

# Unveiling the Evolution of Neural Network-based Approaches for Fake News Detection in Social Media

AKSHAT GAURAV<sup>1</sup>, KWOK TAI CHUI<sup>2</sup>

<sup>1</sup>Ronin Institute, Montclair, USA akshat.gaurav@ieee.org

<sup>2</sup>Hong Kong Metropolitan University (HKMU), Hong Kong Email: jktchui@hkmu.edu.hk

**ABSTRACT** This paper presents a analysis of a dataset obtained from the Scopus database, spanning the period from 2017 to 2023. The dataset comprises 354 documents sourced from 218 different journals, books, and other publications, offering insights into recent research trends and developments. The analysis reveals important information about document types, authorship patterns, collaboration, and keyword frequencies. The dataset encompasses various document types, including articles, book chapters, conference papers, conference reviews, and reviews, reflecting the breadth and diversity of scholarly outputs within the field. Authorship patterns highlight the contributions of different authors to the dataset. The analysis reveals both single-authored and multi-authored documents, with a focus on collaborative research. The collaboration index indicates a moderate level of collaboration, underscoring the significance of teamwork and cooperation in the research field. Moreover, keyword analysis reveals prominent topics within the dataset. Terms like "fake news," "fake news detection," "deep learning," and "social media" emerge as focal points, indicating the thematic areas explored in the research papers. These keywords shed light on the research interests and methodologies employed within the dataset. Overall, this analysis provides a comprehensive overview of the collected dataset, offering valuable insights into research output, collaboration patterns, and topical interests. The findings contribute to a deeper understanding of the current research landscape within the subject area. Researchers can leverage this information to identify key contributors, track research trends, and explore potential areas of collaboration.

**KEYWORDS** Fake news, Misinformation, Disinformation, Machine learning, Natural language processing, Deep learning, Social media analysis

## I. INTRODUCTION

In recent years, the proliferation of misinformation and fake news has emerged as a significant challenge in the era of digital information. The spread of false or misleading information through various online platforms has the potential to influence public opinion, disrupt social harmony, and undermine trust in media sources. As a result, there has been a growing interest in understanding and combating the issue of fake news within the research community [1]–[3].

The detection and mitigation of fake news have become critical objectives in ensuring the reliability and integrity of information sources. Researchers and practitioners have been actively investigating various methods and approaches to tackle this problem. Fake news detection involves the application of computational techniques, such as machine learning, natural language processing, and deep learning, to analyze and identify deceptive or misleading content [4]–[6].

Existing research in the field of fake news detection has made significant progress in developing algorithms and models to distinguish between legitimate and fabricated news

articles. These methods often leverage features such as linguistic patterns, social media metadata, credibility of sources, and propagation patterns to detect deceptive content. Some studies have explored the use of deep learning techniques, such as recurrent neural networks and convolutional neural networks, to capture complex relationships and contextual cues for improved detection accuracy.

However, despite the advancements in fake news detection methods, several research gaps persist [7]. One notable research gap pertains to the adaptability of detection algorithms to evolving forms of fake news. The perpetrators of misinformation constantly refine their techniques, making it challenging for detection models to keep up with the evolving landscape. Additionally, the scalability and real-time applicability of detection methods remain areas of concern, as the volume and speed of information dissemination continue to increase [8]–[10].

To address these research gaps and contribute to the existing body of knowledge, this paper presents a comprehensive analysis of the Scopus database. The analysis examines

recent research trends, document types, authorship patterns, collaboration networks, and keyword frequencies within the field of fake news and fake news detection. By providing an overview of the research landscape and identifying key areas of focus, this study aims to contribute to the understanding of current developments and potential avenues for future research in the fight against fake news.

## II. RESULTS AND DISCUSSION

The dataset collected from the Scopus database provides valuable insights into patterns and research collaboration, as represented in Fig. 1. The data spans from 2017 to 2023, representing recent research trends. Over this period, a total of 354 documents were analyzed, sourced from 218 different journals, books, and other publications.

