ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DECODING ADVERSARIAL MACHINE LEARNING: A BIBLIOMETRIC PERSPECTIVE

JEENA JOSEPH [1], JOBIN JOSE [2], DIVYALAKSHMI S [3], RAJIMOL A[4], GREETY TOMS [5], ANAT SUMAN JOSE [6], GILU G ETTANIYIL[7]

[1] Assistant Professor, Marian College Kuttikkanam Autonomous, Department of Computer Applications, Idukki, Kerala, India
[2] Librarian, Marian College Kuttikkanam Autonomous, Idukki, Kerala, India
[3] Assistant Professor, Marian College Kuttikkanam Autonomous, Department of Computer Applications, Idukki, Kerala, India
[4] Associate Professor, Marian College Kuttikkanam Autonomous, Department of Computer Applications, Idukki, Kerala, India
[5] Bharata Mata College,Thrikkakara, Kerala, India
[6] Librarian, St. Peter's College Kolenchery, Kerala, India
[7] Librarian, St. Thomas College of Teacher Education, Pala, Kottayam, Kerala, India

E-mail: [1]jeena.joseph@mariancollege.org, [2]jobin.jose@mariancollege.org, [3]divyalakshmi.s@mariancollege.org,[4] rajimol.a@mariancollege.org, [5]greetysony@bharatamatacollege.in, [6]anatsumanjose1@gmail.com, [7]gilu@stcte.ac.in

ABSTRACT

The rapid improvement of machine learning techniques has led to an extraordinary rise in the prominence of adversarial attacks and their accompanying defenses. This study performs a comprehensive bibliometric analysis of adversarial machine learning, providing insight into the field's evolution from its foundation to the present. We identify the primary themes, foundational works, and influential figures in this field using state-of-the-art bibliometric technologies and databases. Our findings demonstrate the impressive growth of adversarial machine learning research and emphasize its transdisciplinary nature. We also highlight the collaborative networks and important hubs that have fueled advancements in this discipline. This report offers a thorough perspective on adversarial machine learning, its major turning points, and valuable insights for researchers, educators, and practitioners.

Keywords: *Adversarial Machine Learning, Bibliometric Analysis, VOSviewer, Biblioshiny.*

1. INTRODUCTION

Machine learning (ML) has now evolved into a cornerstone of contemporary technology, influencing sectors ranging from healthcare to finance [1]. As machine learning (ML) models become more complex and necessary to critical systems, protecting them against deliberate manipulations has become a top priority [2]. Adversarial machine learning explores the intentional alteration of inputs to trick machine learning models and aims to provide defenses against these kinds of attacks [3]–[5].

The fundamental investigation of the security and dependability of machine learning algorithms is the source of adversarial machine learning [6], [7]. These algorithms perform exceptionally well in benign contexts but are vulnerable to adversarial interventions, which has revealed exciting issues and sparked a competition to develop new assaults and defences [8]. This dynamic interaction has spurred a wave of study in this field, making it one of the most dynamic and rapidly growing subfields in machine learning [9].

However, the large amount of research and the variety of applications in adversarial machine learning can make it challenging for practitioners and scholars to investigate the area thoroughly, identify the most significant discoveries, and identify overarching trends. A bibliometric study provides a systematic, empirically-based way to understand this process since it uses quantitative evaluations of academic publications to identify patterns in scientific research [10]–[12].

This study aims to delineate the theoretical trajectory of adversarial machine learning, furnishing readers with an all-encompassing viewpoint on its development. We will examine critical subjects, notable publications, and well-known scholars using sophisticated bibliometric techniques, illuminating the complex web of knowledge that characterizes the area. Our goal in going back in time is to lay the groundwork for future research by identifying areas of interest and potential roadblocks.

Bibliometric analysis is a potent approach for quantitatively evaluating and interpreting patterns in academic publications, providing valuable insights into the dynamics of scientific research across diverse fields [12], [13]. These assessments can unveil trends, pivotal works, collaborative networks, and other essential aspects of academic disciplines. Various tools have been created to leverage the potential of bibliometric data [14]–[16]. Take VOSviewer, for instance; it's a specialized software designed for constructing and visually representing bibliometric networks, encompassing co-authorship, co-citation, and keyword co-occurrence networks [17]–[20]. With its user-friendly graphical interface and distinctive clustering algorithm, VOSviewer lets users explore complex networks and intuitively perceive underlying patterns [21], [22].

Conversely, biblioshiny is an R-based application integrated into the 'bibliometrix' package. It furnishes an accessible web interface for conducting comprehensive bibliometric analyses without necessitating direct interaction with R code [23], [24]. Combining data extraction, processing, and visualization functions, biblioshiny simplifies the bibliometric research process, making it approachable even for individuals with limited programming expertise [25], [26]. Tools like VOSviewer and biblioshiny have transformed how researchers engage with bibliometric studies, presenting streamlined solutions for dissecting and comprehending the extensive realm of scientific literature [25].

The objectives of conducting the bibliometric analysis on adversarial machine learning are:

- Identify Key Contributors: Determine the researchers, institutions, and countries that have made significant contributions to the field of adversarial machine learning. This can help in recognizing leading experts and centers of excellence.
- Publication Trends: Analyze the growth in the number of publications related to adversarial machine learning over time. Identify periods of increased research activity and observe if there are any fluctuations.
- Co-authorship Network: Explore collaboration patterns among researchers by constructing co-authorship networks. Identify research clusters or groups that frequently collaborate on adversarial machine learning research.
- Keyword Analysis: Identify the most frequently used keywords and terms in the titles and abstracts of publications. This can reveal emerging topics and areas of interest within adversarial machine learning.
- Journal and Conference Analysis: Determine which journals and conferences are the most popular venues for publishing research on adversarial machine learning. This can help researchers decide where to submit their work and where to look for relevant papers.
- Geographical Analysis: Investigate the geographical distribution of research contributions. Determine which countries are leading in terms of research output and collaboration.
- Evolution of Topics: Analyze how research topics within adversarial machine learning have evolved over time. Identify shifts in focus, emerging subfields, and interdisciplinary connections.
- Identify Influential Papers: Find the most influential or highly cited papers in the field. These papers often provide foundational knowledge or key insights into adversarial machine learning.

