OSINT: A Double-edged Sword

1st Nitin Shrivas
Department of Cyber Security
Noida Institute of Engineering & Technology
Greater Noida, India
nitin.shrivas011@gmail.com

2nd Ajitabh Kalia
Department of Cyber Security
Noida Institute of Engineering & Technology
Greater Noida, India
ajitabhkalia257@gmail.com

3rd Ronit Roy
Department of Cyber Security
Noida Institute of Engineering & Technology
Greater Noida, India
ronitroy778999@gmail.com

4th Sumit Sharma
Department of Cyber Security
Noida Institute of Engineering & Technology
Greater Noida, India
sssumitsssharma@gmail.com

5th Puneet Garg
Department of CSE - AI
KIET Group of Institutions
Delhi NCR, Ghaziabad
puneetgarg.er@gmail.com

6th Gaurav Agarwal
Department of CE
SAITM, Gurugram
Delhi NCR, India
gauravagarwal009@gmail.com

***Abstract*— In this review, the significance of OSINT in contemporary cybersecurity is critically analysed, emphasizing the dual nature of OSINT in terms of being an effective instrument for threat identification and a latent liability in case of the alleged abuse. Our analysis examines the progression of OSINT methodologies, including the shift from raw manual data collection to the blending of AI-powered analytics for real-time intelligence acquisition. We highlight critical challenges, including how to handle an overwhelming amount of unstructured data, how to ensure that accurate data points are used without many erroneous data points, and how privacy and regulatory issues should be addressed. Furthermore, we found a lack of standardization in how data validation processes are handled in OSINT platforms, especially regarding the incorporation of fallback mechanisms. To mitigate these issues, we propose prevention strategies including rigorous quality assurance processes, robust encryption and secured access mechanisms, and autonomous, explainable AI models to map technical data to top-level threat intelligence. This review intends to create a holistic framework to facilitate the ethical and effective application of OSINT, with an eye towards more resilient cybersecurity in collectively informing research, policy, and practice in a rapidly digitalized world of October 2023.**

Keywords—Threat intelligence, Security Risk Assessment, Intelligence Gathering, OSINT Methodologies, Data Theft, Cybersecurity, Information Filtering, Cybercrime, Digital Foot-printing, Social Media Intelligence, Data Aggregation, Reconnaissance

I. Introduction

Open-source Intelligence is often called OSINT, the most crucial and cornerstone element in cybersecurity. As in cybersecurity and ethical hacking, reconnaissance is considered the initial phase. Most of the data is already available on the internet, either on the dark web or the deep web. Open-source Information (OSINF) is essential for collecting big data, which is then processed to extract valuable insights. The collection of information can be done via different methods that can be either active or passive. During active foot-printing or data gathering, an attacker interacts directly with the target system to collect information, which may be logged in the system's records. In contrast, passive foot-printing does not require any interaction between the target and the attacker; instead, it relies on analysing publicly available data from the internet.

OSINT tools can effectively filter Open-source Information (OSINF) using various methods, some of which are reliable while others are not. OSINT can be considered a double-edged sword with major advantages and disadvantages in the cybersecurity field. [1] It can be useful to gain intel about cybercrimes, cyber criminals, and cyberterrorist activities. But it also has a negative impact if used with malicious intent like for cybercrimes (i.e. DoS attacks, Hacking, Server takedowns, data theft, etc.). Therefore, reviewing details in this field becomes essential as it not only clarifies the ethical use of tools but also offers solutions to mitigate data breaches through OSINT.

Finding information from the different sources varies the complexity and credibility. For example, finding the information using social media often referred as SOCMINT, information gathered through humans referred as HUMINT and technology & information assets are used to gather enemy intelligence is known as TECHINT. Also, gathering data from the Deep Web or Dark Web is often challenging because standard browser tools and techniques are unable to access those encrypted websites. An in-depth review of these topics falls beyond the scope of this paper.