One important measure of a document's impact is the average number of citations it receives. In this dataset, the average citations per document stand at 13.97, indicating that the included documents have been cited frequently by other researchers. This suggests that the research within these documents has made a significant contribution to the scholarly community.

Furthermore, the average citations per year per document, which is 3.66, gives us an idea of the rate at which the documents continue to be cited over time. This metric reflects the enduring relevance and influence of the research findings.

Fig. 1 encompasses various document types, including articles, book chapters, conference papers, conference reviews, and reviews. This diversity highlights the broad scope of research covered in the analysis, representing a wide range of scholarly outputs.

Fig. 1 also provides insights into the collaboration patterns among authors. Out of the 354 documents, only 8 were single-authored, while the remaining 979 documents were multi-authored. On average, each document had 2.79 authors, indicating a tendency towards collaborative research.

Further analysis reveals that the collaboration index, which compares the average number of authors per document to the average number of authors for single-authored documents, is 3.04. This suggests a moderate level of collaboration among authors. The presence of multiple authors per document and a collaboration index above 1 demonstrates the importance of teamwork and cooperation in contemporary research.

In terms of keywords, the dataset contains both Keywords Plus (ID) and Author's Keywords (DE). The Keywords Plus (ID) category includes 1,524 unique terms, while the Author's Keywords (DE) category comprises 625 unique terms. These keywords provide additional insights into the main themes and concepts addressed within the documents, facilitating future research and exploration.

Overall, the analysis of the Scopus database presents a comprehensive overview of recent research trends. The high average citations per document, along with the varied document types and collaborative authorship, underscores the significance and impact of the included research. The findings from this dataset serve as a valuable resource for

researchers seeking to understand the current landscape of scholarly publications and the collaborative dynamics within academic circles.

The annual production of papers in Fig. 2 shows a notable growth trend over the years. In 2017, there were only 2 articles published, but the number increased to 12 in 2018 and further to 27 in 2019, representing a gradual rise. However, the growth rate accelerated significantly from 2020 onwards. In that year, the number of articles surged to 49, followed by a substantial increase to 103 in 2021. The growth trend continued in 2022, with 112 articles published.

The annual growth rate, calculated as 70.42%, highlights the rapid expansion in research output observed within the dataset. This indicates an increasing interest and engagement in the subject area or the research field under consideration.

The substantial growth in the number of articles over the years suggests several conclusions. Firstly, it indicates an active and dynamic research environment, with researchers contributing significantly to the literature. The steady rise in publication numbers from 2017 to 2022 reflects a consistent effort to disseminate new findings and insights.

The sharp increase in the annual production of papers from 2020 onwards may indicate a surge in research activity, potentially driven by emerging trends, breakthroughs, or increased funding and support. This could suggest a growing interest and investment in the subject area or research field, stimulating a higher volume of publications.

The notable growth rate underscores the potential for future research development and expansion in the field. The increasing number of papers implies an expanding knowledge base and a growing pool of information for researchers and practitioners to draw upon.

The findings from this analysis emphasize the importance of keeping up with the latest research in the field, given the substantial growth observed in recent years. Researchers should be aware of the evolving landscape and stay informed about the latest publications to remain at the forefront of knowledge and contribute to the field's advancements.

Overall, the significant annual growth rate in the number of papers published in the dataset indicates a vibrant research landscape and a promising outlook for further developments and discoveries in the field.

The information provided in Fig. 3 offers insights into the authors' contributions and their respective article counts within the dataset. The data is presented both in terms of the absolute number of articles authored and the fractionalized representation of their contributions.

Among the authors, KALIYAR RK has the highest number of articles with 13, representing a significant contribution to the dataset. This accounts for 4.09% of the total articles. Following closely is GOSWAMI A with 11 articles, contributing 2.93% of the total. NARANG P and KUMAR A have authored 9 and 8 articles, respectively, contributing 2.87% and 3.08% to the overall dataset.