2. REVIEW OF LITERATURE

Adversarial machine learning is a rapidly evolving field with a focus on understanding and mitigating vulnerabilities in machine learning systems, especially in the face of malicious inputs or alterations.

Functionality-preserving models are essential for robust classification in adversarial machine learning, especially in cybersecurity and intrusion detection applications. These models are designed to maintain their functionality while addressing potential vulnerabilities to adversarial attacks [27]. Adversarial training can enhance the robustness of machine learning models. However, certain models, like WordCNN and LSTM-based models, show varying degrees of susceptibility to temporal changes, which impacts their resilience to adversarial tactics [28].

Adversarial examples pose a significant threat to machine-learning-based systems, especially deep neural networks. This concern is heightened by the inclusion of semantic, contextual, and system-specific specifications in adversarial strategies [29]. Adversarial machine learning involves various attack types and defense mechanisms, particularly in security applications, where traditional algorithms can be manipulated to underperform or fail [30]. Robust learning techniques, developed through game-theoretic approaches, are being increasingly used to address adversarial attacks in cybersecurity applications [31].

Adversarial attacks can be conducted by manipulating data during training or testing, significantly impacting the effectiveness of classifiers, such as those used in malware detection systems [32]. Adversarial examples, created by slightly modifying input data, are a major challenge for the robustness of modern machine learning systems, posing significant security concerns [33].

Adversarial machine learning encompasses a range of techniques and approaches aimed at enhancing the resilience of machine learning systems against adversarial attacks. This involves a comprehensive understanding of the types of attacks, the vulnerabilities of various models, and the development of robust defense mechanisms. The field is characterized by its interdisciplinary nature, combining elements from cybersecurity, deep learning, semantic analysis, and game theory.

3. MATERIALS AND METHODS

We opted for the Scopus database for our investigation due to its comprehensive collection of academic papers spanning various disciplines. Our methodical inquiry utilized the phrase "adversarial machine learning." We included documents in all languages, focusing solely on journal articles and conference papers. Duplicate entries were identified and removed. We also rectified any inconsistencies in the authors' names and their affiliations. From 509 distinct sources, we amassed 1071 papers, encompassing the years 2011 through 2023. These findings were archived in a CSV file, and we conducted a bibliometric evaluation employing VOSviewer version 1.6.19 along with Biblioshiny software.

4. RESULTS

4.1 Annual Scientific Production

The Figure 1 showcases the trend of annual scientific production on adversarial machine learning from 2011 to 2023. From the visual, it's evident that there was a relatively minimal number of articles during the early years, especially between 2011 and 2015. Starting in 2016, there's a noticeable upward trajectory in the production of papers. This growth became even more pronounced around 2018, with articles increasing sharply—the production peaks around 2022, after which there appears to be a slight dip in 2023. The overall trend underscores the growing importance and attention towards adversarial machine learning in the scientific community over the years.

Figure 1. Annual scientific production.

4.2 Most Relevant Authors

Based on their publication count, the most prominent authors in the field are Biggio B and Sagduyu Ye lead the list, each with 25 articles. They are closely followed by Shi Y, who has contributed 20 articles. Wang Y has authored 18 pieces, while Roli F has 16. Wang J has penned 15 articles, and Catak Fo follows with 14. Di Noia T and Erpek T have each written 13 articles, and rounding off the list is Chen Y with 12 publications. These individuals have showcased their profound expertise in the domain and have significantly impacted the field with their extensive contributions.

4.3 Most relevant sources and affiliations

Figure 2 shows the number of documents associated with various sources, likely academic journals, conference proceedings, or other publication outlets. Each horizontal bar corresponds to a different source, with the length of the bar representing the number of documents from that source. The source with the most documents is "Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)" with a notable 73 documents, as it has the longest bar on the chart. "IEEE Access" appears to be the second most relevant source with 39 documents. The "Proceedings of the ACM Conference on Computer and Communications Security" has 28 documents.

Other sources listed include various proceedings and transactions, with document counts ranging from 11 to 23.

Figure 2. The top ten sources

Figure 3 represents a count of articles associated with various academic or research institutions. The affiliations are listed along the y-axis, while the x-axis represents the number of articles. Each bar's length corresponds to the number of articles related to the respective affiliation. The University of Cagliari with the most articles is at the top of the y-axis, with a significantly longer bar suggesting 40 articles. The universities listed vary in academic contributions, with several having close numbers of articles, such as those with around 21 to 23 articles. The University of California and Tsinghua University are among the institutions with higher counts, 33 and 26 articles, respectively.

Figure 3. Most relevant affiliations

4.4 Trend Topics

Figure 4 offers a detailed overview of the evolution of critical topics in adversarial machine learning from 2011 to 2023. Throughout this period, consistent trends are observed in terms such as "Machine Learning" and "Deep Learning." Notably, while "Machine Learning" maintains a steady presence, the prominence of "Deep Learning" began rising around 2016-2017, mirroring the real-world surge in deep learning research and applications. Terms closely intertwined, namely "Adversarial Machine Learning," "Adversarial Attacks," and "Adversarial Learning," have seen an uptick in the latter years, underscoring the escalating research interest in adversarial techniques within machine learning. Security-centric terminologies, including "Malware Detection," "Computer Security," and "Privacy," also feature in the graph, emphasizing the growing necessity to secure prevalent machine learning systems and uphold user privacy. Recent years have marked a spike in terms like "Poisoning Attacks" and "Exploratory Attacks," suggesting these might be burgeoning areas or potential threats in adversarial machine learning research. Additionally, "Game Theory" and "Classification" have emerged as trending topics, with the former potentially highlighting its role in modelling adversarial scenarios and the latter indicating research into the impact of adversarial attacks on classification models. This graph paints a rich tapestry of the shifting focus and concerns in the adversarial machine-learning landscape over the past decade.

