International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 11-No.3 (2025) pp. 5639-5647
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

AI-Augmented Big Data Analytics for Real-Time Cyber Attack Detection and Proactive Threat Mitigation

Sharmin Sultana[1], Mukther Uddin[2], Md Asikur Rahman Chy[3], Syed Nazmul Hasan[4], Emran Hossain[5], Harleen Kaur[6] , Md Nazibullah Khan[7], Jobanpreet Kaur [8*]

[1*] School of Business, International American University, Los Angeles, CA 90010, USA
Email: sharminanis369@gmail.com- **ORCID:** 0009-0005-7213-4504

[2] School of Business, International American University, Los Angeles, CA 90010, USA
Email: muktheruddin1996@gmail.com - **ORCID**: 0009-0009-0995-8761

[3] School of Business, International American University, Los Angeles, CA 90010, USA
Email: mdasikurrahmanchy21@gmail.com - **ORCID**: 0009-0002-6550-7104

[4] College of Technology & Engineering, Westcliff University, CA 92614, USA
Email: s.hasan.104@westcliff.edu- **ORCID**: 0009-0008-0977-595X

[5] Department of Business Administration, Humphreys University, 6650 Inglewood Ave, Stockton, California, USA
Email: hu0112358@student.humphreys.edu - **ORCID**: 0009-0005-2080-780X

[6] College of Technology & Engineering, Westcliff University, CA 92614, USA
Email: H.kaur.4088@westcliff.edu- **ORCID**: 0009-0009-7062-0700

[7] School of Business, International American University, Los Angeles, CA 90010, USA
Email: khanroyal2014@gmail.com- **ORCID**: 0009-0008-4132-1413

[8] College of Technology & Engineering, Westcliff University, CA 92614, USA
*** Corresponding Author Email:** j.kaur.244@westcliff.edu- **ORCID**: 0009-0008-0083-8205

Article Info:

DOI: 10.22399/ijcesen.3564
Received : 22 May 2025
Accepted : 23 July 2025

Keywords

Big Data Analytics
Cybersecurity
Machine Learning
Threat Detection
Anomaly detection

Abstract:

Big data analytics, as used in defense, is the capacity to gather vast amounts of digital data for analysis, visualization, and decision-making that might aid in anticipating and preventing cyberattacks. When combined with security technologies, it improves it position in terms of cyber defense. They enable companies to identify behavioral patterns that point to network dangers. With its potent capabilities to tackle the increasing scope, variety, and complexity of cyberthreats, big data analytics has become a disruptive force in contemporary cybersecurity. Traditional data processing methods fall short in managing the massive volumes, varieties, and velocities (3Vs) characteristic of big data. This paper explores the foundational principles of big data analytics, including its core dimensions and key application areas such as healthcare, transportation, finance, education, and social media. The study further investigates the classification of cyberattacks malware, phishing, ransomware, and advanced persistent threats (APTs) and their evolving complexity due to AI-powered automation, IoT proliferation, and multi-vector intrusion techniques. It is highlighted how crucial big data is to supporting real-time threat detection, predictive modelling, and automated incident response. Techniques such as behavioral analysis, threat intelligence integration, and anomaly detection are examined for their effectiveness in identifying sophisticated attacks like polymorphic malware and zero-day exploits. Ultimately, this paper highlights how big data analytics enhances cybersecurity capabilities by delivering predictive, prescriptive, diagnostic, and cyber-specific insights that empower proactive threat mitigation and ensure digital resilience.

1. Introduction

The term "Big Data" describes data sets that are too large or complex to work with using conventional data set processing application software. Scale, velocity, and variety are the primary ways that big data differs from regular data. Volume, velocity, and variance are used to represent various forms of organized and unstructured data, respectively. Volume is the total amount of data generated. These days, big data is a popular issue for researching practically every sector, including cybersecurity. The main sources of this information are smart gadgets and social networking websites. Data is currently being generated. The startling number of malware infections in a single industry each year illustrates the extent of the danger to the world economy [1]. It is clear that the cybersecurity of online apps, corporate networks, and IT systems may not be sufficient given the rapid rise in cyberattacks. "Big data" describes datasets that are too large or complex to handle with standard data set processing application software. In terms of volume, velocity, and variation, big data is very different from traditional data [2]. Variation refers to the types of organized and unstructured data, Volume indicates the amount of data created, and Velocity indicates the rate at which the data is produced. Big data is becoming a hot issue for research across practically all disciplines, notably cybersecurity.

Cybersecurity is also an electronic information security or information technology security which is the process of preventing unauthorized access to private data and breaches into networks, computers, apps, and data. Increasing the advanced technologies used by Cyber criminals, traditional tools used for data security with huge volume of data becomes hard. Cybersecurity is a technology that prevents assaults on networks, systems, devices, data, and applications by controlling and processing [3,4]. Its main aim is to lower the cyber-attacks risks and guard against the unauthorized usage of technologies, systems and networks. Cybersecurity with strong strategies has protection layers to guard against cybercrimes, includes accessing the data, destroying or changing confidential or sensitive data [5]. Now a day's, organizations are using more proactive and adaptive techniques due to the rapidly and fast – changing nature of security risks. Traditional cybersecurity systems are a group of network and computer security tools, such as IDS, firewalls, and antivirus software.