Additionally, there are several authors who have contributed a smaller number of articles. For instance, authors

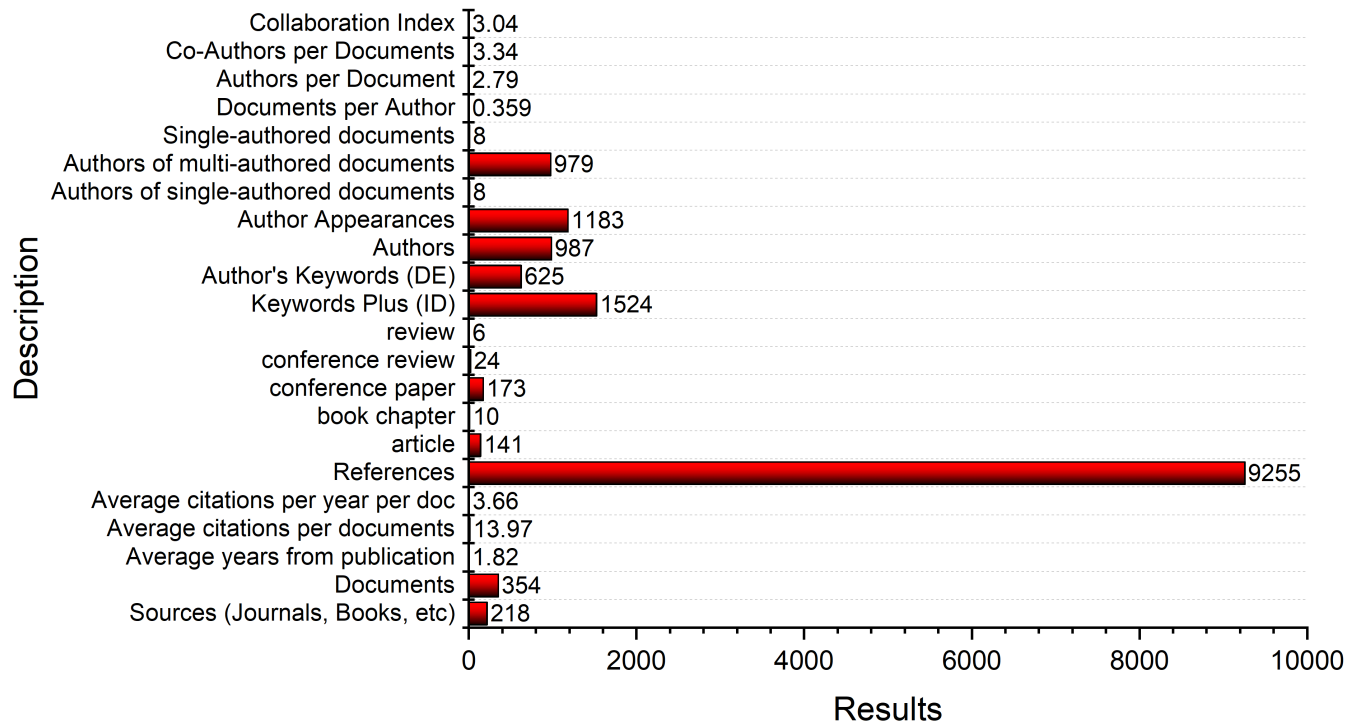


FIGURE 1: Main Information

TABLE 1: Highly Cited Papers

Paper	DOI	Total Citations
WANG Y, 2018, PROC ACM SIGKDD INT CONF KNOWL DISCOV DATA MIN [11]	10.1145/3219819.3219903	451
JIN Z, 2017, MM - PROC ACM MULTIMED CONF [12]	10.1145/3123266.3123454	265
ISLAM AKMN, 2020, TECHNOL FORECAST SOC CHANGE [13]	10.1016/j.techfore.2020.120201	203
NASIR JA, 2021, INT J INF MANAG DATA INSIGHTS [14]	10.1016/j.ijime.2020.100007	184
AJAO O, 2018, ACM INT CONF PROC SER [15]	10.1145/3217804.3217917	167
KALIYAR RK, 2021, MULTIMEDIA TOOLS APPL [16]	10.1007/s11042-020-10183-2	165
KALIYAR RK, 2020, COGN SYS RES [17]	10.1016/j.cogsys.2019.12.005	165
MA J, 2018, WEB CONF - COMPANION WORLD WIDE WEB CONF, WWW [18]	10.1145/3184558.3188729	134
ALVAREZ-RISCO A, 2020, AM J TROP MED HYG [19]	10.4269/ajtmh.20-0536	127
LU Y-J, 2020, PROC ANNU MEET ASSOC COMPUT LINGUIST [20]	NA	119
QIAN F, 2018, IJCAI INT JOINT CONF ARTIF INTELL [21]	10.24963/ijcai.2018/533	111
GHANEM B, 2020, ACM TRANS INTERNET TECHNOL [22]	10.1145/3381750	109
SHU K, 2019, COMPUT MATH ORGAN THEORY [23]	10.1007/s10588-018-09280-3	101
QI P, 2019, PROC IEEE INT CONF DATA MIN ICDM [24]	10.1109/ICDM.2019.00062	92
GIRGIS S, 2019, PROC - INT CONF COMPUT ENG SYST, ICCES [25]	10.1109/ICCES.2018.8639198	86
KULA S, 2020, LECT NOTES COMPUT SCI [26]	10.1007/978-3-030-50423-6_49	76