Figure 4. Trend Topics.

4.5 Thematic Map

The thematic map in Figure 5 comprehensively visualizes various themes in a specific research domain, categorizing them based on their relevance and developmental progress. On one side of the spectrum, we have the Niche Themes, which encompass specialized topics such as "adversarial learning," "feature extraction," and "features extraction." These themes are unique to specific research areas and may not be as widespread as others. Centrally positioned are Emerging or Declining Themes, representing areas still evolving or undergoing transformation. These middle-ground themes include "machine learning," "network security," and "learning algorithms." Shifting focus to the Motor Themes, these signify the core subjects within the research domain. They are the driving forces of the domain and include pivotal topics like "machine learning," "adversarial machine learning," and "learning systems." Lastly, at the foundation of the field lie the Basic Themes, which serve as the bedrock upon which other themes build. The entire map is a tool to help academics and professionals gauge the prominence and trajectory of various themes within their domain, offering insights into the current landscape and potential future directions.

Figure 5. Thematic Map

4.6 Top Ten Cited Articles

Adversarial machine learning has witnessed remarkable research contributions over the past years, with numerous papers delineating innovative techniques and insights. Among the most influential works in this domain is "One Pixel Attack for Fooling Deep Neural Networks" by Su J., Vargas D.V., and Sakurai K., published in 2019. This paper stands out with an impressive 1,113 citations. Another seminal work from 2013 titled "Evasion attacks against machine learning at test time" by Biggio B., Corona I., Maiorca D., Nelson B., Srndic N., Laskov P., Giacinto G., and Roli F. follows closely with 991 citations. As in Table 1, the list continues with various other impactful contributions, including works on poisoning attacks, back-gradient optimization, and applications to neural network attacks. The continuous growth in citations for these papers underscores the importance and evolving nature of adversarial machine learning in the contemporary research landscape.

Table 1. Top ten cited papers

4.7 Co-occurrence of keywords

The visualization in Figure 6 offers a comprehensive insight into the interrelated keywords associated with adversarial machine learning. Central to the discussion is "adversarial machine learning," emphasizing its significance in the depicted network. Closely linked to this are "deep learning" techniques, which underscore the integration of adversarial strategies in sophisticated machine learning models. A notable emphasis on "computer vision" suggests a deep exploration of adversarial tactics within the image recognition domain. A particular type of adversarial strategy, "poisoning attacks," which pertains to deliberately manipulating data to misguide models, also stands out. Security concerns in adversarial machine learning are evident with terms like "security," "cybersecurity," and "computer crime," highlighting the growing importance of this domain in safeguarding digital infrastructures. Moreover, the network suggests a keen interest in deploying adversarial techniques for "intrusion detection," crucial for identifying unauthorized data accesses. The domain's expansion into text is evident with terms like "natural language processing" and "text processing." Furthermore, mentioning "IoT" or the Internet of Things indicates the relevance of adversarial approaches in the modern landscape of interconnected devices. The intricate weave of terms such as "white box," "black-box attacks," "reinforcement learning," and "Bayesian networks," among others, portray the rich diversity and depth of discussions around adversarial machine learning across various domains and methodologies.

Figure 6. Co-occurrence of keywords.

4.8 Co-Authorship between Countries

The visualization in Figure 7 provides a vivid representation of country co-authorship patterns, offering a snapshot of international collaborations in academic or research publications. The "United States" stands out as a central node, indicating its robust involvement in collaborative research efforts with numerous countries. Adjacently, "China" showcases significant collaborative tendencies, particularly with neighboring regions such as "Japan," "Hong Kong," and "Singapore," among others.

European nations manifest a dense web of interconnectedness, emphasizing the close-knit research partnerships within the continent. "Germany," "United Kingdom," "Italy," and "France" are some of the prominent actors in this cluster, suggesting their pivotal roles in fostering European research alliances. "Poland" and "Switzerland" are also noticeable, albeit with slightly lesser interconnections.

The Asian landscape, beyond China, showcases countries like "India," "Israel," and "Vietnam actively engaging in collaborative research endeavors. "Australia" adjacent to Asian and European countries signifies its strategic position as a collaborative bridge between the two continents.

In the Americas, "Canada" and "Brazil" emerged as significant contributors, indicating their active roles in transcontinental research. Middle Eastern countries such as "Saudi Arabia" and "Qatar" are also visible, highlighting their growing involvement in global research. Overall, this visualization paints a picture of an international research community that is interconnected and collaborative, transcending geographical boundaries to produce collective knowledge.

Figure 7. Country co-authorship analysis

5. DISCUSSION

The trajectory of scientific production in adversarial machine learning, as observed from 2011 to 2023, provides a compelling narrative of its growing significance in academic circles. The initial phase (2011-2015) witnessed a subdued rate of publications, indicating the nascent stage of this domain. However, post-2016, there's an undeniable surge in research interest. This could be attributed to the broader acceptance and implementation of machine learning techniques, necessitating studies on its vulnerabilities.

Prominent authors, highlighted by publication count, signify the leaders driving the field's evolution. Their contributions offer a roadmap to the domain's development, setting quality and innovative thinking benchmarks. The thematic map underscores the multifaceted nature of research within adversarial machine learning. While niche themes delve into specifics, motor themes highlight the core areas of interest. Observing these themes can assist researchers in identifying gaps and emerging areas that might be the focal points of future studies. The high citation counts of specific papers, like "One Pixel Attack for Fooling Deep Neural Networks," accentuate the critical role such works play in shaping discourse, guiding research directions, and offering breakthroughs. These papers not only contribute knowledge but also inspire subsequent research endeavors.