II. Objective and Scope

This paper discusses the different tools, techniques, and methodologies for OSINT with legal and ethical considerations. What challenges are we facing in reconnaissance, and how can we overcome these obstacles to leverage our tools in the most effective and impressive way possible? Additionally, this paper suggests preventive measures to protect your assets from being exposed on public platforms such as social media and public forums. It also suggests explains the advantages, disadvantages and ethical use of the tools in this modern era of cybersecurity. How the new trends overcome the previous one and what are the current tools we can use to filter out the data in a most effective way? How we can automate the entire process of locating and extracting data, followed by refining it to focus on the most relevant information in an efficient manner.

III. Reasons for Review

The rapid expansion of digital information and the advent of sophisticated data analytics have dramatically reshaped the OSINT environment. With open-source data now being employed to both pre-empt and perpetrate cybercrimes, there

exists a pressing need to reassess the balance between operational utility and security risks.

This review is motivated by the recognition of the fact that while OSINT offers unprecedented opportunities for real-time intelligence gathering and decision support, its misuse can lead to serious breaches of privacy and cybersecurity vulnerabilities. Evaluating these contrasting dimensions is essential for establishing robust governance frameworks and developing countermeasures that mitigate potential threats without stifling innovation.

IV. Previous and New Trends

Historically, data was gathered through media and other rudimentary search techniques. In 2001, Wikipedia was established, a non-profit organization often provide information which are publicly available. Whereas in this new era of OSINT whole manual process shifted towards Big-data collection techniques. Rapid development of the technology opens different techniques to filter out the data quickly and efficiently, as shown in TABLE I. New trends incorporate AI-driven tools for real-time data harvesting and managing. Use of customized regex to filter out the Big-data in a predominant way.

TABLE I *A comparison between past and current trends.*

Criteria	*Previous Trends*	*New Trends*
Data Collection Method	This approach entails directly obtaining information by thoroughly examining documents, public records, and various online sources. Traditionally, analysts perform this process manually, which is inherently time-consuming and labour-intensive. The reliance on human assessment restricts scalability and rapid responsiveness, thereby reducing its effectiveness for capturing intelligence in rapidly changing environments.	AI-driven analytics utilizes state-of-the-art machine learning algorithms to rapidly and accurately process extensive datasets. This automated strategy excels at uncovering complex patterns, trends, and anomalies that may be missed during manual review. By integrating artificial intelligence into their analytical frameworks, organizations can derive deeper insights, forecast future developments, and base their decisions on robust data—thereby markedly enhancing operational efficiency and responsiveness.
Search Techniques	Basic search methods rely heavily on elementary keyword queries and straightforward database lookups to retrieve information. These approaches typically lack sophisticated filtering and context-aware analysis, which can result in less accurate outcomes. They often demand extensive manual verification to confirm findings, thereby slowing the process and limiting the depth of insights—especially amid the rapid expansion of digital data.	Real-time data harvesting involves the continuous collection of information from a range of digital sources as events unfold. Automated tools extract data from platforms like social media, news feeds, and online resources without delay, ensuring that intelligence remains current. This method empowers organizations to conduct immediate analysis and maintain situational awareness, which is essential for swift decision-making in dynamic environments.
Data Sources	Sourcing intelligence from public records and traditional media involves gathering information primarily from government publications, newspapers, television, and radio. This method leverages the reliability and organized structure of official records and long-established news outlets. However, it may miss the rapid updates and detailed nuances offered by modern digital platforms, limiting both the timeliness and depth of the intelligence collected.	Big data integration involves combining extensive and diverse datasets from various sources, incorporating both structured and unstructured information. This holistic approach uncovers hidden correlations and insights that might be overlooked when data remains segregated. By merging different data streams, organizations can conduct more comprehensive analyses and gain a deeper, more nuanced understanding of complex issues.
Approach	These practices involve responding to events only after they have taken place, relying on sporadic data collection from disparate sources. The resulting information is often fragmented and incomplete. Such an ad hoc method lacks systematic integration and foresight, which can lead to delayed reactions and overall reduced effectiveness, as it fails to present a cohesive view of rapidly evolving situations.	In contrast, proactive approaches leverage the expertise of multiple disciplines to foresee and address challenges before they fully emerge. By combining technological, social, and analytical insights, this strategy creates a holistic understanding of complex environments. Such collaboration and forward-thinking enable organizations to detect risks and opportunities early, ensuring prompt and effective responses.
Information Processing	Static information retrieval refers to the practice of accessing archived or static datasets that are not regularly updated. This approach relies on periodic, manual searches through established records, which may not reflect the most current information. While it offers dependable historical data, its absence of dynamic analysis and infrequent updates limit its effectiveness in monitoring fast-changing situations or responding promptly to emerging trends.	Dynamic, continuous monitoring refers to the persistent observation and systematic analysis of data streams to swiftly identify emerging risks and opportunities. By leveraging automated systems and advanced analytical tools, this approach offers an up-to-date snapshot of the operational landscape. It facilitates prompt responses and well-informed decision-making, enabling organizations to efficiently mitigate risks and exploit favourable trends as they arise.