The ability of AI to do complicated jobs more quickly is on par with or even better than that of humans. Numerous scholars have put out attack or threat intelligence models to gather data about threats from the websites and social media accounts of attackers [6, 7]. In order to extract and portray intelligence in a graph structure, they relied on AI techniques. Nonetheless, gathering intelligence about cyberattacks and conventional IT is the primary goal.

In the effort to improve cybersecurity, ML and big data analytics are becoming powerful technologies. Using the massive amounts of data generated in cloud environments, big data analytics assists companies in identifying complex patterns and anomalies that may indicate security risks [8][9]. Cloud security, where massive data volumes and quick reactions are critical, benefits greatly from ML, techniques like clustering, anomaly detection, and predictive modelling. For example, clustering algorithms can group similar patterns of network behavior, helping to detect anomalies that may signify security breaches [10][11]. Organizations may proactively reduce risks by using predictive analytics, which forecast possible security events using past data. Additionally, reinforcement learning is increasingly used to enable automated security responses, dynamically adapting defense mechanisms as the threat landscape evolves.

Organization of the Paper

The following paper is organized as: Section II and Section III provide the fundamentals of big data analytics and classification of cyberattacks, then Section IV provides the role of big data in cybersecurity, also Section V and Section VI give the evaluation of the relevant literature and a conclusion with recommendations for further research.

2. Fundamentals of Big Data Analytics

Information that is too complex to manage, query, and analyze with standard data storage systems, algorithms, and query techniques is referred to as big data [12]. Figure 1 illustrates how the 3V's characterize the "complexity" of big data:

Figure 1. *The 3Vs of big data Volume, variety, and velocity*

- **Volume:** The most evident feature of big data is its enormous volume, or vertical scalability, as measured by the amount of data produced and stored. The quantity of data created daily is expected to expand by 300 times by 2020 compared to 2005, with a current prediction of 2.5 quintillion bytes per day. Big data typically surpasses the capacity of conventional column-and-row relational databases, necessitating the use of novel storage systems.
- **Variety:** There are three common categories of data, which may be acquired from many domains: unstructured, semi-structured, and structured [13]. The data's horizontal scalability is increased by this variant. Unstructured data seems more random and is more challenging to sort and analyses than structured data, which is typically already marked and is easily mapped into fields that have already been created, such tables in a database or spreadsheet.
- **Velocity:** The velocity is a measure of how quickly data is created and processed to satisfy requirements. In essence, velocity quantifies the pace of data creation, storage, analysis, and display. Big data technologies aim to create and analyze data in real-time or almost real-time, whereas traditional data handling methods can only manage data using batch data snapshots.

The 3Vs, high volume, high variety, and high velocity, have been commonly employed to characterize big data.

A- Big Data Applications

Data sources for Big Data analytics are many application domains. Education, governance, healthcare/medical, retail, history, entertainment, social networking, banking, and transportation are just a few examples of these uses. Figure 2 illustrates these several application areas [14, 15]. The current evaluation does not have the capacity to describe every application area in depth.

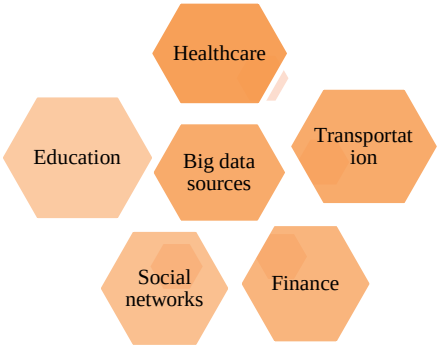

Figure 2. *Big data from different sources*

1. Big Data in Healthcare

The healthcare industry is one that uses the largest and fastest-growing databases found in big data analytics. These days, determining the size and growth rates of healthcare databases is quite challenging [16]. This category of clinical data includes, for instance, imaging data and electronic medical records (EMRs). A variety of data types are found in healthcare, including drug-response profiles, screening data, personal genetic data, and drug-related structures. Data pertaining to personal behaviors and preferences, such as food habits, environmental variables, financial records, and clinical and drug-related data, are also included. By combining all of these data types, big data analytics contribute to bettering both individual healthcare and the health of the world.

2. Big Data in Transportation Systems

One of the most significant uses of big data is in the intelligent transportation system (ITS), where inefficient traffic routing is primarily caused by the increasing amount of data on vehicle flow at crossings [17]. In order to address conventional processing issues, researchers are looking at more effective and efficient processing paradigms like Big Data analytics. This is because traditional data processing and analysis methodologies were unable to handle the massive data scale. In order to address conventional processing issues, researchers are looking at more effective and efficient processing paradigms like analytics of big data. This is a result of conventional data processing and analysis techniques' inability to manage the enormous volume of data.

