Mesopotamian journal of Big Data
Vol. **(2023)**, 2023, **pp**. 138–148
DOI: https://doi.org/10.58496/MJBD/2023/018 ISSN: 2958-6453
https://mesopotamian.press/journals/index.php/BigData

Research Article

Mapping the Evolution of Intrusion Detection in Big Data: A Bibliometric Analysis

Mohanad G. Yaseen[1,*], A. S. Albahri[2]

[1] *Department of Computer, College of Education, Aliraqia University, Baghdad, Iraq*

[2] *Faculty of Computing and Meta-Technology (FKMT), Universiti Pendidikan Sultan Idris (UPSI), Perak, Malaysia*

ARTICLE INFO

Article History
Received 13 Sep 2023
Accepted 19 Nov 2023
Published 03 Dec 2023

Keywords
Intrusion Detection
Big Data Security
Bibliometric Analysis
Cybersecurity Trends

ABSTRACT

This study provides a comprehensive analysis of the dynamic amalgamation of intrusion detection and big data, revealing trends and patterns within cybersecurity research. The investigation reveals a notable surge in scholarly output from 2018 onwards, reflecting heightened interest and exploration within the field. Dominant themes such as "intrusion detection," "big data," and "machine learning" underscore the integration of security concerns with advanced technologies. Geographical influences showcase diverse contributions, with varying citation impacts from countries like India, China, and Saudi Arabia. Author contributions reveal a balance between prolific authors and impactful contributions from authors with fewer publications. Recommendations include fostering interdisciplinary collaborations, integrating advanced computational methods, and conducting longitudinal studies to gauge sustained impacts. This research underscores collaboration dynamics, thematic evolution, and global influences as pivotal facets within the realm of intrusion detection and big data, guiding future research to fortify digital security in an ever-evolving technological landscape.

1. INTRODUCTION

In our contemporary digital era, the proliferation of data has transformed the way information is generated, processed, and utilized across various domains. Big data, a term encompassing vast and complex datasets, has emerged as a pivotal asset driving innovation and insights across industries and academic disciplines[1]. This paradigm shift in data management and analysis has not only revolutionized technological landscapes but has also posed significant challenges in ensuring the security and integrity of these voluminous datasets. Concurrently, the realm of intrusion detection, an essential facet of cybersecurity, has been tasked with safeguarding systems against unauthorized access, malicious activities, and cyber threats. As big data continues to proliferate, its intersection with intrusion detection has become increasingly pivotal, necessitating advanced methodologies and innovative approaches to ensure the resilience of digital infrastructures[2].

This study hypothesizes that the exponential growth in the volume and complexity of big data has influenced a surge in research activities and technological advancements within the field of intrusion detection[3]. Furthermore, it postulates that the collaborative integration of big data analytics and intrusion detection systems has catalyzed innovative solutions to address evolving cybersecurity challenges. This research aims to elucidate the evolving landscape within the intersection of intrusion detection and big data through a comprehensive analysis of scholarly output trends, thematic emphasis, author contributions, geographical influences, and collaboration patterns.

The subsequent sections will delve into an exploration of scholarly output trends over the past decade, a thematic analysis highlighting prevalent research topics, an examination of geographical influences on contributions, an assessment of author contributions and impact, and an analysis of collaborative networks among authors. Finally, recommendations and future directions for research within this dynamic domain will be proposed based on the revealed insights from the analysis.

2. METHODOLOGY

For this bibliometric analysis, the Scopus database was selected as the primary source of scholarly information due to its comprehensive coverage of academic literature in various fields. Scopus was chosen for its extensive indexing of high-

*Corresponding author. Email: maymy832410@gmail.com

quality research articles, conference papers, and other scholarly documents related to intrusion detection and big data. Its wide-ranging coverage and indexing policies ensure a diverse and relevant collection of documents.

1. Search Strategy and Keywords: The search query employed for document retrieval in Scopus utilized the Boolean operator "AND" to combine the keywords "Intrusion Detection" and "Big Data". These keywords were specifically targeted within the article titles to ensure a focused retrieval of relevant publications related to the intersection of intrusion detection systems and big data analytics.
2. Document Collection and Filtering: A total of 101 documents were obtained using the search query from Scopus. Upon retrieval, the documents were subjected to initial screening and filtering based on relevance to the study's focus on intrusion detection within the realm of big data. All retrieved documents were included in the analysis to ensure a comprehensive review of the available literature in this domain.
3. Data Extraction and Analysis: To extract bibliographic information, RStudio, a powerful integrated development environment (IDE), was utilized along with the R programming language. The 'biblioshiny' package was employed for data extraction, including figures and tables generated for analysis purposes.
4. Assessment of Bibliographic Metadata Completeness: A comprehensive evaluation was conducted to assess the completeness of bibliographic metadata across the 101 retrieved documents. Table 1 outlines the status of various metadata elements, highlighting their presence or absence within the dataset. While certain metadata elements, such as affiliation, author information, document type, journal details, publication year, and titles, exhibited high completeness (100%), some elements, including abstracts, DOIs, keywords, author correspondence, and cited references, demonstrated varying degrees of completeness, ranging from good to acceptable levels. Special attention was given to manage the issues with missing metadata to ensure the integrity and accuracy of the analysis.
5. Handling Missing Metadata: Given the issues encountered with missing metadata elements, the analysis did not solely rely on the completeness of bibliographic metadata. Instead, multiple validation methods and cross-referencing techniques were employed to mitigate the impact of missing data on the research findings. Efforts were made to triangulate information from available sources and ensure the reliability and validity of the bibliometric analysis despite incomplete metadata. This methodology was designed to facilitate a comprehensive assessment of the scholarly landscape concerning intrusion detection in the context of big data, despite challenges related to missing bibliographic metadata. The adopted approach aimed to provide meaningful insights and trends in this field based on the available dataset obtained from Scopus.