V. TECHNIQUES AND TOOLS

Advancements in technology have led to the development of sophisticated OSINT tools that includes automation, machine learning, and natural language processing.

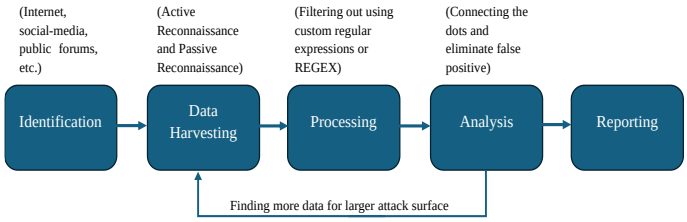

***Fig. 1**. Data flow in the process of OSINT*

Techniques such as network mapping, and geospatial intelligence and social analysis further enhance the analytical capabilities of OSINT systems. Additionally, integration with artificial intelligence frameworks allows for predictive modelling and anomaly detection, i.e. improving both the accuracy and relevance of intelligence outputs. Modern techniques in OSINT follow several common steps to gather, process, analyse, and report the final outcomes. These steps include identification, data collection, processing, analysis, and reporting, as illustrated in **Fig. 1**.

1. **Identification**: The very first logical step is to specifically define your requirements and the type of information you are looking for or form which source. Determine the limits on what you will be investigating (i.e. time or geographical), and define the areas you must cover. It also involves choosing the most suitable sources such as social media platforms, public records, or research articles. Legal and ethical factors are analysed to ensure the subsequent data collection procedure will be in according to the standards. [1], [2] [23]
2. **Data Collection:** After the intelligence requirements have been established, it is time to collect the raw data from the sources identified. This is achieved through automated as well as manual processes. Data extraction methods incorporate automated tools (web crawlers and OSINT tools) for systematic data retrieval from web-based sources and manual processes for precision or when the data source cannot be queried through automation. Data may be in the text, image, or metadata forms. The approach may be active, in which systems and APIs are called, or passive, in which data is pulled with no digital trace. [1] [22]
3. **Processing**: Data may be in various forms and may be obtained through various sources with varying data forms. The raw data should be processed and converted into analysable data. The data is purified by eliminating duplicate fields and extraneous data in this stage, while normalization is performed to ensure consistency in various data types and data forms. The data obtained is placed in databases or spreadsheets, which allows filtering and sorting operations. The process ensures only high-quality data is preserved.
4. **Analysis**: It is the process of in-depth examination of the processed data in order to draw relatable conclusions out of it. Analysts look for patterns and corelations in the data which can be helpful to identify relationships in the data. Verification techniques, including cross-verification with several sources, are applied to ensure the credibility and reliability of the data. Statistical analysis, network mapping, and AI-based analytics can be part of the analytical procedures, with the use of manual as well as automated tools and methods. The purpose of this stage is to filter out the information into a meaningful way possible.
5. **Reporting**: The final stage is documenting of the results in a report to present the findings clearly and effectively. The documentation of the methods, data sources, and the analytical process is incorporated in this stage. To show the relationship between different data, illustration are added like pie chart, diagram, tables, etc. The report is written for the intended audience, whether law enforcement officers, in-house cybersecurity professionals, or business decision-makers, so the analysis is presented in a respectful manner regarding confidentiality and ethical standards.[24]

Modern platforms such as Maltego, theHarvester, Recon-NG, Sherlock and various social media monitoring systems enable correlation and analysis of data from different and various sources. These tools offer a variety of user-friendly interfaces, including Graphical User Interface (GUI), Command Line Interface (CLI), and menu-driven options to ensure users' interaction easily and less frustrating. Simple steps must engage users to interact with the tool. Therefore, using machine learning and advanced regex methods can enhance the effectiveness of any tool. Some custom and common regex patterns are illustrated in **TABLE II**.