3. Big Data in Finance

The shortcomings of processing capabilities, particularly floating-point processing, cause financial organizations to have excessive transaction processing delays. High-performance computing may be utilized to solve this issue in two ways: by adding computing nodes to a pool of computing systems or by adding additional CPUs or RAM to a single computer system [18]. However, the performance of storage systems also became a problem since financial data is growing at an exponential rate. Consequently, the frequent flow of massive data transactions has an impact on latency in addition to processing power.

4. Big Data in Social Networks

Large global user bases are connected via modern

social networks like Facebook, LinkedIn, and Twitter for both social and professional purposes. Because of the contact, exabytes of information are created every day. In order to improve their marketing and sales plans, information system analysts at businesses evaluate and forecast social network user behavior, which is regarded as a Big Data challenge [19].

5. Big Data in Education

Through helping educators understand the reasons behind poor results, such as why students fail or quit courses, fail to master ideas or abilities, and many more, data may be seen as a tool for institutions to improve their performance in order to enhance student outcomes [20]. It is believed that by recognizing notable variations in students' performance and conduct, educational institutions may stop any unfavorable evaluation results earlier.

3. Classification of Cyber Attacks

Cybersecurity is a vital topic within the broader area of information technology, which is committed to protecting computer networks, systems, and data against damage, infiltration, or assault, as shown in Figure 3. It includes a wide range of procedures, methods, and resources intended to preserve the accessibility, privacy, and accuracy of data [21][22]. The main objective of cybersecurity is to provide a safe online space where individuals and companies may conduct business without fear of data loss or cyberattacks. It is becoming more difficult to accomplish this aim because of the speed at which digital technology is developing and the skill of possible enemies [23]. Cyber threats may be classified into numerous main sorts, each having its own strategies, approaches, and aims.

- **Malware:** Trojan horses, worms, viruses, and spyware are examples of malicious software, or malware, which is designed to enter, harm, or take over a computer system without the user's consent [24]. The most common uses of malware are to monitor user behavior, steal private data, and disrupt system operation.
- **Phishing:** This kind of attack is an unethical attempt to get private data, including credit card numbers, usernames, and passwords, by impersonating a reliable business in online conversations [25][26]. Phishing attacks often involve emails asking victims to give personal information on a phony website that seems and feels almost exactly like the real one.
- **Ransomware:** A type of malware that renders the victim's data worthless by encrypting it and demands a payment to unlock it. Attacks utilizing ransomware have disrupted operations and caused significant financial losses for individuals, businesses, and even government organizations [27][28].
- **Advanced Persistent Threats (APTs):** These are highly advanced, high-tech attacks directed against significant targets, including big businesses and nation-states. APTs entail sustained involvement, in which hackers enter a network to steal data or keep tabs on activity covertly. Persistence is the primary trait of APTs; in order to keep access to the target, attackers constantly modify their tactics.

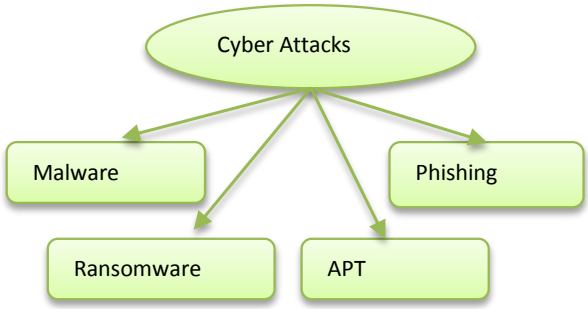

Figure 3. *Cyber Attacks Concept overview*

A- Impact of The Cyber Attack

Cyberattacks have become a pervasive and grave danger in the modern digital environment that impacts people, companies, and society as a whole [29]. The impacts can be classified into the following major areas:

- **Theft of Personal Information:** Hackers get sensitive information without authorization, including passwords, phone numbers, emails, bank records, complete names, addresses, dates of birth, and ID numbers[30].
- **Financial Loss:** Victims may lose money due to fraudulent transactions, ransomware payments, or identity theft exploitation.
- **Emotional and Psychological Impact:** After an assault, people may experience tension, anxiety, and a decline in trust in digital systems [31].
- **Business Revenue Loss:** Attacks like denial-of-service (DoS) prevent customer access, directly reducing online sales and service usage.
- **Operational Disruption:** Cyberattacks can halt business operations, damage systems, and delay services, affecting overall productivity.

B - Evolving Nature and Complexity of Attacks

Cyberattacks are becoming more sophisticated, automated, and stealthy. Attackers use AI-driven tools to bypass security systems, generate zero-day exploits, and mimic legitimate behavior, complicating detection and mitigation. Security defenses are further challenged by the rise of multi-vector assaults and living-off-the-land strategies, in which attackers make advantage of integrated system tools.
Moreover, the adoption of IoT, cloud computing, and 5G networks expands the attack surface, introducing new vulnerabilities and entry points. Cybercriminals now weaponize big data, social engineering, and deepfakes to target specific individuals or infrastructures, escalating the threat level.