3. RESULTS

3.1 Annual Scientific Production

The annual scientific production in the field exhibited intriguing trends and patterns over the past decade. The progression of scholarly output, as reflected in Figure 1, signifies a notable trajectory in research activity. Commencing at a modest rate with 3 articles in 2014, there was a gradual increase in scientific production until 2018, where a significant leap occurred with 12 articles. This surge in publications continued to climb in subsequent years, marking a substantial growth curve. Notably, 2022 and 2023 witnessed a remarkable spike in scientific output, reaching 21 and 24 articles, respectively.

The progression observed in the annual scientific production portrays an upward trend in research activity, indicative of the field's burgeoning interest and developments. The consistent rise from 2014 to 2023 implies a heightened focus on scholarly endeavors, possibly influenced by the evolving landscape of the field and the emerging areas of interest. The spike in 2018 could suggest a pivotal juncture where the field experienced a surge in interest, leading to intensified research initiatives and contributions. Furthermore, the exponential growth from 2020 onwards signifies an accelerated pace of research output, potentially fueled by technological advancements, global collaborations, or paradigm shifts within the scientific community. The correlation between the year-wise article count and the evolutionary course of the field underscores an evident surge in scholarly activity.

This surge may reflect the growing significance of the subject matter, expanding research inquiries, or intensified efforts to bridge gaps in knowledge. The escalating trend from 2018 onwards signals an upward trajectory in the field's scientific output, suggesting a vibrant and dynamic landscape conducive to continual exploration and innovation. As such, the annual scientific production provides a lens into the progressive evolution of the field, marking a phase of increased scholarly attention and engagement.

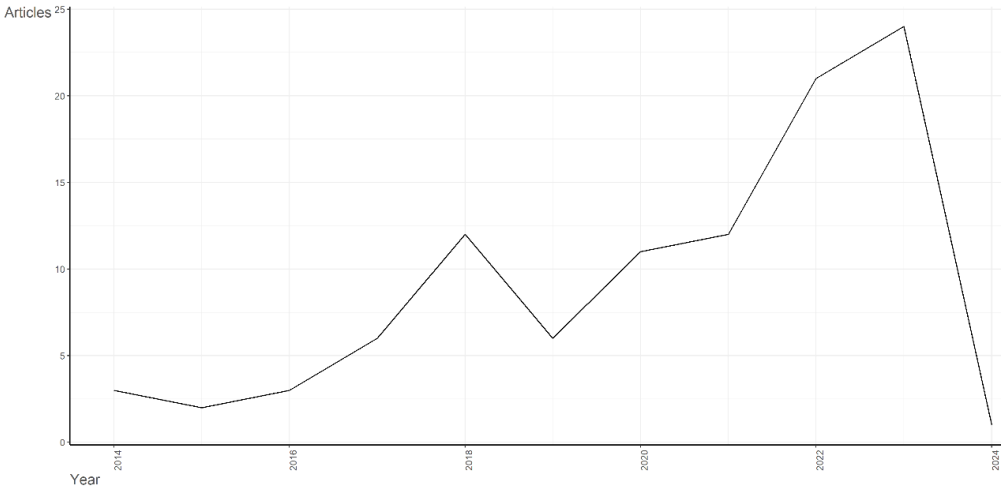

Fig. 1. Annual Scientific Production

3.1 Average Citations per Year

The average citations per year provide insights into the impact and reception of scholarly articles within the field over the past decade, as depicted in Figure 2. The trend in mean citations per article fluctuates considerably across the years. Commencing in 2014 with an average of 16.67 citations per article, the subsequent year, 2015, displayed a notable decline, with a mean of 0.5 citations per article, possibly indicating a temporary drop in the articles' impact. However, this figure is grounded in a limited dataset, with only 2 articles, resulting in an overall low mean citation count. In the following years, the mean citation count underwent variations, revealing fluctuations in the articles' impact. Notably, 2016 demonstrated a considerable surge in the mean citations per article, reaching 36.67, reflective of the higher impact and wider acceptance of the publications. This trend continued through 2017 with an average of 26 citations per article, sustaining a relatively higher level of impact.

However, a striking change occurred in 2018, showcasing a substantial increase to 37.67 citations per article, marking a peak in the average impact. Subsequently, in 2019 and 2020, the mean citation count remained consistently high, hovering around 40 citations per article, suggesting a sustained period of impactful contributions within the field. The following years experienced a decline in the mean citations per article, with 2023 indicating a significant decrease to 0.75 citations per article. This decline may be attributed to the limited time frame since publication, implying that articles from this year might not have had adequate time to accumulate citations. Overall, the mean citations per year fluctuated significantly across the analyzed period, indicating varying degrees of influence and impact of scholarly articles within the field. The variations observed highlight the dynamic nature of research impact, influenced by numerous factors such as article novelty, relevance, and visibility, thereby shaping the scholarly landscape within the field.