KUMAR S, 2020, TRANS EMERG TELECOMMUN TECHNOL [27]	10.1002/ett.3767	75
CHOUDHARY A, 2021, EXPERT SYS APPL [28]	10.1016/j.eswa.2020.114171	72
SONG C, 2021, INF PROCESS MANAGE [29]	10.1016/j.ipm.2020.102437	66
ABDELMINAAM DS, 2021, IEEE ACCESS [30]	10.1109/ACCESS.2021.3058066	65

like KUMAR S, SHU K, ZHANG Y, MEEL P, and MOM-TAZI S have contributed 6 or 5 articles each, representing a range between 1.53% to 2.12% of the dataset.

Further down the list, there are authors who have contributed 4, 3, or fewer articles. These authors, such as LIU H, SHARMA S, WANG Y, and ASHRAF I, have made relatively smaller contributions, representing less than 2

In conclusion, the dataset is marked by a range of author contributions, with some authors significantly contributing to the overall article count. Authors like KALIYAR RK, GOSWAMI A, NARANG P, and KUMAR A have made

substantial contributions, collectively accounting for a considerable proportion of the articles in the dataset.

It is also worth noting that there are authors who have contributed fewer articles individually, but their combined contributions form a substantial fraction of the dataset. This indicates a diverse group of authors making valuable contributions to the field.

Overall, the findings emphasize the significance of the identified authors in the dataset, highlighting their research output and potential influence in the subject area. Researchers and readers interested in the dataset should pay attention to