C - Trends and Evolution of Attack Techniques

The cyber-attack landscape is continuously evolving, with attackers adopting more complex, intelligent, and covert methods, and the key trends are shown in Figure 4:

1. Polymorphic and Fileless Malware

Modern malware variants frequently change their code signatures (polymorphism) to evade detection by traditional antivirus systems. Fileless malware, on the other hand, operates in-memory using legitimate tools like PowerShell, leaving minimal forensic evidence. According to these techniques represent a significant challenge for signature-based and static detection models.

2. AI-Powered Attacks

AI-powered attacks represent a growing and alarming trend in the cyber threat landscape. Adversaries are using AI and ML more and more to automate, enhance, and expand their malevolent actions. One prominent use case involves the generation of highly convincing phishing emails using NLP models, which significantly increases the success rate of social engineering campaigns. Additionally, attackers utilize AI to scrape and analyze social media and publicly available data to identify and profile vulnerable targets, tailoring their attacks for maximum impact. Moreover, ML techniques are employed to bypass traditional security systems, particularly anomaly-based IDS, by imitating typical user behavior and avoiding discovery.

3. IoT-Based Attacks

The attack surface has greatly increased due to the IoT devices' explosive expansion. Attackers can use many IoT devices as entry points into bigger networks or as targets for botnet construction since they are frequently installed with default credentials or without firmware upgrades.

4. Supply Chain Attacks

These attacks aim to undermine a trusted system by focusing on less secure components in the hardware or software supply chain. Orion software upgrades were compromised by the 2020 SolarWinds hack, which had an impact on Fortune 500 businesses and government organizations. This type of attack is particularly dangerous due to the trust placed in software vendors.

5. Hybrid and Multi-Vector Attacks

Modern attacks often employ a combination of methods. For example, an APT might begin with a phishing email, followed by malware deployment and lateral movement across a network. These multi-stage attacks are harder to detect and require correlation across diverse data sources, making big data analytics especially valuable.

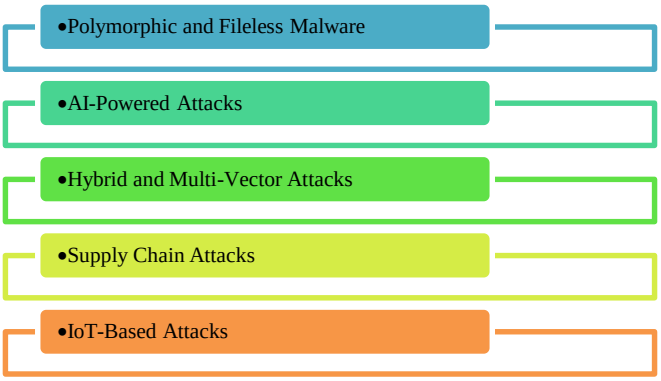

Figure 4. *Attack Techniques*

4. Role of Big Data in Cybersecurity

In cybersecurity, big data analytics offers several advantages over traditional techniques. To begin with, big data analytics enables the analysis of a variety of data types and quantities, such as network traffic, log files, user behavior, and external data sources like threat intelligence feeds and public databases. This makes it possible to see the whole cyber environment, including dangers from the inside as well as the outside [32][33]. Additionally, to find patterns and abnormalities in the data, big data analytics employ cutting-edge methods like machine learning and data mining. These methods can identify established hazards and unearth new ones by using models and

algorithms using historical and current data [34]. The ability to monitor and evaluate data streams in real time is another benefit of big data analytics, which facilitates the early detection of cyberthreats. Rapid threat response and mitigation are made possible by this real-time monitoring capacity, which reduces possible damage. Big Data Analytics and its applications for various kinds are depicted in Figure 5.

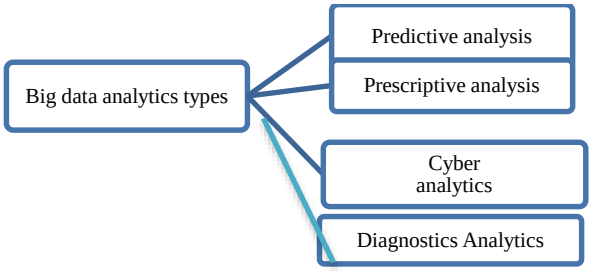

Figure 5. *Big data analytics for cyber security*

- **Predictive Analysis:** This focuses on forecasting future trends and behaviors by analyzing historical data. Common in fraud detection, risk assessment, and customer behavior prediction.
- **Prescriptive Analysis:** This provides actionable recommendations by combining data insights with optimization models. It helps in decision-making by suggesting the best course of action.
- **Cyber Analytics:** Tailored for cybersecurity, this type analyzes data to detect threats, intrusions, and vulnerabilities in real time or proactively.
- **Diagnostics Analytics:** This is used to determine the causes of past events or outcomes. It facilitates the discovery of connections, trends, and anomalies in historical data.