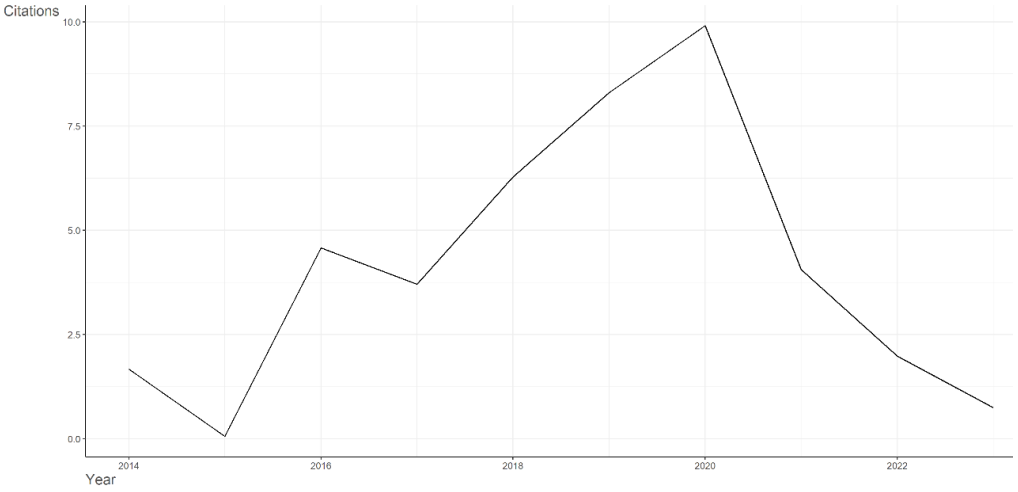

Fig. 2. Average Citations per Year

3.2 Most Relevant Authors

The analysis of the most relevant authors within the field, as illustrated in Figure 3, sheds light on the key contributors and their respective article contributions. The authors with the highest number of articles published possess varying degrees of prominence and influence within the domain.

At the forefront, Jemili F[1] emerges as the most prolific author, with a significant contribution of 10 articles. This substantial contribution signifies their extensive involvement and impact on the scholarly landscape. Following closely behind, Korbaa [2] and Wang [3] exhibit notable authorship, each with 4 articles. Despite slightly fewer publications, these authors maintain a considerable presence and influence within the field, indicating consistent and impactful contributions to the literature.

Further down the list, several authors, such as Abid [4], Leung [5], and Peng [6], among others, exhibit a contribution of 3 articles each. While their contributions are marginally lower than the top authors, their consistent involvement showcases a substantial presence and influence in the scholarly discourse.

Additionally, authors like Ahmad [7], Ajabi [8], and Akhtar [9] have contributed 2 articles each. Although fewer in number, their publications demonstrate a noteworthy presence, contributing significantly to the body of knowledge within the domain.

The fractionalized representation of article contributions emphasizes the proportional impact of authors within the analyzed period. Jemili [1] retains the highest fractionalized contribution of 5.08, underlining their substantial influence and extensive involvement in the scholarly output. Authors such as Wang Y and Abid [2] maintain fractionalized contributions of 1.53 and 1.17, respectively, reaffirming their substantive impact on the field's literature.

This comprehensive overview of the most relevant authors provides a nuanced perspective on their respective contributions and influence within the field, offering valuable insights into the scholarly landscape.

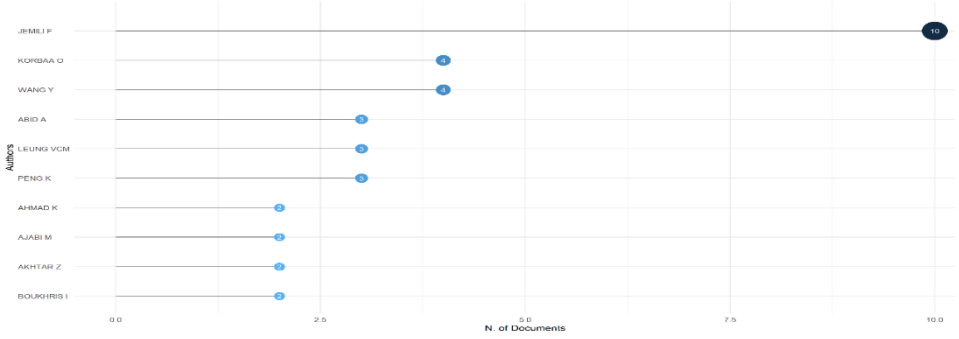

Fig. 3. Most Relevant Authors and Their Article Contributions

3.3 Corresponding Author's Countries

Analyzing the distribution of corresponding authors' countries, as depicted in Figure 4, provides insights into the geographical origins of contributions within the scholarly publications. The countries from which the corresponding authors hail exhibit varying degrees of participation, shedding light on their involvement in the research landscape. India emerges as the leading contributor, serving as the corresponding author for 22 articles. Among these, 18 articles feature solely Indian contributors (SCP - Sole Corresponding Publications), emphasizing a significant reliance on local expertise for these publications. Moreover, 4 articles feature mixed collaboration with authors from other countries (MCP - Mixed Corresponding Publications), demonstrating a balanced international collaboration with a frequency of 0.218.

China follows closely behind, with corresponding authors leading 18 articles. Among these, 11 articles are purely China-led (SCP), highlighting a substantial reliance on internal expertise. However, Chinese authors demonstrate a higher inclination towards international collaboration, as evidenced by 7 articles featuring mixed authorship (MCP), presenting a higher MCP ratio of 0.389. Additionally, Tunisia showcases a noteworthy contribution, with corresponding authors leading 12 articles. The majority of these articles (11) are exclusively authored by Tunisian researchers (SCP), indicating a predominant reliance on local expertise. However, the MCP ratio of Tunisia stands at 0.083, depicting a relatively lower frequency of international collaborations among the publications led by Tunisian authors.

Furthermore, countries such as Korea, the USA, Morocco, Turkey, Brazil, and Algeria also contribute to the scholarly output with varying degrees of involvement in the corresponding authorship. While the SCP for these countries reflects predominantly internal collaborations, the MCP ratio illustrates the extent of their engagement in international research collaborations, ranging from 0.250 to 1.000. This analysis of corresponding author's countries demonstrates the varying degrees of international collaboration and the reliance on local expertise within the publications, offering valuable insights into the geographical landscape of research collaborations in this domain.