## Documents by year

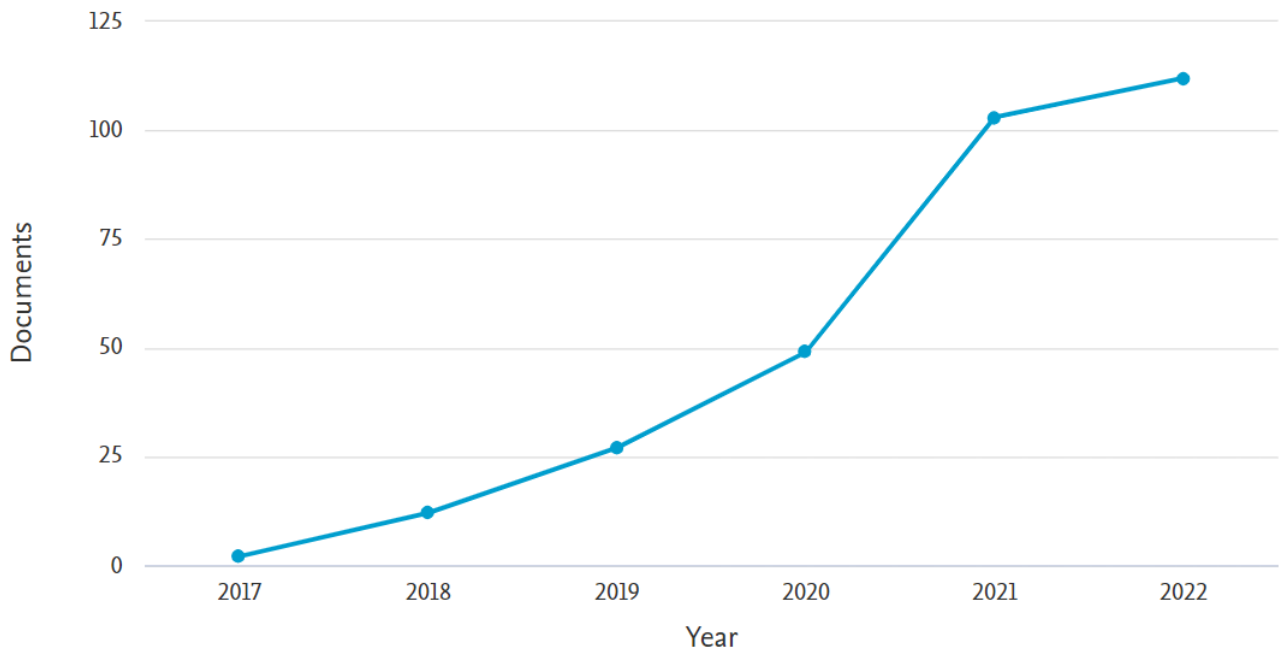


FIGURE 2: Annual Scientific Production

the works of these authors to gain insights into the subject matter and potential collaborations within the field.

Fig. 4 highlights the frequency of certain keywords within the dataset, indicating the prominence of specific terms in the research papers. The dataset encompasses a range of keywords, with some appearing more frequently than others.

The most frequently occurring keyword is "fake news," which appears 121 times within the dataset. This indicates a significant focus on the topic of fake news, suggesting that the papers in the dataset likely revolve around this subject. The closely related term "fake news detection" is also prominent, appearing 106 times. This suggests a strong emphasis on developing methods and techniques to identify and combat fake news within the scholarly discourse.

Another notable keyword is "deep learning," which appears 95 times. This indicates a substantial interest in the application of deep learning techniques within the field under consideration. Similarly, "social media" appears 79 times, highlighting the relevance and influence of social media platforms within the research context.

Other keywords that appear with lower frequencies include "machine learning" (57), "natural language processing" (32), "neural networks" (25), and "neural network" (24). These terms signify the utilization of machine learning techniques, natural language processing algorithms, and neural networks within the research papers, indicating the technical aspects and methodologies explored in the dataset.

Additionally, more specific terms like "twitter" (20) and "LSTM" (19) also emerge, suggesting a focus on the anal-

ysis of Twitter data and the utilization of Long Short-Term Memory (LSTM) networks within the research papers.

In conclusion, the dataset showcases a strong emphasis on the topic of fake news, along with a notable interest in deep learning, social media, and various machine learning techniques. These keywords reflect the main themes and technical aspects addressed within the papers, offering insights into the research trends and areas of focus within the field.

Researchers and readers interested in the dataset should pay attention to the papers related to fake news, as well as those involving deep learning, social media, and machine learning techniques. These keywords provide a glimpse into the current research landscape and the relevant methodologies employed by scholars.

### III. CONCLUSION

In conclusion, this analysis based on the Scopus database provides valuable insights into the research landscape within the chosen subject area. The dataset covers a timespan from 2017 to 2023, indicating recent research trends and developments. The collected information sheds light on various aspects of the dataset, including document types, authorship, collaboration patterns, and keyword frequencies.