***TABLE II** Common regex patterns*

Use Case	Regex Pattern	Example Input	Extracted Output
Email Address Extraction	`\b[A-Za-z0-9._%+-]+@[A-Za-z0-9.-]+.[A-Za-z]{2,}\b`	echo "Contact us at support@example.com"	support@example.com
IP Address Detection	\b(?:[0-9]{1,3}\.){3}[0-9]{1,3}\b	Suspicious activity from 192.168.1.1	192.168.1.1
URL Extraction	https?:\/\/(?:www\.)?[^\s/$.?#].[^\s]*	Visit https://www.cybersecurity.com	https://www.cybersecurity.com
Credit Card Number Detection	\b(?:\d[-]*?){13,16}\b	Card details: 4111-1111-1111-1111	4111-1111-1111-1111
SSN Detection (US)	\b\d{3}-\d{2}-\d{4}\b	John Doe's SSN is 123-45-6789	123-45-6789
Phone Number Extraction (US)	\(?\d{3}\)?[-.\s]?\d{3}[-.\s]?\d{4}	Call me at (555) 123-4567	(555) 123-4567

Bitcoin Address Detection	\b[13][a-km-zA-HJ-NP-Z1-9]{25,34}\b	Send BTC to 1A1zP1eP5QGefi2DMPTfTL5SLmv7DivfNa	1A1zP1eP5QGefi2DMPTfTL5SLmv7DivfNa

Above Regex patterns are most commonly and widely used to extract the valuable information quickly and efficiently. **Fig. 2** illustrates the use of regex pattern to extract the URL from the result and give the tailored output as per the

Fig. 2 *Execution of URL extraction using REGEX pattern.*

requirement. In this, sherlock is used to find the possible account on the web corresponding to 'test_user123' and then grep is used to filter out URL based on REGEX provided where '-Eo' option is used for 'Extended regex' and 'only matching' respectively.[18]

VI. ADVANTAGES AND DISADVANTAGES

OSINT has several important advantages:

- Timeliness and Accessibility: OSINT platforms excel at rapidly collecting the most up-to-date publicly available information from a broad variety of sources. The rapid accumulation of information allows organizations to identify emerging trends and probable threats almost in real-time, giving decision-makers the most up-to-date information in order to respond appropriately to changing conditions.
- Cost-Effectiveness: OSINT is less costly than traditional intelligence gathering. By using open sources of information in combination with low-cost analytical tools, organizations can acquire required information at low costs, offering a cost-effective choice for large organizations as well as for smaller organizations.
- Global Reach: Gathering data from a broad cross-section of international open sources, OSINT presents a balanced international affairs outlook and trends. The extensive capability to gather data in different domains and geographic regions creates a richer, more comprehensive view of the general environment.[15] [17]
- Scalability: OSINT systems are designed to process huge amounts of data automatically. Since the quantity of information released to the public is growing, the systems can increase their data-processing capacity to maintain the intelligence cycle effective and efficient despite the growing data volumes.

Nonetheless, these advantages are counterbalanced by substantial drawbacks:

- Overload of Data and Quality Problems: The quantity of data can be so vast as to be intimidating, with the result being it is hard to determine which data is accurate. Sorting, verification, and removal of old or inaccurate data require additional work for the resulting intelligence to be accurate and dependable.
- Privacy and Ethical Issues: The open nature of OSINT creates the risk of inadvertently disclosing confidential business or personal data. Unintended disclosure can be very serious in its ethical and privacy implications with potentially long-lasting effects for individuals as well as organizations.[21]
- Security Threats: OSINT can be intentionally misused or manipulated, resulting in its misuse in cyberattacks or other malicious activities. Thus, it must be authenticated in minute detail to prevent any vulnerabilities being taken advantage of by the enemies. [19] [20]

Regulatory and Compliance Issues: Various data protection legislation in various geographic areas present difficulties for organizations relying on OSINT. It is difficult to be in compliance with such shifting legal requirements, with non-compliance likely to be a regulatory issue, which will require ongoing legal supervision.