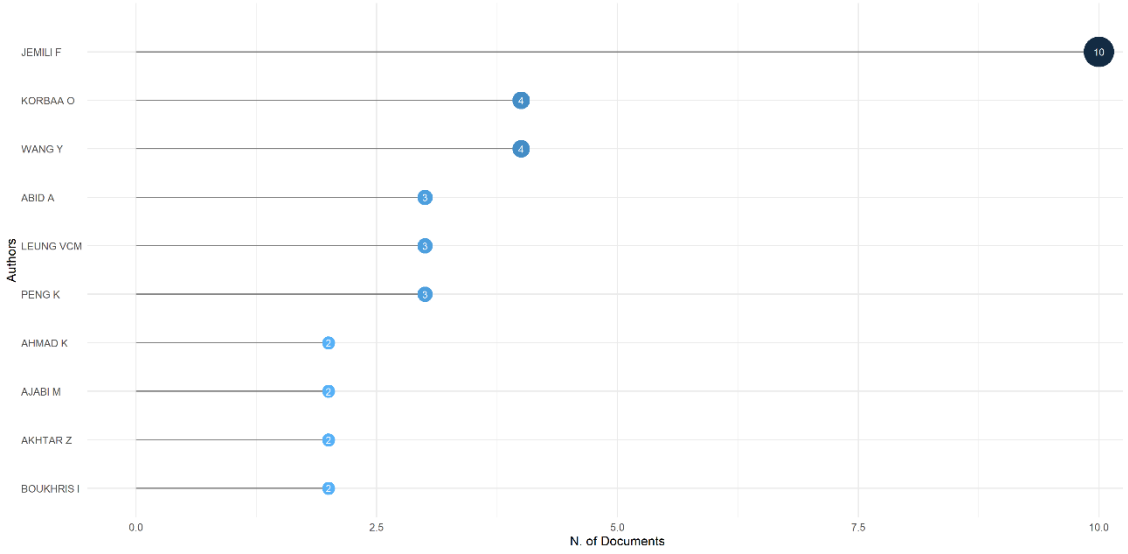

Fig. 4. Most Relevant Authors and Their Article Contributions

3.4 Abbreviations and Acronyms

Analyzing the distribution of corresponding authors' countries, as depicted in Figure 4, provides insights into the geographical origins of contributions within the scholarly publications. The countries from which the corresponding authors hail exhibit varying degrees of participation, shedding light on their involvement in the research landscape. India emerges as the leading contributor, serving as the corresponding author for 22 articles. Among these, 18 articles feature solely Indian contributors (SCP - Sole Corresponding Publications), emphasizing a significant reliance on local expertise for these publications. Moreover, 4 articles feature mixed collaboration with authors from other countries (MCP - Mixed Corresponding Publications), demonstrating a balanced international collaboration with a frequency of 0.218.

China follows closely behind, with corresponding authors leading 18 articles. Among these, 11 articles are purely China-led (SCP), highlighting a substantial reliance on internal expertise. However, Chinese authors demonstrate a higher inclination towards international collaboration, as evidenced by 7 articles featuring mixed authorship (MCP), presenting a higher MCP ratio of 0.389. Additionally, Tunisia showcases a noteworthy contribution, with corresponding authors leading 12 articles. The majority of these articles (11) are exclusively authored by Tunisian researchers (SCP), indicating a predominant reliance on local expertise. However, the MCP ratio of Tunisia stands at 0.083, depicting a relatively lower frequency of international collaborations among the publications led by Tunisian authors.

Furthermore, countries such as Korea, the USA, Morocco, Turkey, Brazil, and Algeria also contribute to the scholarly output with varying degrees of involvement in the corresponding authorship. While the SCP for these countries reflects predominantly internal collaborations, the MCP ratio illustrates the extent of their engagement in international research collaborations, ranging from 0.250 to 1.000. This analysis of corresponding author's countries demonstrates the varying degrees of international collaboration and the reliance on local expertise within the publications, offering valuable insights into the geographical landscape of research collaborations in this domain.

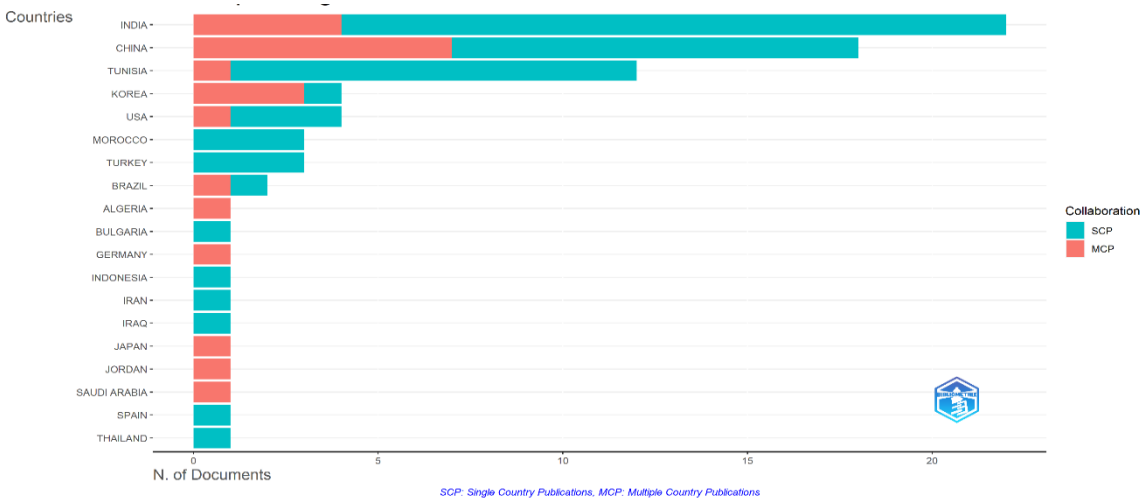

Fig. 5. Corresponding Author's Countries and Their Contributions

3.5 Most Cited Countries

The analysis of the most cited countries, illustrated in Figure 5, presents a panorama of their impact and influence within the scholarly sphere. The Total Citations (TC) and Average Article Citations offer a comprehensive overview of the prominence of these countries in terms of their research output and its impact. The United States (USA) stands out as a leading force in academic influence, accumulating a total of 212 citations across its articles, resulting in an impressive average of 53 citations per article. This high average signifies the substantial impact and recognition of research originating from the USA within the academic community.