The analysis of trend topics presents a comprehensive trajectory of pivotal areas in adversarial machine learning from 2011 to 2023. A consistent interest in "Machine Learning" is evident, with a notable surge in "Deep Learning" around 2016-2017, reflecting its broader academic and industrial embrace. The increasing attention towards terms like "Adversarial Machine Learning" and associated adversarial terminologies in recent years signifies the growing concerns and research fervour surrounding adversarial threats in the AI domain. Security-focused terms further accentuate the urgency to fortify machine learning systems. The emergence of specific adversarial tactics, such as "Poisoning Attacks" and "Exploratory Attacks," indicates evolving challenges, while the rise of "Game Theory" suggests its potential utility in modeling these adversarial dynamics. This analysis encapsulates the evolving priorities and challenges in the adversarial machine learning domain over the past decade, spotlighting areas of heightened interest and concern.

Lastly, the visualization of keywords and country co-authorship patterns emphasizes the interdisciplinary nature of adversarial machine learning and the global collaborative spirit driving its advancements. The integration of terms like "IoT" and "computer vision" with "adversarial machine learning" suggests the domain's extensive reach. The dense web of country collaborations highlights the universality of the challenges posed by adversarial attacks and the collective efforts to address them. In essence, the domain of adversarial machine learning is dynamic and expansive and holds immense potential for future explorations, innovations, and collaborations.

The study offers invaluable insights that have wide-ranging practical implications. It highlights emerging trends for academics and industry professionals, guiding future research and product development. Educational institutions can refine curriculums to stay abreast of the latest methodologies. Governments and regulatory bodies can craft informed policies to address potential threats, while businesses can align R&D strategies with cutting-edge advancements. Furthermore, the insights foster collaboration, ethical application, and public awareness, ensuring that adversarial machine learning is harnessed responsibly and effectively across various sectors.

6. LIMITATIONS

This study, centered on the Scopus database, potentially overlooks pertinent research available in other databases, leading to a possible selection bias. Focusing exclusively on journal articles and conference papers, we may have missed valuable insights from other document types like books, theses, or technical reports.

7. CONCLUSION

The bibliometric analysis of adversarial machine learning research has provided a thorough understanding of the field's evolution, highlighting its rapid growth, diverse influences, and innovative developments. The notable surge of papers in the past few years highlights the increasing recognition of adversarial risks and the critical requirement for robust defense mechanisms in machine learning systems. The study highlighted important papers and innovative scholars whose contributions have been crucial in shaping the field's course. Furthermore, our analysis revealed a vibrant and highly networked research community, emphasizing the interdisciplinary nature of adversarial machine learning across computer science, cybersecurity, artificial intelligence, and cognitive science disciplines. Our work provides a roadmap for navigating the vast and complex adversarial machine learning research field. Understanding our history helps us better prepare for the future by anticipating obstacles and ensuring that reliability and resilience are given equal weight as machine learning advances.

REFERENCES

[1] P. Ambika, “Machine learning,” Handbook of Research on Cloud and Fog Computing Infrastructures for Data Science. 2018.

[2] A. Ławrynowicz and V. Tresp, “Introducing machine learning,” Perspectives on Ontology Learning, vol. 18. 2014.

[3] J. D. Tygar, "Adversarial Machine Learning," IEEE Internet Comput., vol. 15, no. 5, pp. 4–6, Sep. 2011,

[4] Y. Vorobeychik, M. Kantarcioglu, R. Brachman, P. Stone, and F. Rossi, Adversarial machine learning, vol. 12. Springer, 2018.

[5] K. Wang, "Adversarial Machine Learning with Double Oracle," in Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, Macao, China: International Joint Conferences on Artificial Intelligence Organization, Aug. 2019, pp. 6472–6473. doi: 10.24963/ijcai.2019/925.

[6] C. J. Hernández-Castro, Z. Liu, A. Serban, I. Tsingenopoulos, and W. Joosen, "Adversarial Machine Learning," in Security and Artificial Intelligence: A Crossdisciplinary Approach, L. Batina, T. Bäck, I. Buhan, and S. Picek, Eds., Cham: Springer International Publishing, 2022, pp. 287–312. doi: 10.1007/978-3-030-98795-4_12.

[7] N. Martins, J. M. Cruz, T. Cruz, and P. Henriques Abreu, "Adversarial Machine Learning Applied to Intrusion and Malware Scenarios: A Systematic Review," IEEE Access, vol. 8, pp. 35403–35419, 2020, doi: 10.1109/ACCESS.2020.2974752.

[8] A. Kumar, S. Mehta, and D. Vijaykeerthy, "An Introduction to Adversarial Machine Learning," in Big Data Analytics, P. K. Reddy, A. Sureka, S. Chakravarthy, and S. Bhalla, Eds., Cham: Springer International Publishing, 2017, pp. 293–299.

[9] V. Duddu, "A Survey of Adversarial Machine Learning in Cyber Warfare," Def. Sci. J., vol. 68, no. 4, p. 356, Jun. 2018, doi: 10.14429/dsj.68.12371.

[10] A. H. Alsharif, N. Z. Salleh, and R. Baharun, "Research Trends of Neuromarketing: A Bibliometric Analysis," J. Theor. Appl. Inf. Technol., vol. 98, no. 15, pp. 2948–2962, 2005.

[11] O. Ellegaard and J. A. Wallin, "The bibliometric analysis of scholarly production: How great is the impact?," Scientometrics, vol. 105, no. 3, pp. 1809–1831, Dec. 2015, doi: 10.1007/s11192-015-1645-z.

[12] W. W. Hood and C. S. Wilson, "The Literature of Bibliometrics, Scientometrics, and Informetrics," 2001.