VII. LITERATURE SURVEY

TABLE III *Comparative analysis of various research papers related to Open-Source Intelligence (OSINT).*

Author(s)	*Title*	*Summary*	*Research Gaps*
Spencer P. Chainey Arantza Alonso Berb otto	A structured, methodical process for populating a crime script of organised crime activity using OSINT [2]	This paper brings to light the fact that there is a lack of systematic methods for incorporating data obtained through OSINT tools and techniques into crime scripts, which results in inconsistent quality of analysis. This study talks about how hard it is to check and validate unstructured data gathered through OSINT, which can lead to mistakes. It also talks about how structured OSINT processes should be integrated	The key issue remains that the data available online is seldom reliable. The paper also doesn't provide ways to try and remedy it.

		with existing crime analysis frameworks to make the data more reliable.	
Yong-Woon Hwang, Im-Yeong Lee, Hwankuk Kim, Hyejung Lee, Donghyun Kim	Current Status and Security Trend of OSINT [1]	This research points out the duality of OSINT tools and techniques, how it is an essential art of ethical hacking, and also how much of a risk it can be if misused. It also identifies the unnoticed gaps present in the sources of OSINT data, thus raising questions about the validity of the said data. This paper urged for improved security measures and validation protocols that are much more robust than what we currently have to mitigate the threats that might arise due to it.	The study majorly focuses on basic security measures and doesn't give any specific technologies that could improve OSINT
João Rafael Gonçalves Evangelista, Renato José Sassi, Márcio Romero & Domingos Napolitano	Systematic Literature Review to Investigate the Application of Open-Source Intelligence (OSINT) with Artificial Intelligence [3]	This study stresses the fact that there is a major gap in the field that can be filled by combining OSINT with AI. It also emphasises the need for a unified framework that follows the latest research trends from certified sources. It urges the development of systematic methodologies that leverage AI for improved and more efficient processing of data gathered through OSINT.	The paper doesn't give technical demonstration of the topics covered.
Randa Basheer and Bassel Alkhatib	Threats from the Dark: A Review over Dark Web Investigation Research for Cyber Threat Intelligence [4]	This study brings to attention the difficulty faced while gathering intelligence from the dark web due to the intentional anonymisation of data. It also points out the fact that there currently isn't a methodology that effectively and reliably combines the data from the dark web with more mainstream threat intelligence frameworks. This study also urges the development of improved techniques that can overcome the challenges faced while OSINT processing data from the dark web to ensure higher reliability and authenticity.	The paper suggests use of tools like social media monitoring tools, but doesn't suggest ways to improve low precision and recall rates of such tools
Yi-Ting Huang, Chi Yu Lin, Ying-Ren Guo, Kai-Chieh Lo, Yeali S. Sun, and Meng Chang Chen	Open-Source Intelligence for Malicious Behaviour Discovery and Interpretation [5]	The study highlights the challenge of linking low-level malware execution traces to high-level threat intelligence descriptions, which hampers actionable insight generation. It underlines difficulties in automatically correlating API call data with known adversary tactics, leading to potential	In the study, the primary focus is given to windows, overlooking other operating systems and environments.

		misinterpretation. The paper advocates for the development of explainable AI models to bridge the gap between raw OSINT data and clear, interpretable threat intelligence.	
Ashok Yadav, Atul Kumar, Vrijendra Singh	Open-source intelligence: a comprehensive review of the current state, applications and future perspectives in cyber security [6]	This review discusses the challenge of extracting actionable insights from a vast array of unstructured public data, emphasizing the need for automated analysis methods. It identifies the absence of standardized workflows and benchmarks in OSINT, which limits consistent application across the cybersecurity field. The paper points to a gap in developing autonomous models capable of real-time OSINT processing, calling for future research to address these shortcomings.	Generalisations don't account for changes in legal frameworks of different countries