Saudi Arabia follows closely, albeit with a smaller number of articles, but an equivalent impact, accruing a total of 191 citations across its articles, resulting in a remarkable average of 191 citations per article. This exceptional average suggests a concentrated influence and high citation rates for scholarly work hailing from Saudi Arabia. China, with a larger corpus of articles, accumulates a substantial total of 335 citations, averaging 18.60 citations per article. Despite a slightly lower average compared to the USA, China maintains a strong presence and influence within the academic sphere due to its high research output. Other notable contributors include Yemen, Korea, Japan, India, Tunisia, Turkey, and Brazil, showcasing varying levels of impact in terms of total citations and average citations per article. These countries exhibit diverse research landscapes, with differences in both the volume of scholarly output and its impact within the academic community. This analysis highlights the varying degrees of impact and influence exhibited by different countries based on their citation metrics, elucidating their significance in shaping scholarly discourse and knowledge dissemination within the field.

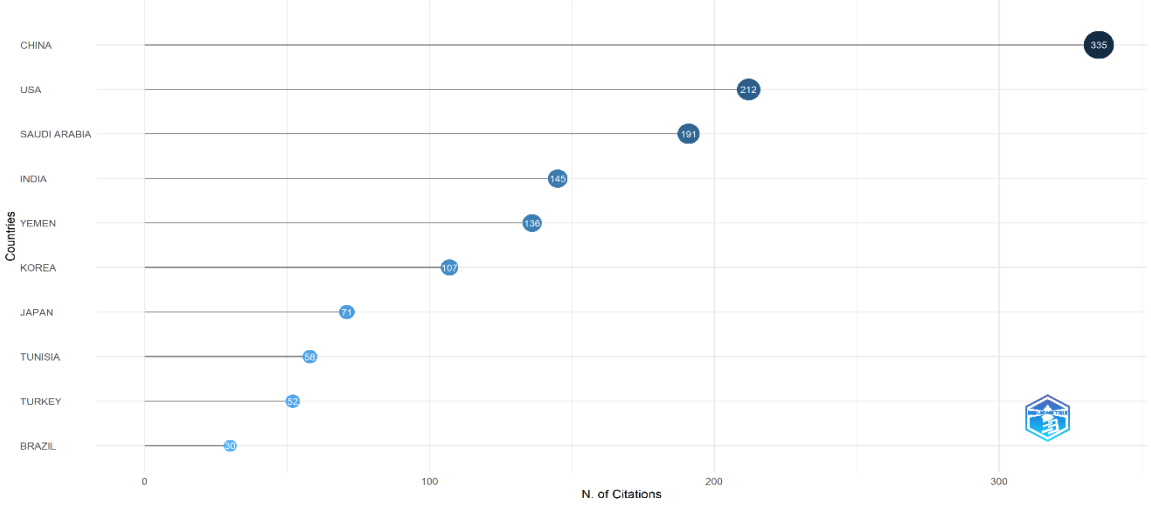

Fig.6. Most Cited Countries and Their Impact

3.6 Word Cloud

The Word Cloud depicted in Figure 6 offers a visual representation of the most frequently occurring terms extracted from research publications. The size of each term corresponds to its frequency within the dataset, providing insights into the prevalent themes and focuses within the field. Terms such as "intrusion detection," "big data," and "network security" emerge as dominant keywords, signifying their significance and prevalence in scholarly discourse. These terms denote the primary areas of interest and research focus within the domain, highlighting the emphasis on security and the utilization of large datasets in addressing challenges related to intrusion detection and network security.

The inclusion of terms like "deep learning," "machine learning," and "learning algorithms" underscores the growing interest and application of advanced computational methods, particularly within security-related domains. The prevalence of these terms signifies the evolving landscape of intrusion detection systems, leveraging cutting-edge technologies like deep learning and machine learning for enhanced efficacy. Moreover, phrases such as "cybersecurity," "data analytics," "cloud computing," and "internet of things" underscore the multidimensional nature of the field, emphasizing the integration of various technological aspects in addressing security challenges in modern digital environments. The presence of these terms signifies a comprehensive approach to tackle security issues, encompassing data analysis, cloud-based solutions, and emerging technologies like the Internet of Things (IoT).

The WordCloud offers a succinct representation of the dominant themes and key research areas within the field of intrusion detection and network security. It serves as a valuable visual tool to comprehend the prevalent research focus and the interplay of diverse concepts and technologies within the domain.

Fig. 7. Word Cloud of Key Terms

3.7 Co-occurrence Network

The Co-occurrence Network visualized in Figure 7 offers insights into the interconnectedness and relationships between significant terms prevalent in research publications. This network graphically represents the associations between terms based on their co-occurrence within scholarly literature.

The network reveals clusters of terms that frequently appear together within research articles related to intrusion detection and network security. Notably, terms like "intrusion detection," "big data," "intrusion detection systems," and "network security" emerge as central nodes within the network, exhibiting high betweenness and closeness centrality. These terms signify their critical roles and frequent co-occurrence within scholarly discourse, indicating their interdependence and relevance in the domain. Additionally, terms such as "computer crime," "deep learning," "data analytics," and "cybersecurity" form sub-clusters or satellite nodes within the network, showcasing their associations with the central themes. While these terms may not have the highest centrality measures, their presence highlights their contextual relevance and co-occurrence with the central nodes, emphasizing their significance in the broader discourse of intrusion detection and network security.