A- Applications of Big Data Analytics in Cybersecurity

The way security teams identify, look into, and the way that issues is evolving due to big data analytics. Its uses extend well beyond basic log analysis, whether it's thwarting insider threats or thwarting zero-day assaults [35]. Here's how it's significantly influencing [36].

1. Threat Intelligence and Predictive Analytics

Digital footprints left by cyberattacks include unsuccessful login attempts, dubious IP connections, odd data flows, and big data analytics that comb through billions of events in real time to identify new trends. Threat intelligence platforms (TIPs) gather information from event logs, malware reports, and dark web forums to help organizations anticipate possible attacks weeks in advance [37]. For instance, businesses may fix weaknesses before an attacker strikes by seeing a growing ransomware trend in financial institutions early.

2. Real-Time Fraud Detection

The multitrillion-dollar problem of financial fraud necessitates a flexible strategy. Big data analytics use machine learning to continually adjust to novel fraud strategies, in contrast to static, rule-based systems. For instance, payment processors are able to examine millions of transactions per second and identify anomalous activity, such 15 transactions from several nations in five minutes, as possibly fraudulent.

3. Advanced Malware and Ransomware Detection

Antivirus programs that rely on signatures frequently fail to detect zero-day malware. The emphasis is shifted to behavioral analysis via big data analytics. The system can recognize the warning indications of ransomware in operation and stop the process if a program begins encrypting thousands of files in quick succession. Additionally, by comparing current activity with previous attack patterns, these tools can identify and even catch polymorphic malware, which alters its code constantly to avoid detection.

4. Automated Incident Response and Threat Hunting

It might be too slow to respond to incidents manually. A deluge of signals, many of which are false positives, usually overwhelms security experts. By automating answers to high-confidence threats, big data analytics simplifies the procedure. For instance, the system may instantly isolate the device, remove its credentials, and notify the security operations center if an endpoint starts sending encrypted data to an unknown external IP.

5. Literature Review

The literature reviewed section highlights big data, artificial intelligence, and cybersecurity combined to improve threat detection, automate risk assessment, ensure data privacy, and address evolving cyber threats across diverse sectors.

Chityala et al. (2025) Traditional security solutions often struggle to keep up with the growing complexity and quantity of cyberthreats, necessitating a more advanced, data-driven approach. The suggested

approach makes use of anomaly detection, predictive modelling, and real-time data processing to increase system scalability, decrease reaction times, and improve threat detection accuracy. Technology finds possible threats more quickly than traditional techniques by gathering and evaluating information from several network sources, such as security device logs and endpoint protection software. Big data platforms like Apache Spark and ML methods like supervised and unsupervised learning allow for real-time analysis of massive amounts of data, guaranteeing prompt detection of known and undiscovered cyberthreats [38].

Helser and Hwang (2025), offer a thorough investigation of the relationship between cybersecurity, AI, and big data (CAB) across six manufacturing and public service industries. It draws attention to the revolutionary potential of these technologies in transforming sectors and increasing productivity, but it also underscores the difficulties they face, particularly in relation to privacy and data security. A three-dimensional security model (security aim, security control, and data state) is created and applied to six industries in order to contextualize these difficulties [39].

Alawadhi et al. (2024) highlights the critical importance of BDCA systems while exploring how Big Data may improve cybersecurity measures. The study specifically looks at the primary problems that the cybersecurity field faces and how Big Data concepts and techniques may be used to solve them. The paper evaluates the possible advantages and inherent challenges of using Big Data in cybersecurity, with an emphasis on the ethical and privacy implications. They also offer suggestions for the ethical and efficient use of big data in cybersecurity, highlighting the need to strike a balance between data security and innovation [40].

Sudan (2023) Data processing and storage is the main application of big data analytics. People, systems, and objects may all communicate with one another easily thanks to the IoT. This research paper looks at the existing frameworks for using big data analytics to build a safe IoT. Big data is information amassed from several sources, including sensors installed in everyday things. It may utilize this data for inference-based analysis and planning of the environment. By collecting and analyzing this information, the Internet of Things may automate the functioning of the many electrical devices present in the immediate vicinity. Yet as automation expands, so does exposure to cyberattacks [41].

Jilani et al. (2023) The use of phony calls, messages, and other methods by cybercriminals to get consumers' personal information is growing. These assaults are happening online more and more these days. Their previous research references have led this study to discover a variety of security technologies and tactics being used to increase the security of big data. Digital forensics technologies that can be used to determine who, where, and how an attack was carried out have also been included to this research. In this project, cybersecurity and data science are merged. This study's primary goals are to detect big data risks and protect and mitigate data security through the application of cybersecurity and digital forensics techniques [42].