VIII. CHALLENGES FACED

In OSINT environment, more and more data are stored on the internet and many companies also steal users' data privately. As Google, Microsoft and Amazon secretly abduct users' personal information, such as name, number, date of birth, etc. These companies claim that all data should be anonymised and encrypted but the transparency is still missing. Because of such amount of Big-data available on the internet many problems and challenges arise while capturing suitable data from OSINT tool, which are as follows:

- **Data from the Underground:** In fact, with millions of users on various dark web marketplaces, you have access to a wealth of information you may never realize. Imagine the raw, unprocessed pools of Big-data available on the web, from which it is difficult to draw meaningful and reliable conclusions. Companies may not know of confidential employee or customer data that has unintentionally leaked over the internet which is less visible to the common users that can be found using advanced search methods or tor network, using deep web dedicated search engines. There's just so much information flooding into the digital realm that more advanced filtering systems are required to slice through it to reach the relevant content. Most open-source intelligence tools fail to rank and prioritize relevant data effectively, leaving end-users digging through tons of irrelevant content to focus on high strategic-value intelligence. The credibility and authenticity of data are a great challenge. There is fake data, as well as manipulated content online, which leads analysts to the long cross-verification cycle with other sources. There must be no inaccuracy when it comes to putting in place a strict verification process that evaluates the intelligence gathered as reliable and actionable. [4][14]
- **https code of ethics:** One we need to pay attention to the ethical code of conduct concerning our practice. It's on technology firms to provide privacy for user data without ever being totally transparent while also designing the systems to encrypt/encrypt data. At the same time, the lack of transparent data practices makes it challenging to safeguard people's privacy, especially under strict regulatory frameworks like GDPR.[13]
- **Computational Requirements and Scalability**: The volume of OSINT data in its unstructured form forces you to leverage powerful computing power and complex algorithms. Intensive mixing various forms of data from various sources leads to increased complexity, which necessitates complex normalization and transformation processes in the field of efficient scalability management.
- **Real-Time Surveillance and Data Synthesis:** The dynamic nature of digital data is an obstacle to conducting real-time analysis. Fusing disparate data into a cohesive narrative requires strong analytical abilities, observation in real-time. If emerging trends and threats are not identified, there may be a delay between the collection of intelligence and the usefulness of the data.[12]

IX. PREVENTION

- **Control Access:** Implement a Role-based access for all the employees in an organization or a company to avoid insider threat which supresses the threat of leaking confidential information on the internet. Zero-trust policy and MFA (Multi-factor Authentication) should be adopted to secure company information.
- **Encrypt and Protect Data:** The data should be secured with sophisticated encryption, limited access, and ongoing authentication in order to protect confidential information and prevent misuse.

Encryption of data in-transit as well as in-rest for robust security.

- **Regular Monitoring:** Use of different tools and software available on the web which can easily find your leaked information. Do act on it and immediately delete your personal information leaked on the web through public forums, social media, etc. So, regular monitoring is the most important in all the mitigation. Avoid oversharing personal details on platforms such as social media. [1]

- **Maintain Backup and Recovery:** Three backup technique should be used in practice to avoid the future risks of data loss and ransomware attack caused due to data leakage. As we know *"Prevention is better than cure". [10]*

- **Integrating AI with Machine Learning for Proactive Detection:** Use of Artificial Intelligence is proliferated and utilize for data filtering and threat monitoring in order to minimize false positives and the timely detection of probable cyber threats. Develop interpretable AI models that map low-level technical traces back to high-level threat intelligence, enabling clearer understanding of malicious behaviours in order to facilitate rapid countermeasures. Incorporating artificial intelligence into existing OSINT tools to enhance their efficiency and accuracy in data collection and analysis. [7]

- **opt for Anonymity:** Use pseudonyms or alias names wherever possible. [1],[3] It hides the person's PII and maintain the integrity and privacy. Use of temp mail and alias for different platforms can prevent your data to be shared over other advertisement companies.

- **Proper Awareness:** Educate your surroundings as well as self-awareness is necessary. For example, avoid use of public Wi-Fi and other shared network. Keep update yourself about latest techniques and mitigation of current trends in cybersecurity. Organization should implement the awareness programme for their employees.