The network structure reflects the intricate interplay and interconnected nature of diverse concepts and technologies within the field. It visually depicts the relationships between key terms, facilitating an understanding of their associations and their collective influence on research in intrusion detection and network security. This visualization aids researchers in identifying prevalent themes and exploring the underlying patterns and connections among various concepts in the domain.

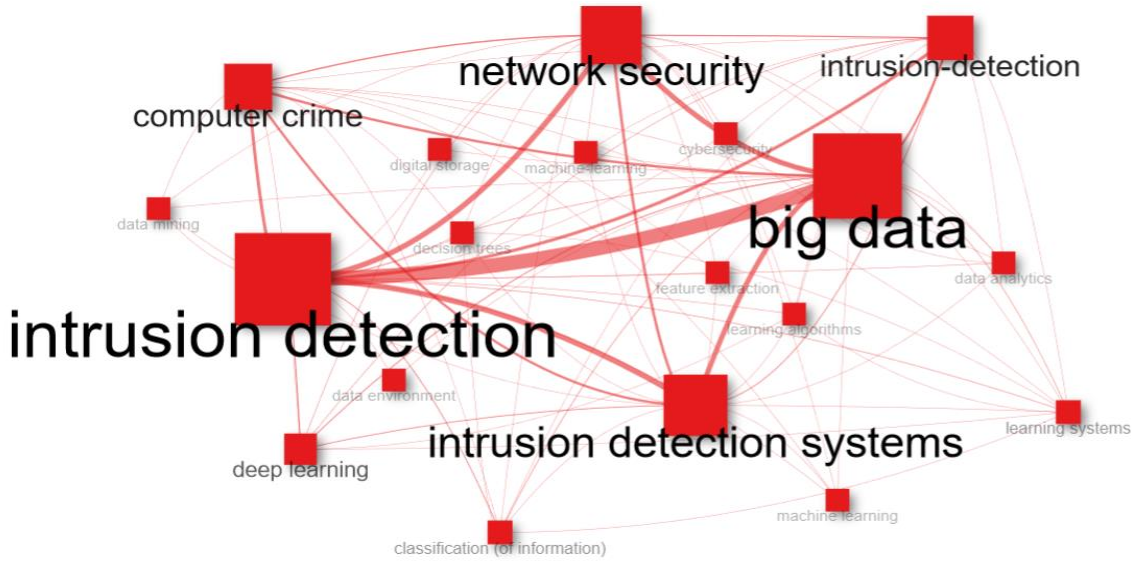

Fig. 8. Co-occurrence Network of Key Terms

3.8 Collaboration Network

The Collaboration Network displayed in Figure 8 illustrates the relationships and collaborations between authors within the field. This network graphically represents the co-authorship patterns and interactions based on their collaborations in research publications.

The network reveals distinct clusters of authors who have collaborated significantly in the field. Author nodes such as jemili [1], korbaa [2],abid [4] and meddeb [10] form central clusters, highlighting their collaborative efforts within their respective groups. These authors exhibit higher betweenness centrality, closeness centrality, and PageRank, indicating their substantial contributions and prominent positions in collaborative networks. Moreover, there are distinct author clusters, denoted as different clusters or subgroups within the network, reflecting different collaborative patterns among researchers. For instance, authors like li [11], meng [12], ahmad [13] and "wahid a" belong to separate clusters, indicating potential isolated or distinct collaborations within their groups.

Additionally, within the network, certain authors like "leung [5], peng [6], zheng [14], akhtar [9], kim [15], siddique [9], ajabi [8], boukhris [16] and elouedi [8] exhibit unique positions in their respective clusters, showcasing their collaborative connections with multiple groups or individuals. The Collaboration Network visualizes the intricate collaborations among authors, offering insights into their cooperative efforts within the domain. It delineates the collaborative patterns, cluster formations, and individual author roles within different collaborative networks, providing a comprehensive overview of the collaborative landscape in the field.

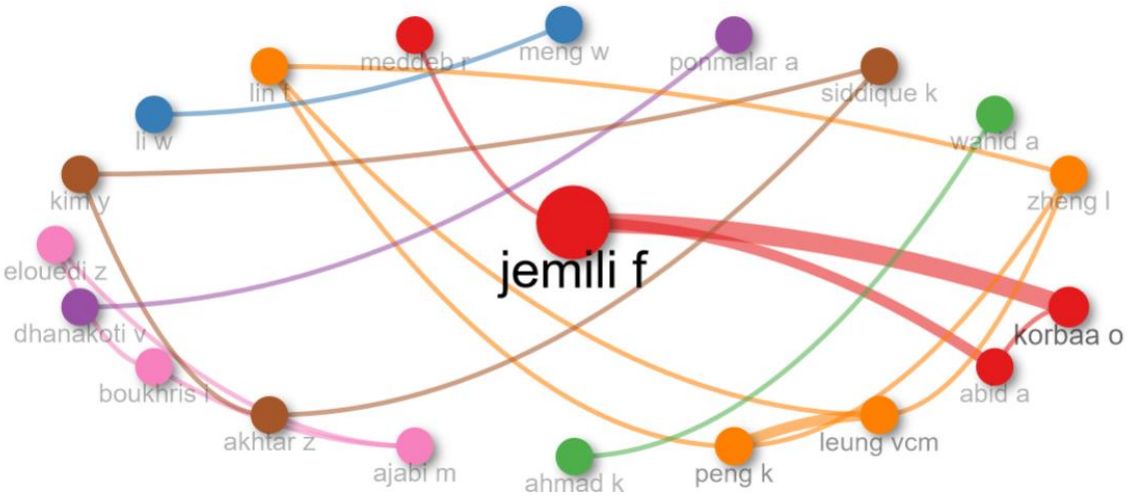

Fig. 9. Collaboration Network among Authors

4. DISCUSSION

The analysis of scholarly output trends depicted a notable trajectory in research activity within the field of intrusion detection and big data over the past decade. Commencing modestly with three articles in 2014, a gradual increase in scientific production ensued, peaking significantly in 2023 with 24 articles. This surge suggests a burgeoning interest and continuous developments within the field.