The dataset comprises a diverse range of document types, such as articles, book chapters, conference papers, conference reviews, and reviews. This diversity reflects the breadth of scholarly outputs and the multidisciplinary nature of the research field.

The analysis of authorship patterns reveals the contribu-

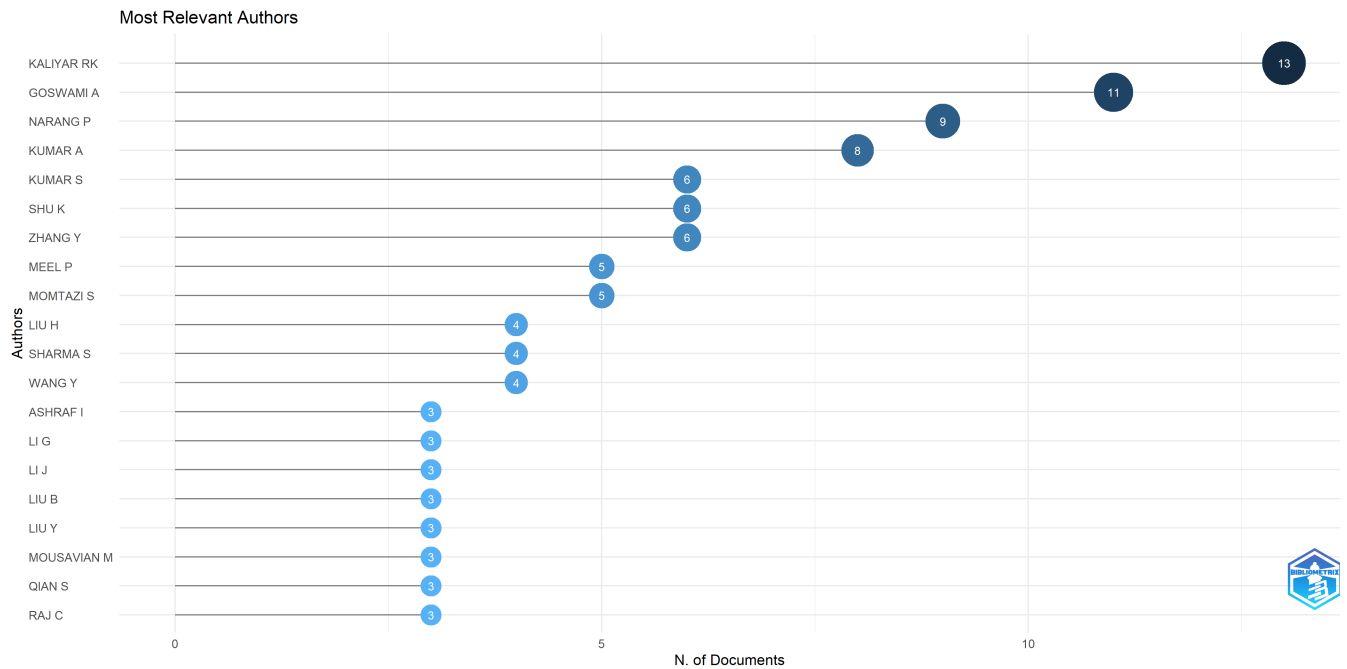


FIGURE 3: Most Relevant Authors



FIGURE 4: Important Keywords

tions of different authors to the dataset. Several authors stand out with a notable number of articles, demonstrating their significant involvement in the research field. Collaboration among authors is also apparent, with most documents being multi-authored. The collaboration index indicates a moderate level of collaboration, emphasizing the importance of teamwork and cooperation in the field.

The dataset also highlights key keywords that are frequently mentioned within the papers. Terms such as "fake news," "fake news detection," "deep learning," and "social media" emerge as prominent topics. These keywords provide insights into the main themes and technical aspects explored within the dataset, indicating the focus of research and the methodologies employed.