X. CONCLUSION

Open-source intelligence (OSINT) is a very significant part of cybersecurity in present times as it is the first step in reconnaissance that an attacker or an ethical hacker should perform further investigates its target on the network which offers low-cost, and in-depth information for the proactive detection of threats. Too much raw data over the internet gives rise to many problems to filter it effectively. But data collection is also the foremost problem, such as the credibility of the data, complexity, unstructured and standards.

This review points out the importance of OSINT for cybersecurity for the purpose of analysing cyber threats. Its impact can be limited by problems such as the management of unstructured data and the accidental leakage of confidential data. In response to such issues, there is a requirement for strong security in the form of standardizing data validation procedures, the implementation of strong encryption, and strict access control in order to maintain data integrity. Additionally, the application of artificial intelligence, machine learning, and complex regex can screen the data more efficiently, which enhances the speed of threat detection and converts RAW data into actionable intelligence with higher clarity. Real-time monitoring also facilitates proactive defence by enabling the detection of emerging threats in organizations at the nascent stage. [11]

REFERENCES

[1] Hwang, Y. W., Lee, I. Y., Kim, H., & Lee, H. (2022). **Current status and security trend of OSINT**. Wireless and Mobile Computing. Wiley Online Library. https://onlinelibrary.wiley.com/doi/abs/10.1155/2022/1290129

[2] Chainey, S. P., & Alonso Berbotto, A. (2022). **A structured methodical process for populating a crime script of organized crime activity using OSINT**. Trends in Organized Crime. Springer. https://link.springer.com/article/10.1007/s12117-021-09428-9

[3] Evangelista, J. R. G., Sassi, R. J., Romero, M., & others. (2021). **Systematic literature review to investigate the application of open-source intelligence (OSINT) with artificial intelligence**. Journal of Applied .... Taylor & Francis. https://www.tandfonline.com/doi/abs/10.1080/19361610.2020.1761737

[4] Basheer, R., & Alkhatib, B. (2021). **Threats from the dark: A review over dark web investigation research for cyber threat intelligence**. *Journal of Computer Networks and Communications, 2021*, Article ID 1302999. https://onlinelibrary.wiley.com/doi/pdf/10.1155/2021/1302999

[5] Huang, Y.-T., Lin, C. Y., Guo, Y.-R., Lo, K.-C., Sun, Y. S., & Chen, M. C. (n.d.). **Open-source intelligence for malicious behavior discovery and interpretation.** *IEEE Transactions on Dependable and Secure Computing*. https://www.iis.sinica.edu.tw/papers/mcc/24312-F.pdf

[6] Yadav, A., Kumar, A., & Singh, V. (2023). **Open-source intelligence: A comprehensive review of the current state, applications, and future perspectives in cybersecurity**. *Artificial Intelligence Review, 56*, 12407–12438. https://link.springer.com/content/pdf/10.1007/s10462-023-10454-y.pdf

[7] 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, 2911–2938. https://doi.org/10.1007/s10207-024-00868-2

[8] Dincelli, E., Van Slyke, C., & Yayla, A. (2023). **Ethical Hacking for a Good Cause: Finding Missing People using Crowdsourcing and Open-Source Intelligence (OSINT) Tools**. Communications of the Association for Information Systems, 53, 1052-1071. https://doi.org/10.17705/1CAIS.05345

[9] Pastor-Galindo, J., Nespoli, P., Mármol, F. G., & others. (2020). ***The not yet exploited goldmine of OSINT: Opportunities, open challenges, and future trends*.** IEEE Access. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8954668

[10] Gautam, V. K., Gupta, S., & Garg, P. (2024, March). Automatic Irrigation System using IoT. In 2024 International Conference on Automation and Computation (AUTOCOM) (pp. 100-103). IEEE. DOI: 10.1109/AUTOCOM60220.2024.10486085

[11] [26] Ramasamy, L. K., Khan, F., Joghee, S., Dempere, J., & Garg, P. (2024, March). Forecast of Students' Mental Health Combining an Artificial Intelligence Technique and Fuzzy Inference System. In 2024 International Conference on Automation and Computation (AUTOCOM) (pp. 85-90). IEEE. https://doi.org/10.1109/AUTOCOM60220.2024.10486194