In conjunction with the annual scientific production, the mean citations per article fluctuated considerably. Notably, 2018 marked a peak in impact with an average of 37.67 citations per article, followed by consistent high averages in 2019 and 2020, hovering around 40 citations per article. However, in 2023, a substantial decline to 0.75 citations per article was observed, possibly due to the limited timeframe since publication. The correlation between annual scientific production and mean citations reveals intriguing insights. Years witnessing high publication rates might not necessarily translate into immediately high citation rates. The impact of scholarly work appears to fluctuate and might take time to manifest, evident in the lower citations in the year of publication (e.g., 2023). However, sustained periods of high citations (2018-2020) hint at the lasting influence of research published in those years.

6. Key Themes and Collaborations: The WordCloud and Co-occurrence Network analyses provided a comprehensive overview of prevalent themes and interrelationships within research publications. The WordCloud highlighted dominant terms such as "intrusion detection," "big data," "network security," and "machine learning," reflecting the primary areas of focus. Meanwhile, the Co-occurrence Network showcased clusters around core themes, indicating associations between terms like "intrusion detection systems," "data analytics," and "cybersecurity."
Moreover, the Collaboration Network among authors demonstrated distinct clusters of collaboration. Authors like jemili , korbaa, abid, and meddeb formed central clusters, indicating substantial collaborations within their respective groups. Conversely, authors such as li w, meng, ahmad and wahid appeared in separate clusters, suggesting potential isolated collaborations or distinct research pursuits.
7. Geographical Influence and Impact: Figure 4 illustrated the contributions of authors from different countries as corresponding authors. India emerged as the leading contributor (22 articles), followed by China (18 articles) and Saudi Arabia (13 articles). However, the impact of their publications, represented by citations per article, revealed interesting nuances. While the USA had fewer publications (4 articles), its impact was significantly high, averaging 53 citations per article. In contrast, China and Saudi Arabia had higher publication counts but relatively lower impact per article, with averages of 18.60 and 191 citations per article, respectively.
8. Comparing Author Contributions and Impact: The analysis of authors' contributions and their impact sheds light on their relative influence within the field. Jemili, despite contributing the most articles (10), had a fractionalized contribution of 5.08, indicating a significant impact comparable to authors with fewer publications. Conversely, authors like Wang and Abid, with fewer articles (4 and 3, respectively), had fractionalized contributions of 1.53 and 1.17, showcasing their substantive impact with fewer contributions.

9. Interpreting Results and Implications: The surge in scholarly output might denote heightened interest driven by technological advancements or evolving security concerns in the digital landscape. However, a high publication rate doesn't always correlate directly with immediate impact. It's essential to consider the lag time for citations to accrue and the longevity of influential publications over time.

The prominence of certain terms (e.g., "intrusion detection," "big data") and their co-occurrence signifies the primary focus areas. The interconnectedness of terms reveals the interdisciplinary nature of the field, blending security concepts with emerging technologies like machine learning and data analytics. Geographically, while certain countries contributed more publications, the impact varied. Countries with fewer publications might yield more impactful research per article, highlighting the quality of contributions over quantity. The collaboration network underscores the importance of cohesive research groups, while also indicating potential isolated efforts. Collaboration can lead to prolific outputs, but diverse collaborations might indicate varied research interests or specialized pursuits within the field. The field of intrusion detection and big data exhibits a dynamic landscape characterized by a continuous surge in scholarly output, evolving research themes, diverse collaborations, and varying impacts across different countries and authors. Understanding the correlations and disparities between publication trends, impact, collaboration patterns, and thematic foci provides valuable insights for researchers and practitioners aiming to navigate and contribute to this evolving domain.

This discussion, showcases the multifaceted nature of scholarly endeavors within the intersection of intrusion detection and big data. While the findings elucidate several trends and patterns, they also underline the complexities and diverse facets present within the field, inviting further exploration and nuanced analyses for a deeper understanding of this dynamic domain.

5. CONCLUSION

The analysis conducted based on the provided dataset sheds light on the multifaceted landscape of intrusion detection and big data research over the past decade. The findings reveal intriguing trends and patterns, offering valuable insights into the evolution, thematic emphasis, author contributions, and geographical influences within this dynamic field. The investigation into scholarly output trends illustrates a consistent increase in publications over the years, particularly notable from 2018 onwards, peaking significantly in 2023. However, while the volume of publications surged, the immediate citation impact displayed variations, indicating that the recognition and influence of research might take time to manifest within the academic sphere. Thematic analysis delineates dominant topics such as "intrusion detection," "big data," "network security," and "machine learning" as central themes within the research discourse. These terms signify the core areas of focus, portraying the integration of security concerns with cutting-edge technologies, suggesting a multi-dimensional approach to address security challenges. Geographical influences on research contributions showcase countries like India, China, and Saudi Arabia as major contributors to publications. Interestingly, despite having fewer publications, the USA exhibits significantly higher citation averages per article, indicating the potency and impact of research originating from this region. Author contributions and their subsequent impact demonstrate intriguing dynamics. Prolific authors like Jemili showcase extensive contributions coupled with substantial impact. Conversely, authors with fewer publications often demonstrate notable influence per article, suggesting a focus on quality contributions over sheer quantity. Moving forward, fostering diverse collaborations across disciplines remains pivotal for innovation within the field. Encouraging interdisciplinary research initiatives can further integrate advanced computational methods, such as machine learning and data analytics, to fortify security measures.