Nisha et al. (2022) In order to predict, identify, characterize, and Cybersecurity analysts use enormous to manage security threats, a lot of events and security data is gathered from several sources. To find correlations that will enable them to make informed decisions and foresee problems before they occur, these researchers will need to use automated methods to understand and analyze these enormous datasets. Cyber protection may be impacted by big data analytics and AI. Massive data sets with a wide range of data types are subjected to big data analytics approaches. The objective is to find trends, correlations, patterns, and other pertinent data [43].

Table I summarizes the objective areas, main findings, challenges, limits, and future gaps of each study on big data in cyberattacks.

Table 1. *Summary Of the Study on Big Data Analytics in Cyber Attack Detection*

Author	Focus On	Key Findings	Difficulties	Limitations / Gap
Chityala et al. (2025)	Big data-driven cybersecurity in real time	combines predictive modelling, anomaly detection, and real-time data processing to identify threats; uses Apache Spark and ML for identifying known and unknown threats	Integration and timely processing of large, diverse data streams	Scalability in real-time threat identification and false positive reduction needs further enhancement
Helser and Hwang (2025)	Cybersecurity, AI, and big data in manufacturing & public services	Proposes a three-dimensional security model (goal, control, state); identifies transformational potential in six sectors	Balancing technological innovation with data privacy	Context-specific model adaptability and generalization across sectors need refinement
Alawadhi et al. (2024)	Big Data Cybersecurity Analytics (BDCA) systems	Highlights ethical, privacy, and operational challenges; provides responsible use recommendations	Ethical issues, data privacy, system integration	Lack of real-world implementation case studies and long-term impact analysis

Sudan (2023)	Big Data & IoT Security	Reviews frameworks using big data for secure IoT; emphasizes inference-based automation	Growing attack surface due to IoT expansion	Lack of robust standardization and secure-by-design architectures
Jilani et al. (2023)	Cyber threats and digital forensics in big data	Combines digital forensics with cybersecurity for threat detection; highlights increasing online frauds	Identification of attacker traceability in complex environments	Insufficient integration of forensic techniques in real-time systems
Nisha et al. (2022)	Big data analytics for cyber risk prediction	Emphasizes automation for analyzing massive security datasets; uses AI to forecast and alert	Need for advanced tools to handle heterogeneous data	Limited automation maturity; real-time alert systems still evolving

6. Conclusion And Future Work

Big Data analytics for security seeks to gather current, useful intelligence. Big Data may significantly impact your present organization in three different ways. Big Data Analytics' incorporation into cybersecurity frameworks signifies a paradigm change away from reactive defenses and towards proactive, clever threat mitigation techniques. Utilizing the 3Vs, Volume, Variety, and Velocity, Big Data Analytics allows businesses to handle and evaluate large, varied datasets instantly. This capability is essential in detecting, classifying, and responding to advanced cyber threats such as polymorphic malware, ransomware, and AI-powered attacks. Additionally, the use of automated incident response, behavioral analytics, and ML models has greatly increased detection accuracy and decreased response latency. Healthcare, finance, education, and transportation are just a few of the areas where big data analytics has significantly improved digital security and operational resilience.

Future studies should concentrate on creating analytics frameworks that protect privacy and guarantee data security while facilitating cooperative threat intelligence sharing amongst enterprises. Additionally, integrating explainable AI (XAI) into cybersecurity systems will enhance transparency and trust, especially in high-stakes decision-making. The increasing deployment of IoT and edge computing devices calls for lightweight big data solutions capable of performing real-time analytics in resource-constrained environments. Future work will also explore hybrid architectures that combine cloud, edge, and fog computing for scalable and efficient cybersecurity analytics. Lastly, advancing self-healing security systems through reinforcement learning and adaptive models represents a promising frontier for fully autonomous threat detection and mitigation.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Sachi, S., & Kumar, M. (2023, November). A review on big data analytics on cyber security. *Tuijin Jishu/Journal of Propulsion Technology, 44*(4), 4128–4133. https://doi.org/10.52783/tjjpt.v44.i4.1632

[2] Shukla, S. T. K. M., Patel, N., & Patel, V. (2024). AI based cyber security data analytic device. *(Makale veya konferans yayımlanma bilgileri eksik olduğundan tam referans verilemiyor.)*

[3] Prasanthi, K., Rani, P. K. S., & Krishna, P. P. V. (2024). BACADA – Big data architecture for cyber security attack detection applications. *Asian Finance, Banking and Business Studies, 6*(6), 7702–7714. https://doi.org/10.33472/AFJBS.6.6.2024.7702-7714

[4] Rao, D. D., Waoo, A. A., Singh, M. P., Pareek, P. K., Kamal, S., & Pandit, S. V. (2024). Strategizing IoT network layer security through advanced intrusion detection systems and AI-driven threat analysis. *Journal of Intelligent Systems and Internet of Things, 24*(2), 195–207. https://doi.org/10.54216/JISIoT.120215