[12] [27] Rajput, R., Sukumar, V., Patnaik, P., Garg, P., & Ranjan, M. (2024, March). The Cognitive Analysis for an Approach to Neuroscience. In 2024 International Conference on Automation and Computation (AUTOCOM) (pp. 524-528). IEEE. DOI: 10.1109/AUTOCOM60220.2024.10486081

[13] [28] Dixit, A., Sethi, P., Garg, P., Pruthi, J., & Chauhan, R. (2024, July). CNN based lip-reading system for visual input: A review. In AIP Conference Proceedings (Vol. 3121, No. 1). AIP Publishing. https://doi.org/10.1063/5.0221717

[14] [29] Bose, D., Arora, B., Srivastava, A. K., & Garg, P. (2024, May). A Computer Vision Based Framework for Posture Analysis and Performance Prediction in Athletes. In 2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE) (pp. 942-947). IEEE. DOI: 10.1109/IC3SE62002.2024.10593041

[15] [30] Singh, M., Garg, P., Srivastava, S., & Saggu, A. K. (2024, April). Revolutionizing Arrhythmia Classification: Unleashing the Power of Machine Learning and Data Amplification for Precision Healthcare. In 2024 Sixth International Conference on Computational Intelligence and Communication Technologies (CCICT) (pp. 516-522). IEEE. DOI: 10.1109/CCICT62777.2024.00086

[16] [31] Kumar, R., Das, R., Garg, P., & Pandita, N. (2024, April). Duplicate Node Detection Method for Wireless Sensors. In 2024 Sixth International Conference on Computational Intelligence and Communication Technologies (CCICT) (pp. 512-515). IEEE. DOI: 10.1109/CCICT62777.2024.00085

[17] [32] Bhardwaj, H., Das, R., Garg, P., & Kumar, R. (2024, April). Handwritten Text Recognition Using Deep Learning. In 2024 Sixth International Conference on Computational Intelligence and Communication Technologies (CCICT) (pp. 506-511). IEEE. DOI: 10.1109/AIST55798.2022.10065348

[18] [33] Gill, A., Jain, D., Sharma, J., Kumar, A., & Garg, P. (2024, May). Deep learning approach for facial identification for online transactions. In 2024 International Conference on Emerging Innovations and Advanced Computing (INNOCOMP) (pp. 715-722). IEEE. DOI: 10.1109/INNOCOMP63224.2024.00123

[19] [34] Mittal, H. K., Dalal, P., Garg, P., & Joon, R. (2024, May). Forecasting Pollution Trends: Comparing Linear, Logistic Regression, and Neural Networks. In 2024 International Conference on Emerging Innovations and Advanced Computing (INNOCOMP) (pp. 411-419). IEEE. DOI: 10.1109/INNOCOMP63224.2024.00074

[20] [35] Malik, T., Nandal, V., & Garg, P. (2024, May). Deep Learning-Based Classification of Diabetic Retinopathy: Leveraging the Power of VGG-19. In 2024 International Conference on Emerging Innovations and Advanced Computing (INNOCOMP) (pp. 645-651). IEEE. DOI: 10.1109/INNOCOMP63224.2024.00111

[21] [36] Srivastava, A. K., Verma, I., & Garg, P. (2024, May). Improvements in Recommendation Systems Using Graph Neural Networks. In 2024 International Conference on Emerging Innovations and Advanced Computing (INNOCOMP) (pp. 668-672). IEEE. DOI: 10.1109/INNOCOMP63224.2024.00115

[22] [37] Aggarwal, A., Jain, D., Gupta, A., & Garg, P. (2024, May). Analysis and Prediction of Churn and Retention Rate of Customers in Telecom Industry Using Logistic Regression. In 2024 International Conference on Emerging Innovations and Advanced Computing (INNOCOMP) (pp. 723-727). IEEE. DOI: 10.1109/INNOCOMP63224.2024.00124

[23] [38] Mittal, H. K., Arsalan, M., & Garg, P. (2024, May). A Novel Deep Learning Model for Effective Story Point Estimation in Agile Software Development. In 2024 International Conference on Emerging Innovations and Advanced Computing (INNOCOMP) (pp. 404-410). IEEE. DOI: 10.1109/INNOCOMP63224.2024.00073