[13] L. Waltman, N. J. van Eck, and E. C. M. Noyons, "A unified approach to mapping and clustering of bibliometric networks," J. Informetr., vol. 4, no. 4, pp. 629–635, Oct. 2010, doi: 10.1016/j.joi.2010.07.002.

[14] F. J. Agbo, S. S. Oyelere, J. Suhonen, and M. Tukiainen, "Scientific production and thematic breakthroughs in smart learning environments: a bibliometric analysis," Smart Learn. Environ., vol. 8, no. 1, p. 1, Dec. 2021, doi: 10.1186/s40561-020-00145-4.

[15] S. Babu and B. Thomas, "Bibliometric Analysis and Visualization of Scientific Literature on Random Forest Regression," Sci. Vis., vol. 14, no. 5, 2022.

[16] M. E. Bales, D. N. Wright, P. R. Oxley, and T. R. Wheeler, "Bibliometric visualization and analysis software: State of the art, workflows, and best practices," 2020.

[17] T. A. FAUZAN and E. S. SOEGOTO, "COMPUTATIONAL BIBLIOMETRIC ANALYSIS OF EDUCATION TECHNOLOGY USING VOSVIEWER APPLICATION WITH PUBLISH OR PERISH (USING GOOGLE SCHOLAR DATA)," J. Eng. Sci. Technol., vol. 18, no. 3, pp. 1498–1508, 2023.

[18] R. Jumansyah, E. S. Soegoto, and C. N. Albar, "COMPUTATIONAL BIBLIOMETRIC ANALYSIS OF EVOLUTIONARY GAME THEORY (EGT) RESEARCH USING VOSVIEWER," vol. 18, 2023.

[19] R. Maryanti, A. B. D. Nandiyanto, A. Hufad, S. Sunardi, D. Al Husaeni, and D. Al Husaeni, "A computational bibliometric analysis of science education research using vosviewer," J. Eng. Sci. Technol., vol. 18, no. 1, pp. 301–309, 2023.

[20] N. J. Van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," Scientometrics, vol. 84, no. 2, pp. 523–538, Aug. 2010, doi: 10.1007/s11192-009-0146-3.

[21] A. B. D. Nandiyanto and D. F. Al Husaeni, "Bibliometric analysis of engineering research using vosviewer indexed by google scholar," J. Eng. Sci. Technol., vol. 17, no. 2, pp. 883–894, 2022.

[22] Y. Yu et al., "A bibliometric analysis using VOSviewer of publications on COVID-19," Ann. Transl. Med., vol. 8, no. 13, pp. 816–816, Jul. 2020, doi: 10.21037/atm-20-4235.

[23] J. S. Racine, "RStudio: a platform-independent IDE for R and Sweave." JSTOR, 2012.

[24] N. Salim, K. Gopal, and A. Ayub, "Effects of using RStudio on statistics performance of

Malaysian undergraduates," Malays. J. Math. Sci., vol. 13, no. 3, pp. 419–437, 2019.

[25] E. Herrera-Viedma, A. Santisteban-Espejo, M. J. Cobo, and others, "Software tools for conducting bibliometric analysis in science: An up-to-date review," Prof. Inf., vol. 29, no. 1, 2020.

[26] A. Nasir, K. Shaukat, I. A. Hameed, S. Luo, T. M. Alam, and F. Iqbal, "A bibliometric analysis of corona pandemic in social sciences: a review of influential aspects and conceptual structure," Ieee Access, vol. 8, pp. 133377–133402, 2020.

[27] A. McCarthy, E. Ghadafi, P. Andriotis, and P. Legg, "Functionality-Preserving Adversarial Machine Learning for Robust Classification in Cybersecurity and Intrusion Detection Domains: A Survey," Journal of Cybersecurity and Privacy, vol. 2, no. 1, pp. 154–190, 2022, doi: 10.3390/jcp2010010.

[28] M. Omar, S. Choi, D. Nyang, and D. Mohaisen, "Quantifying the Performance of Adversarial Training on Language Models with Distribution Shifts," in CySSS '22, New York, NY, USA: Association for Computing Machinery, 2022, pp. 3–9. doi: 10.1145/3494108.3522764.

[29] S. A. Seshia, S. Jha, and T. Dreossi, "Semantic Adversarial Deep Learning," IEEE Design & Test, vol. 37, no. 2, pp. 8–18, Apr. 2020, doi: 10.1109/MDAT.2020.2968274.

[30] J. D. Tygar, "Adversarial Machine Learning," IEEE Internet Computing, vol. 15, no. 5, pp. 4–6, Oct. 2011, doi: 10.1109/MIC.2011.112.

[31] Y. Zhou, M. Kantarcioglu, and B. Xi, "A survey of game theoretic approach for adversarial machine learning," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 9, no. 3, p. e1259, 2019, doi: 10.1002/widm.1259.

[32] K. Aryal, M. Gupta, and M. Abdelsalam, "A survey on adversarial attacks for malware analysis," arXiv preprint arXiv:2111. 08223, 2021.

[33] Center for Security and Emerging Technology, T. Rudner, and H. Toner, "Key Concepts in AI Safety: Robustness and Adversarial Examples," Center for Security and Emerging Technology, Mar. 2021. doi: 10.51593/20190041.