[5] Arora, S., Khare, P., & Gupta, S. (2024, July). AI-driven DDoS mitigation at the edge: Leveraging machine learning for real-time threat detection and response. In *2024 International Conference on Data Science and Network Security (ICDSNS)* (pp. 1–7). IEEE.

https://doi.org/10.1109/ICDSNS62112.2024.10690930

[6] Majumder, R. Q. (2025, April). A review of anomaly identification in finance frauds using machine learning systems. *International Journal of Advanced Research in Science, Communication and Technology*, 101–110. https://doi.org/10.48175/IJARSCT-25619

[7] Neeli, S. S. S. (2022, November). Critical cybersecurity strategies for database protection against cyber attacks. *Journal of Artificial Intelligence, Machine Learning and Data Science, 1*(1), 2102–2106. https://doi.org/10.51219/JAIMLD/sethu-sesha-synam-neeli/461

[8] Abbas, M., & Zafer, A. (2024). Big data analytics for cybersecurity: Enhancing cloud infrastructure protection with machine learning techniques. *ResearchGate*. https://doi.org/10.13140/RG.2.2.20093.99040

[9] Duggasani, A. R. (2025, May). Scalable and optimized load balancing in cloud systems: Intelligent nature-inspired evolutionary approach. *International Journal of Innovative Science, Research and Technology, 10*(5), 2153–2160. https://doi.org/10.38124/ijisrt/25may1290

[10] Garg, S. (2019). Predictive analytics and auto remediation using artificial intelligence and machine learning in cloud computing operations. *International Journal of Innovative Research in Engineering and Multidisciplinary Physical Sciences, 7*(2), 1–5. https://doi.org/10.5281/zenodo.15362327

[11] Patel, N. (2024). Secure Access Service Edge (SASE): Evaluating the impact of converged network security architectures in cloud computing. *Journal of Emerging Technologies and Innovative Research, 11*(3), e703–e714.

[12] Mahmood, T., & Afzal, U. (2013, December). Security analytics: Big data analytics for cybersecurity: A review of trends, techniques and tools. In *2013 2nd National Conference on Information Assurance (NCIA)* (pp. 129–134). IEEE. https://doi.org/10.1109/NCIA.2013.6725337

[13] Yang, L., Li, J., Elisa, N., Prickett, T., & Chao, F. (2019, December). Towards big data governance in cybersecurity. *Data-Enabled Discovery and Applications, 3*(1), 1–12. https://doi.org/10.1007/s41688-019-0034-9

[14] Rassam, M. A., Maarof, M. A., & Zainal, A. (2017). Big data analytics adoption for cyber-security: A review of current solutions, requirements, challenges and trends. *Journal of Information Assurance and Security, 11*, 124–145.

[15] Duary, S., Choudhury, P., Mishra, S., Sharma, V., Rao, D. D., & Aderemi, A. P. (2024, February). Cybersecurity threats detection in intelligent networks using predictive analytics approaches. In *2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM)* (pp. 1–5). IEEE. https://doi.org/10.1109/ICIPTM59628.2024.10563348

[16] Shah, S. B. (2025). Machine learning for cyber threat detection and prevention in critical infrastructure. *Department of Operations, Business Analytics & Information Systems, 2*(2), 1–7. https://doi.org/10.5281/zenodo.14955016

[17] Polinati, A. K. (2025). AI-based big data management device. *UK Intellectual Property Office*.

[18] Prajapati, N. (2025). The role of machine learning in big data analytics: Tools, techniques, and applications. *ESP Journal of Engineering and Technology Advances, 5*(2), 16–22. https://doi.org/10.56472/25832646/JETA-V5I2P103

[19] Majumder, R. Q. (2025, April). Machine learning for predictive analytics: Trends and future directions. *International Journal of Innovative Science, Research and Technology, 10*(4), 3557–3564. https://doi.org/10.38124/ijisrt/25apr1899

[20] Malali, N. (2022). Using machine learning to optimize life insurance claim triage processes via anomaly detection in Databricks: Prioritizing high-risk claims for human review. *International Journal of Engineering Technology Research and Management, 6*(6). https://doi.org/10.5281/zenodo.15176507

[21] Happa, J., Glencross, M., & Steed, A. (2019, April). Cyber security threats and challenges in collaborative mixed-reality. *Frontiers in ICT, 6*. https://doi.org/10.3389/fict.2019.00005

[22] Polinati, A. K. (2025). AI-powered anomaly detection in cybersecurity: Leveraging deep learning for intrusion prevention. *International Journal of Communication Networks and Information Security, 17*(3), 301–323.