Overall, this paper provides a comprehensive overview of the collected dataset, offering valuable information about research output, collaboration patterns, and topical interests within the chosen subject area. The findings contribute to the understanding of the current state of research, providing a foundation for further exploration and analysis. Researchers can leverage this information to identify significant contributors, track research trends, and identify potential collaboration opportunities.

## REFERENCES

- [1] A. Tewari and et al., "Cryptanalysis of a novel ultra-lightweight mutual authentication protocol for iot devices using rfid tags," *The Journal of Supercomputing*, vol. 73, pp. 1085–1102, 2017.
- [2] E. Alomari and et al., "Botnet-based distributed denial of service (ddos) attacks on web servers: classification and art," *arXiv preprint arXiv:1208.0403*, 2012.
- [3] M. Al-Ayyoub and et al., "Deep learning for arabic nlp: A survey," *Journal of computational science*, vol. 26, pp. 522–531, 2018.
- [4] L. Zou and et al., "A novel coverless information hiding method based on the average pixel value of the sub-images," *Multimedia tools and applications*, vol. 78, pp. 7965–7980, 2019.
- [5] B. Gupta, S. Gupta, S. Gangwar, M. Kumar, and P. Meena, "Cross-site scripting (xss) abuse and defense: exploitation on several testing bed environments and its defense," *Journal of Information Privacy and Security*, vol. 11, no. 2, pp. 118–136, 2015.
- [6] A. Almomani and et al., "Phishing dynamic evolving neural fuzzy framework for online detection zero-day phishing email," *arXiv preprint arXiv:1302.0629*, 2013.
- [7] e. a. Plageras, Andreas P and, "Efficient large-scale medical data (ehealth big data) analytics in internet of things," in *2017 IEEE 19th Conference on Business informatics (CBI)*, vol. 2. IEEE, 2017, pp. 21–27.
- [8] e. a. Poonia, Vikas and, "Drought occurrence in different river basins of india and blockchain technology based framework for disaster management," *Journal of Cleaner Production*, vol. 312, p. 127737, 2021.
- [9] B. B. Gupta, K. Yadav, I. Razzak, K. Psannis, A. Castiglione, and X. Chang, "A novel approach for phishing urls detection using lexical based machine learning in a real-time environment," *Computer Communications*, vol. 175, pp. 47–57, 2021.
- [10] A. Singh and B. B. Gupta, "Distributed denial-of-service (ddos) attacks and defense mechanisms in various web-enabled computing platforms: issues, challenges, and future research directions," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 18, no. 1, pp. 1–43, 2022.
- [11] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, L. Su, and J. Gao, "Eann: Event adversarial neural networks for multi-modal fake news detection," *Association for Computing Machinery*, 2018, p. 849 – 857, all Open Access, Bronze Open Access.
- [12] Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo, "Multimodal fusion with recurrent neural networks for rumor detection on microblogs," *Association for Computing Machinery, Inc*, 2017, p. 795 – 816.
- [13] A. N. Islam, S. Laato, S. Talukder, and E. Sutinen, "Misinformation sharing and social media fatigue during covid-19: An affordance and cognitive load perspective," *Technological Forecasting and Social Change*, vol. 159, 2020, all Open Access, Bronze Open Access, Green Open Access.
- [14] J. A. Nasir, O. S. Khan, and I. Varlamis, "Fake news detection: A hybrid cnn-rnn based deep learning approach," *International Journal of Information Management Data Insights*, vol. 1, no. 1, 2021, all Open Access, Gold Open Access.
- [15] O. Ajao, D. Bhowmik, and S. Zargari, "Fake news identification on twitter with hybrid cnn and rnn models," *Association for Computing Machinery*, 2018, p. 226 – 230, all Open Access, Green Open Access.