Future directions could involve longitudinal studies to gauge the lasting impact of recent publications, emphasizing the assessment of research quality alongside quantity, and promoting global collaborations to enrich the field's diversity of ideas. The provided data analysis offers a glimpse into the vibrant landscape of intrusion detection and big data research. By addressing collaboration dynamics, thematic evolution, and global influences, the field can advance its efficacy in addressing the intricate and evolving security challenges inherent in our increasingly digitized world.

Data Source

The bibliometric data utilized in this study, titled "Mapping the Evolution of Intrusion Detection in Big Data: A Bibliometric Analysis," was sourced from a BibTeX file available on GitHub. This BibTeX file contains a curated collection of scholarly articles sourced from reputable databases, such as Scopus, spanning the intersection of intrusion detection and big data research. The dataset comprises essential bibliographic metadata, including publication titles, authors, publication years, abstracts, keywords, and other relevant information related to the publications. The dataset's inclusion criteria focused on articles specifically addressing intrusion detection within the context of big data analytics.

The dataset underwent rigorous quality checks to ensure relevance and reliability in reflecting the scholarly landscape within this interdisciplinary domain.

Conflicts of Interest

The author's paper clearly states that no conflicts of interest exist in relation to the research or its publication.

Funding

The author's paper explicitly states that no funding was received from any institution or sponsor.

Acknowledgment

The author acknowledges the assistance and guidance received from the institution in various aspects of this study.

References

[1] F. Jemili, "Towards data fusion-based big data analytics for intrusion detection," (in English), Journal of Information and Telecommunication, Article vol. 7, no. 4, pp. 409-436, 2023.

[2] A. Abid, F. Jemili, and O. Korbaa, "Real-time data fusion for intrusion detection in industrial control systems based on cloud computing and big data techniques," (in English), Cluster Computing, Article 2023.

[3] Y. Wang and S. Zhang, "Simulation of Cloud Computing Network Security Intrusion Detection Model Based on Neural Network Algorithm Driven by Big Data," in Proceedings - 2023 2nd International Conference on Artificial Intelligence and Autonomous Robot Systems, AIARS 2023, 2023, pp. 37-40: Institute of Electrical and Electronics Engineers Inc.

[4] A. Abid and F. Jemili, "Intrusion Detection based on Graph oriented Big Data Analytics," in Procedia Computer Science, 2020, vol. 176, pp. 572-581: Elsevier B.V.

[5] K. Peng, V. C. M. Leung, L. Zheng, S. Wang, C. Huang, and T. Lin, "Intrusion detection system based on decision tree over big data in fog environment," (in English), Wireless Communications and Mobile Computing, Article vol. 2018, 2018, Art. no. 4680867.

[6] K. Peng, V. C. M. Leung, and Q. Huang, "Clustering Approach Based on Mini Batch Kmeans for Intrusion Detection System over Big Data," (in English), IEEE Access, Article vol. 6, pp. 11897-11906, 2018.

[7] M. M. Rathore, A. Ahmad, and A. Paul, "Real time intrusion detection system for ultra-high-speed big data environments," (in English), Journal of Supercomputing, Article vol. 72, no. 9, pp. 3489-3510, 2016.

[8] M. Ajabi, I. Boukhris, and Z. Elouedi, "Big data classification using belief decision trees: application to intrusion detection," in Advances in Intelligent Systems and Computing, 2016, vol. 407, pp. 369-379: Springer Verlag.

[9] K. Siddique, Z. Akhtar, and Y. Kim, "Intrusion detection in high-speed big data networks: A comprehensive approach," in Lecture Notes in Electrical Engineering, 2018, vol. 474, pp. 1364-1370: Springer Verlag.

[10] F. Jemili, R. Meddeb, and O. Korbaa, "Intrusion detection based on ensemble learning for big data classification," (in English), Cluster Computing, Article 2023.

[11] J. Xiang, M. Westerlund, D. Sovilj, and G. Pulkkis, "Using extreme learning machine for intrusion detection in a big data environment," in Proceedings of the ACM Conference on Computer and Communications Security, 2014, vol. 2014-November, pp. 73-82: Association for Computing Machinery.

[12] W. Meng and W. Li, "A Review of Network Intrusion Detection in the Big Data Era: Challenges and Future Trends," in Networking For Big Data: CRC Press, 2015, pp. 195-214.

[13] K. Ahmad, G. Kumar, A. Wahid, and M. M. Kirmani, "Intrusion detection and prevention on flow of big data using bacterial foraging," in Handbook of Research on Securing Cloud-Based Databases with Biometric Applications: IGI Global, 2014, pp. 386-411.

[14] K. Peng, L. Zheng, X. Xu, T. Lin, and V. C. M. Leung, "Balanced iterative reducing and clustering using hierarchies with principal component analysis (PBirch) for intrusion detection over big data in mobile cloud environment," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2018, vol. 11342 LNCS, pp. 166-177: Springer Verlag.

[15] K. Siddique, Z. Akhtar, M. A. Khan, Y. H. Jung, and Y. Kim, "Developing an intrusion detection framework for high-speed big data networks: A comprehensive approach," (in English), KSII Transactions on Internet and Information Systems, Article vol. 12, no. 8, pp. 4021-4037, 2018.

[16] I. Boukhris, Z. Elouedi, and M. Ajabi, "Toward intrusion detection using belief decision trees for big data," (in English), Knowledge and Information Systems, Article vol. 53, no. 3, pp. 671-698, 2017.