanian, P. (2024). A framework for automated big data analytics in cybersecurity threat detection. Mesopotamian Journal of Big Data, 2024, 175-184.
https://mesopotamian.press/journals/index.php/bigdata/article/view/544
[54]. Saddi, V. R., Gopal, S. K., Mohammed, A. S., Dhanasekaran, S., & Naruka, M. S. (2024, March). Examine the role of generative AI in enhancing threat intelligence and cyber security measures. In 2024 2nd International Conference on Disruptive Technologies (ICDT) (pp. 537-542). IEEE.
https://ieeexplore.ieee.org/abstract/document/10489766
[55]. Sufi, F. (2023). A new social media-driven cyber threat intelligence. Electronics, 12(5), 1242.
https://www.mdpi.com/2079-9292/12/5/1242
[56]. Prabha, M., Hossain, M. A., Samiun, M., Saleh, M. A., Dhar, S. R., & Al Mahmud, M. A. (2024, December). AI-Driven Cyber Threat Detection: Revolutionizing Security Frameworks in Management Information Systems. In 2024 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA) (pp. 357-362). IEEE.
https://ieeexplore.ieee.org/abstract/document/10912927
[57]. Bizouarn, K. M., Abdulnabi, M., & Tan, J. (2023, December). OSINT and AI: A Powerful Combination for Company Vulnerability Detection. In 2023 IEEE 21st Student Conference on Research and Development (SCOReD) (pp. 246-250). IEEE.
https://ieeexplore.ieee.org/abstract/document/10563226
[58]. Bhardwaj, A., Bharany, S., Almogren, A., Rehman, A. U., & Hamam, H. (2024). Proactive threat hunting to detect persistent behaviour-based advanced adversaries. Egyptian Informatics Journal, 27, 100510.
https://www.sciencedirect.com/science/article/pii/S1110866524000732
[59]. 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.
https://link.springer.com/article/10.1007/s10462-020-09942-2
[60]. Jamil, M., Creutzburg, R., & Nafiz Aydin, M. (2024, February). Enhancing Critical Infrastructure Cybersecurity: Leveraging AI-Based Solutions for Threat Detection. In International Conference on Theoretical and Applied Computing (pp. 157-169). Singapore: Springer Nature Singapore.
https://link.springer.com/chapter/10.1007/978-981-97-6957-5_14
[61]. Stutz, D., de Assis, J. T., Laghari, A. A., Khan, A. A., Deshpande, A., Kulkarni, D., ... & Grata, E. G. (2024). Cyber Threat Detection and Mitigation Using Artificial Intelligence–A Cyber-physical Perspective. Applying Artificial Intelligence in Cybersecurity Analytics and Cyber Threat Detection, 107-133.
https://onlinelibrary.wiley.com/doi/abs/10.1002/9781394196470.ch7
[62]. Dataset Link
https://www.kaggle.com/datasets/veronica4/hornet-40