Table 1. Top ten cited papers

Authors	**Title**	**Year**	**Cited by**
Su J.; Vargas D.V.; Sakurai K.	One Pixel Attack for Fooling Deep Neural Networks	2019	1113
Biggio B.; Corona I.; Maiorca D.; Nelson B.; Šrndić N.; Laskov P.; Giacinto G.; Roli F.	Evasion attacks against machine learning at test time	2013	991
Kurakin A.; Goodfellow I.J.; Bengio S.	Adversarial machine learning at scale	2017	710
Biggio B.; Roli F.	Wild patterns: Ten years after the rise of adversarial machine learning	2018	697
Huang L.; Joseph A.D.; Nelson B.; Rubinstein B.I.P.; Tygar J.D.	Adversarial machine learning	2011	678
Zügner D.; Akbarnejad A.; Günnemann S.	Adversarial attacks on neural networks for graph data	2018	471
Jagielski M.; Oprea A.; Biggio B.; Liu C.; Nita-Rotaru C.; Li B.	Manipulating Machine Learning: Poisoning Attacks and Countermeasures for Regression Learning	2018	394
Muñoz-González L.; Biggio B.; Demontis A.; Paudice A.; Wongrassamee V.; Lupu E.C.; Roli F.	Towards poisoning of deep learning algorithms with back-gradient optimization	2017	298
Chen P.-Y.; Sharma Y.; Zhang H.; Yi J.; Hsieh C.-J.	EAD: Elastic-net attacks to deep neural networks via adversarial examples	2018	273
Wang Z.; Wang J.; Wang Y.	An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition	2018	258

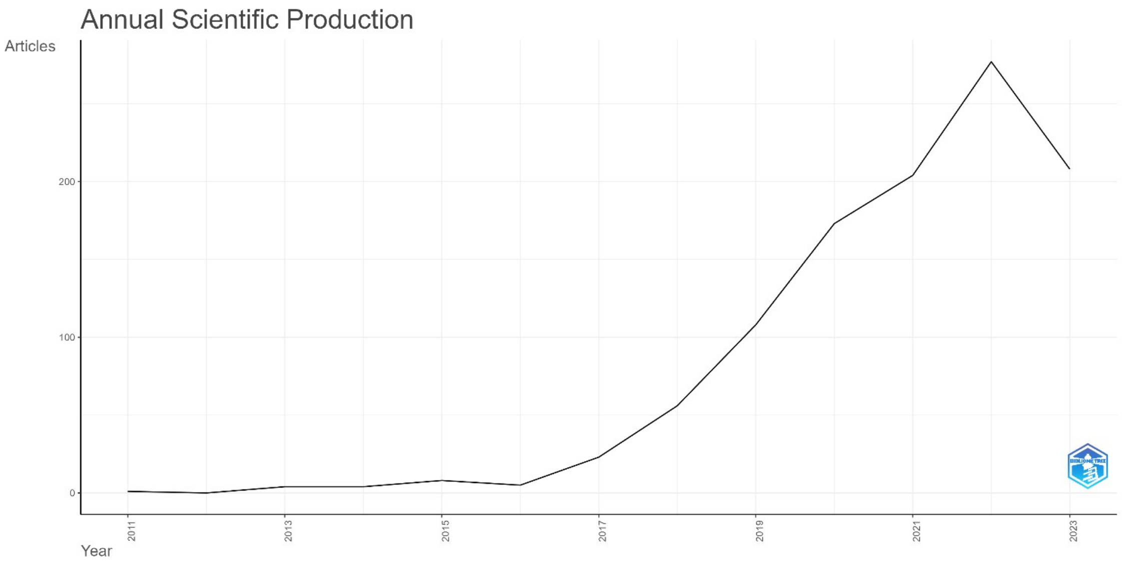

Figure 1. Annual scientific production.

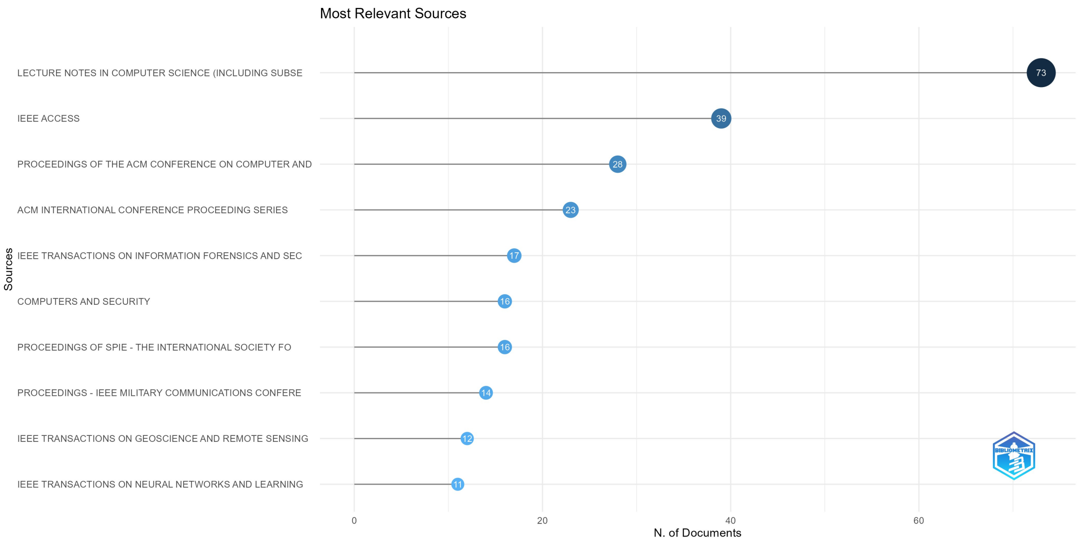

Figure 2. The top ten sources.