[23] Ezra, L. A. (2023). Big data analytics in cyber threat intelligence: A comprehensive literature survey on methodologies, challenges, and future directions. *International Journal of Computational Engineering Research and Trends, 10*(2). https://doi.org/10.22362/ijcert/2023/v10/i2/v10i205

[24] Aslan, O., & Samet, R. (2020). A comprehensive review on malware detection approaches. *IEEE Access, 8*, 6249–6271. https://doi.org/10.1109/ACCESS.2019.2963724

[25] Gupta, B. B., Tewari, A., Jain, A. K., & Agrawal, D. P. (2017, December). Fighting against phishing attacks: State of the art and future challenges. *Neural Computing and Applications, 28*(12), 3629–3654. https://doi.org/10.1007/s00521-016-2275-y

[26] Singamsetty, S. (2019). Fuzzy-optimized lightweight cyber-attack detection for secure edge-based IoT. *Journal of Critical Reviews, 6*(7), 1028–1033. https://doi.org/10.53555/jcr.v6:i7.13156

[27] Beaman, C., Barkworth, A., Akande, T. D., Hakak, S., & Khan, M. K. (2021, December). Ransomware: Recent advances, analysis, challenges and future research directions.

Computers & Security, 111. https://doi.org/10.1016/j.cose.2021.102490

[28] Prajapati, V. (2025). Cloud-based database management: Architecture, security, challenges and solutions. *Journal of Global Research in Electronics and Communication, 1*(1), 7–13.

[29] Fadli, Z., Yong, S., Kee, L., & Ching, G. (2022). Cyber attack awareness and prevention in network security. *International Journal of Informatics and Communication Technology, 11*, 105. https://doi.org/10.11591/ijict.v11i2.pp105-115

[30] Nurmuslimah, S., Saidatin, N., & Bagus, P. (2023). The future of HR cybersecurity: AI-enabled anomaly detection in Workday security. *International Journal of Recent Technology and Scientific Management, 8*(6).

[31] Sola, R. P., Malali, N., & Madugula, P. (2025). Cloud database security: Integrating deep learning and machine learning for threat detection and prevention. Notion Press.

[32] Singh, S., & Kumar, D. (2024). Data fortress: Innovations in big data analytics for proactive cybersecurity defense and asset protection. *International Journal of Research and Publication Review, 5*(6), 1026–1031. https://doi.org/10.55248/gengpi.5.0624.1425

[33] Prajapati, N. (2025). Federated learning for privacy-preserving cybersecurity: A review on secure threat detection. *International Journal of Advanced Research in Science, Communication and Technology, 5*(4), 520–528. https://doi.org/10.48175/IJARSCT-25168

[34] Chatterjee, P. (2022). Machine learning algorithms in fraud detection and prevention. *Eastern-European Journal of Engineering and Technology, 1*(1), 15–27.

[35] Thangaraju, V. (2025). Enhancing web application performance and security using AI-driven anomaly detection and optimization techniques. *International Research Journal of Innovative Engineering and Technology, 9*(3), 205–212. https://doi.org/10.47001/IRJIET/2025.903027

[36] Alani, M. M. (2021, June). Big data in cybersecurity: A survey of applications and future trends. *Journal of Reliability and Intelligent Environments, 7*(2), 85–114. https://doi.org/10.1007/s40860-020-00120-3

[37] Prajapati, V. (2025). Enhancing threat intelligence and cyber defense through big data analytics: A review study. *Journal of Global Research in Mathematical Archives, 12*(4), 1–6.

[38] Chityala, U. R., Shnain, A. H., Govindaraj, M., Johri, P., Kuppuraj, T., & Devi, N. L. (2025, [Month]). Big data for enhancing cybersecurity in enterprise environments: Proactive threat detection and prevention. In *2025 International Conference on Automation and Computation (AUTOCOM)* (pp. 1396–1401). https://doi.org/10.1109/AUTOCOM64127.2025.10957069

[39] Helser, S., & Hwang, M. (2025). AI and big data: Synergies and cybersecurity challenges in key sectors. *IEEE Transactions on Technology and Society, 6*(1), 54–63. https://doi.org/10.1109/TTS.2024.3465935

[40] Alawadhi, R., Aalmohamed, H., Alhashemi, S., & Alkhazaleh, H. A. (2024). Application of big data in cybersecurity. In *2024 7th International Conference on Signal Processing and Information Security (ICSPIS)* (pp. 1–6). https://doi.org/10.1109/ICSPIS63676.2024.10812589

[41] Sudan, P. (2023). A model for big data analytics using Internet of Things and cybersecurity. In *2023 International Conference on Power Energy, Environment & Intelligent Control (PEEIC)* (pp. 27–31). https://doi.org/10.1109/PEEIC59336.2023.10450652

[42] Jilani, S., Kishore, N. N., Chand, N. N., Varma, R. D., Raja, G., & Rao, P. V. (2023, January). Big data security: Detect and prevent the data from attacks with digital forensic tools. In *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)* (pp. 783–787). IEEE. https://doi.org/10.1109/ICSSIT55814.2023.10060947

[43] Nisha, S. S., Patil, H., Bag, A., Singh, A., Kumar, Y., & Kumar, J. S. (2022). Critical information framework against cyber-attacks using artificial intelligence and big data analytics. In *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*. https://doi.org/10.1109/ICACITE53722.2022.9823779