- [16] R. K. Kaliyar, A. Goswami, and P. Narang, "Fakebert: Fake news detection in social media with a bert-based deep learning approach," *Multimedia Tools and Applications*, vol. 80, no. 8, p. 11765 – 11788, 2021, all Open Access, Bronze Open Access, Green Open Access.
- [17] R. K. Kaliyar, A. Goswami, P. Narang, and S. Sinha, "Fndnet – a deep convolutional neural network for fake news detection," *Cognitive Systems Research*, vol. 61, p. 32 – 44, 2020.
- [18] J. Ma, W. Gao, and K.-F. Wong, "Detect rumor and stance jointly by neural multi-task learning," *Association for Computing Machinery, Inc*, 2018, p. 585 – 593, all Open Access, Bronze Open Access, Green Open Access.
- [19] A. Alvarez-Risco, C. R. Mejia, J. Delgado-Zegarra, S. Del-Aguila-Arcentales, A. A. Arce-Esquivel, M. J. Valladares-Garrido, M. R. Del Portal, L. F. Villegas, W. H. Curioso, M. C. Sekar, and J. A. Yáñez, "The peru approach against the covid-19 infodemic: Insights and strategies," *American Journal of Tropical Medicine and Hygiene*, vol. 103, no. 2, p. 583 – 586, 2020, all Open Access, Green Open Access.
- [20] Y.-J. Lu and C.-T. Li, "Gean: Graph-aware co-attention networks for explainable fake news detection on social media," *Association for Computational Linguistics (ACL)*, 2020, p. 505 – 514.
- [21] F. Qian, C. Gong, K. Sharma, and Y. Liu, "Neural user response generator: Fake news detection with collective user intelligence," L. J., Ed., vol. 2018-July. *International Joint Conferences on Artificial Intelligence*, 2018, p. 3834 – 3840, all Open Access, Bronze Open Access.
- [22] B. Ghanem, P. Rosso, and F. Rangel, "An emotional analysis of false information in social media and news articles," *ACM Transactions on Internet Technology*, vol. 20, no. 2, 2020, all Open Access, Green Open Access.
- [23] K. Shu, D. Mahudeswaran, and H. Liu, "Fakenewstracker: a tool for fake news collection, detection, and visualization," *Computational and Mathematical Organization Theory*, vol. 25, no. 1, p. 60 – 71, 2019.
- [24] P. Qi, J. Cao, T. Yang, J. Guo, and J. Li, "Exploiting multi-domain visual information for fake news detection," W. J., S. K., and W. X., Eds., vol. 2019-November. *Institute of Electrical and Electronics Engineers Inc.*, 2019, p. 518 – 527, all Open Access, Green Open Access.
- [25] S. Girgis, E. Amer, and M. Gadallah, "Deep learning algorithms for detecting fake news in online text," Z. A.M., T. M., E.-K. M.W., S. A., B. E.-D. A.M., and A. H.M., Eds. *Institute of Electrical and Electronics Engineers Inc.*, 2019, p. 93 – 97.
- [26] S. Kula, M. Choraś, R. Kozik, P. Ksieniewicz, and M. Woźniak, "Sentiment analysis for fake news detection by means of neural networks," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12140 LNCS, p. 653 – 666, 2020, all Open Access, Bronze Open Access.
- [27] S. Kumar, R. Asthana, S. Upadhyay, N. Upreti, and M. Akbar, "Fake news detection using deep learning models: A novel approach," *Transactions on Emerging Telecommunications Technologies*, vol. 31, no. 2, 2020.
- [28] A. Choudhary and A. Arora, "Linguistic feature based learning model for fake news detection and classification," *Expert Systems with Applications*, vol. 169, 2021.
- [29] C. Song, N. Ning, Y. Zhang, and B. Wu, "A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks," *Information Processing and Management*, vol. 58, no. 1, 2021.
- [30] D. S. Abdelmaam, F. H. Ismail, M. Taha, A. Taha, E. H. Houssein, and A. Nabil, "Coaid-deep: An optimized intelligent framework for automated detecting covid-19 misleading information on twitter," *IEEE Access*, vol. 9, p. 27840 – 27867, 2021, all Open Access, Gold Open Access, Green Open Access.