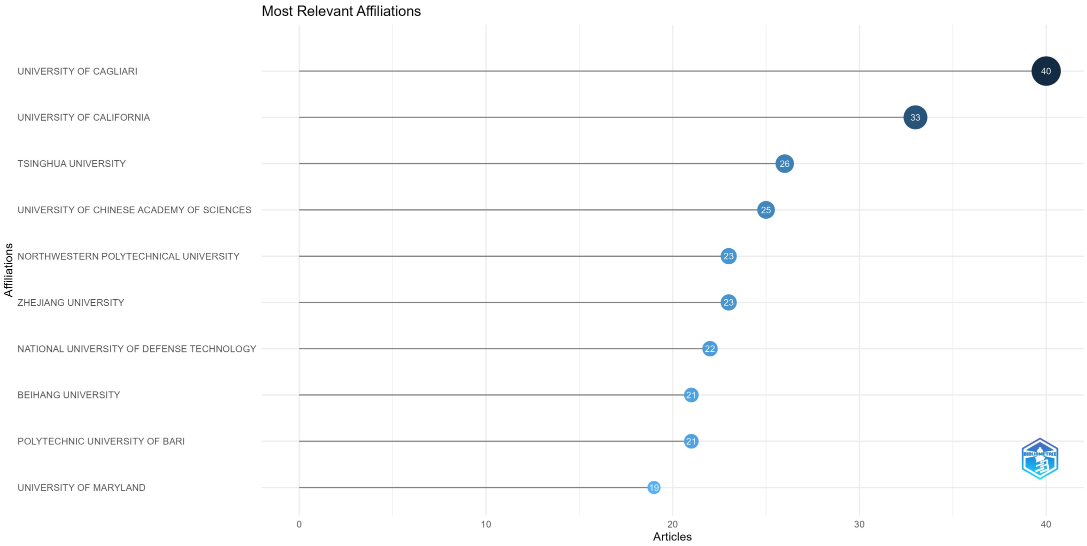

Figure 3. Most relevant affiliations.

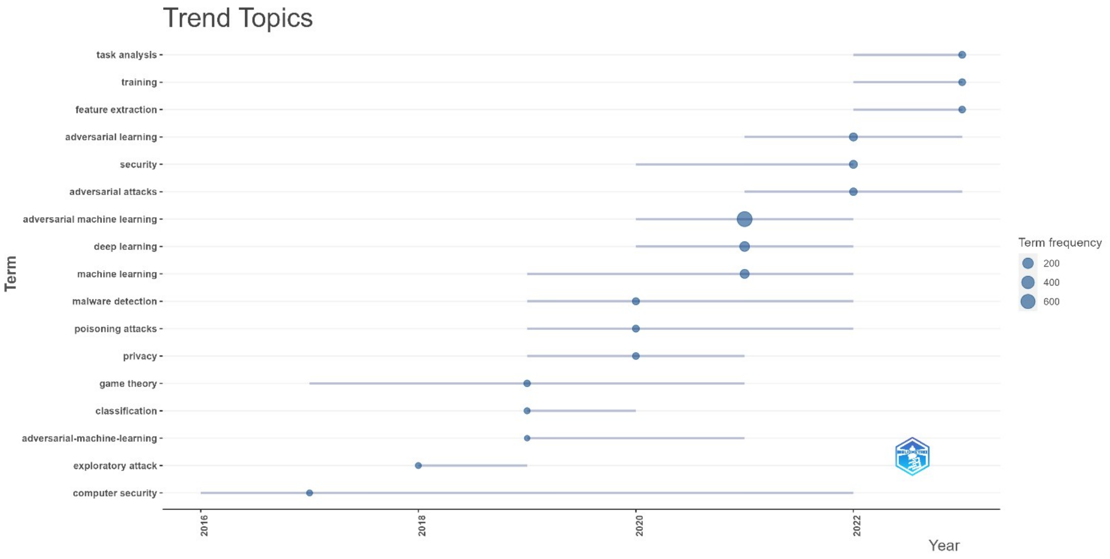

Figure 4. Trend Topics

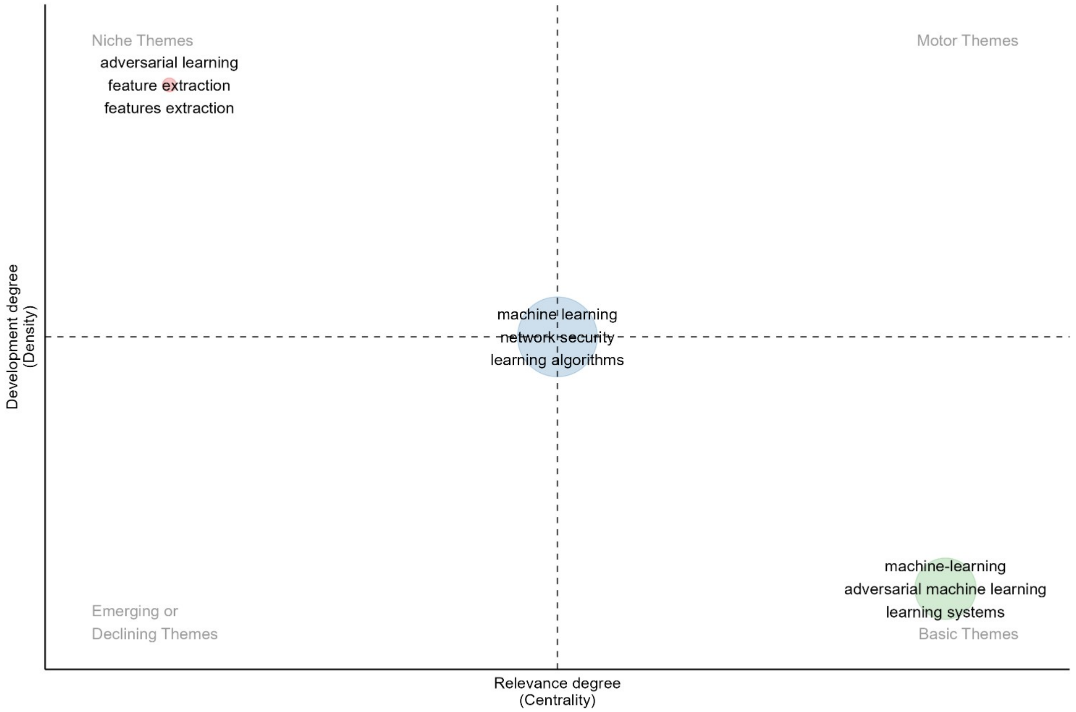

Figure 5. Thematic Map.

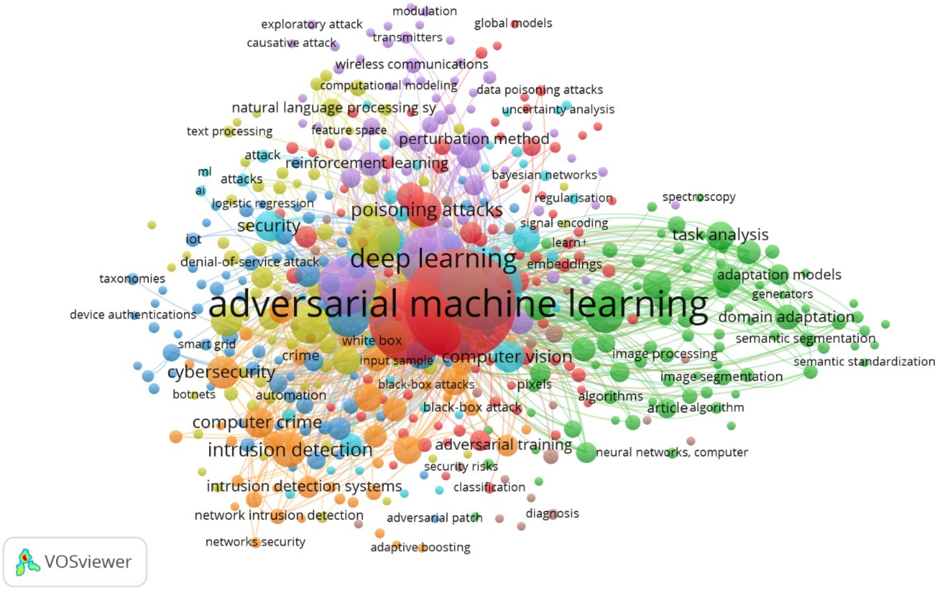

Figure 6. Co-occurrence of keywords

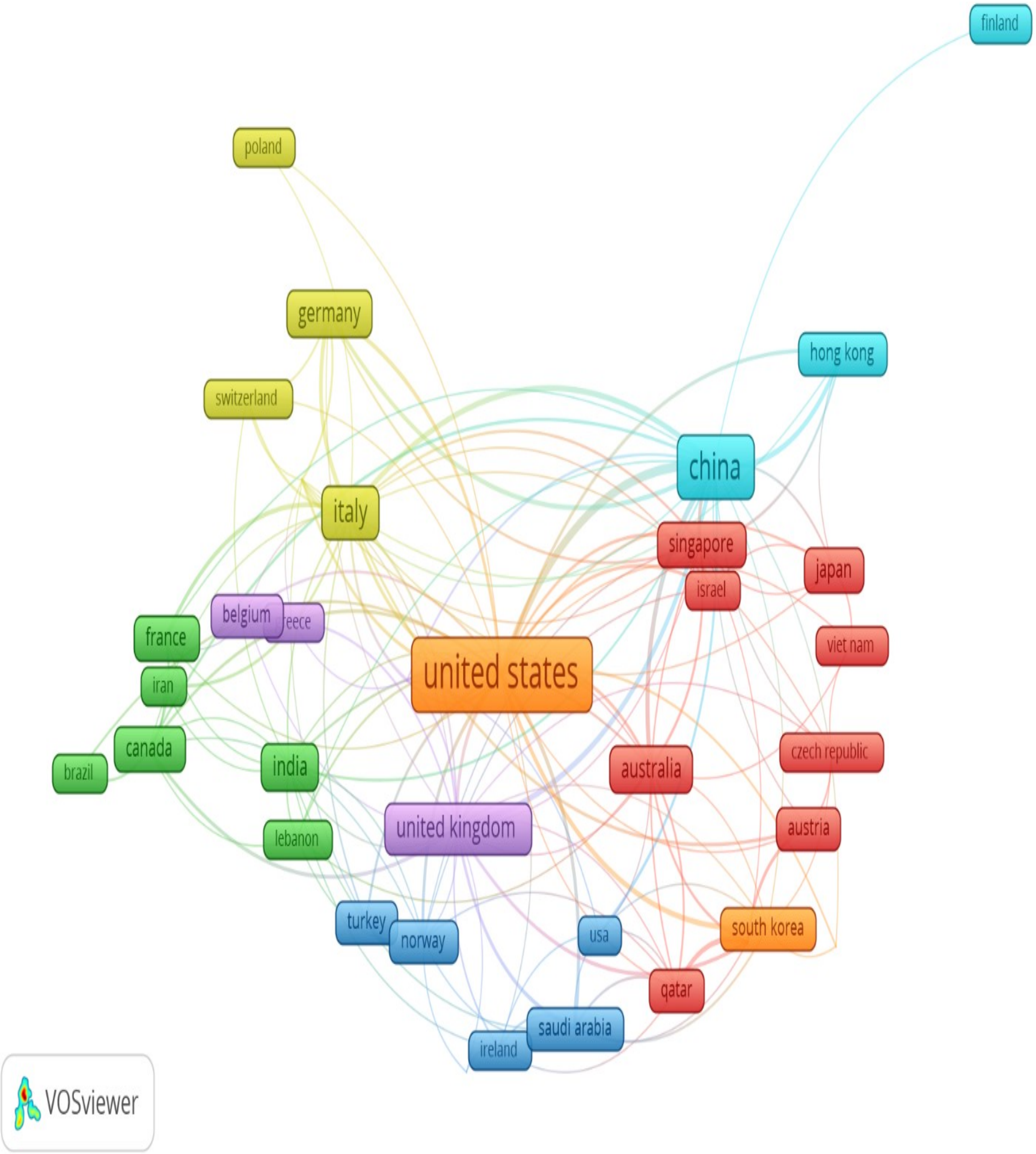

Figure 7. Country co-authorship analysis.