Journal of Science and Technology
ISSN: 2456-5660 Volume 8, Issue 07 (July -2023)
www.jst.org.in DOI:https://doi.org/10.46243/jst.2023.v8.i07.pp184- 192

IDENTIFYING AND PREVENTING THE DISSEMINATION OF FAKE NEWS

Syed Salman Hussain[1] | Boddupalli Anvesh[2] | Dr. V. Bapuji[3]

[1]Department of MCA, Vaageswari College of Engineering,
[2]Asst Professor, Department of MCA, Vaageswari of Engineering,
[3]Professor and Hod, Department of MCA, Vaageswari College of Engineering,

To Cite this Article
Syed Salman Hussain | Boddupalli Anvesh | Dr. V. Bapuji, **"IDENTIFYING AND PREVENTING THE DISSEMINATION OF FAKE NEWS "** *Journal of Science and Technology, Vol. 08, Issue 07,- July 2023, pp184-192*
Article Info
Received: 04-06-2023 Revised: 05-07-2023 Accepted: 15-07-2023 Published: 27-07-2023

ABSTRACT

Misinformation poses a significant threat to democratic societies, particularly in today's interconnected digital world, as it has the potential to shape public opinion. Researchers from various disciplines, including computer science, political science, information science, and linguistics, have been investigating the spread of fake news, methods for detecting it, and strategies to mitigate its impact. However, effectively identifying and preventing the dissemination of false information remains a complex endeavor. Given the increasing role of Artificial Intelligence (AI) systems, it is vital to offer clear and user – ,bfriendly explanations for the decisions made by fake news detectors, particularly on social media platforms. Therefore, this paper conducts a systematic analysis of the latest approaches employed to detect and combat the spread of fake news. By examining these approaches, we uncover key challenges and propose potential future research directions, with a particular emphasis on integrating AI explain ability into fake news credibility systems.

KEYWORDS: *AI explainability, fabricated news, synthetic media, misleading content, dishonest information, fake profiles, online platforms, biased information, block – chain powered identification.*

INTRODUCTION

Numerous solutions[1] have been suggested to tackle various security and privacy issues, whether they are related to the Internet of Things(IoT)[2], user authentication problems[3], enhancing road traffic safety or other cyber-crime threats. However, as people and organizations increasingly rely on real-time information from diverse sources such as user- generated content and social media platforms there is a new risk emerging: the abuse of these sources and dissemination methods through the spread of fake news.

Currently, research in the field of fake news is still in its preliminary stages. We define fake news as content intentionally created to deceive users, aiming to mislead[4], deceive or defame individuals, groups, organizations, and governments. Fake news can have various consequences on our society, as illustrated in figure 1.

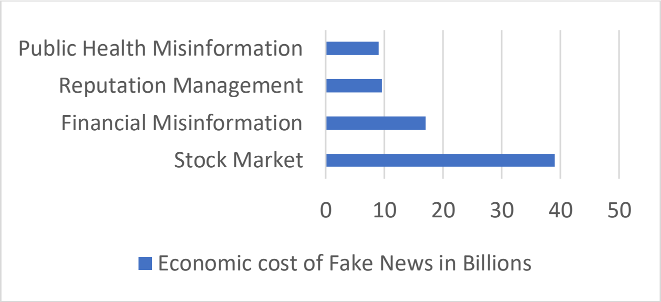

Figure 1Potential Economical Impacts

For instance, a study conducted by CHEQ, a cybersecurity company and the University of Baltimore[5] revealed that globally fake news has an economic impact of approximately $78 Billion, with a direct annual loss of about $39 Billion in the stock market . Notable real-world events such as the COVID – 19 vaccine rollout and the 2016 US presidential Election[6] have been significantly influenced by fake news. These incidents highlight the urgent need for developing effective techniques and approaches to detect and counteract the dissemination of fake news. There are several difficulties in determining whether content is genuine or fabricated, as a news article can contain elements of truth and falsehood.

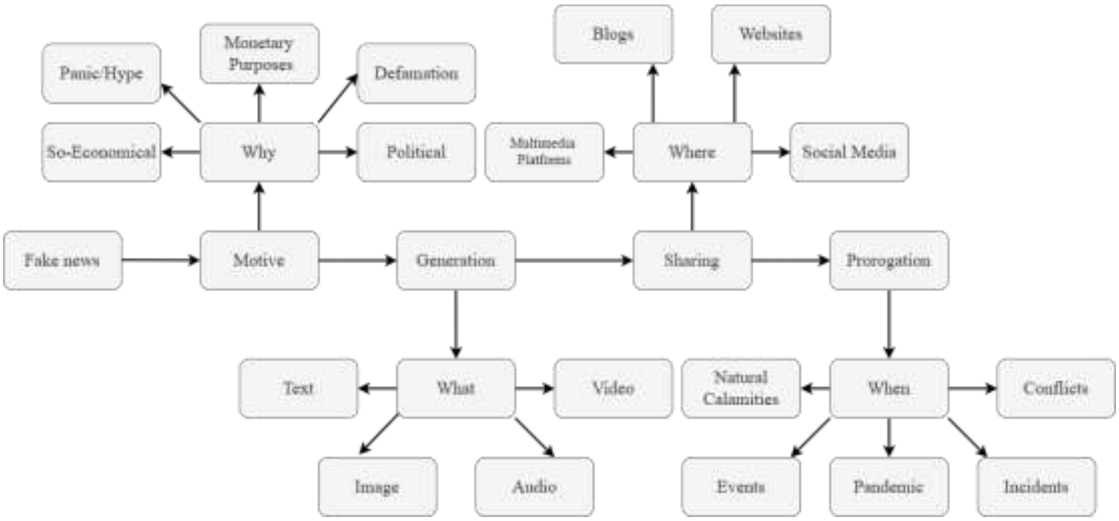

Figure 2Fake news life Cycle

Figure 2 illustrates the life cycle of fake news, outlining how it is generated and spread across various platforms. The motivation behind creating and disseminating fake news can vary, including financial gain, political agendas (both left and right wing), promoting terrorism to incite panic or unrest and defamation, among others. Fake news can take different forms such as text, images, audio, and videos. To combat this issue, numerous solutions have been proposed and fact – checking websites like "PolitiFact" and "TruthorFiction" have gained popularity. Through our analysis , we propose a new taxonomy for categorizing fake news detection approaches. Additionally, we discuss the existing and emerging challenges in detecting fake news and present potential research directions.

The structure of the paper is as follows: we explain existing system, followed by proposed system and finally conclusion.

1. EXISTING SYSTEM

To narrow down the research scope, we specifically concentrated on scientific journal articles from 2019 that were indexed in Scientific Citation Index (SCI), excluding conference papers and book chapters. Our objective was to summarize the latest studies on fake news detection, categorizing them into seven distinct detection categories, as illustrated in Figure 3

The taxonomy consists of four tiers, with the first tier organizing studies according to their research focus. Each color represents a different research focus for detecting fake news.

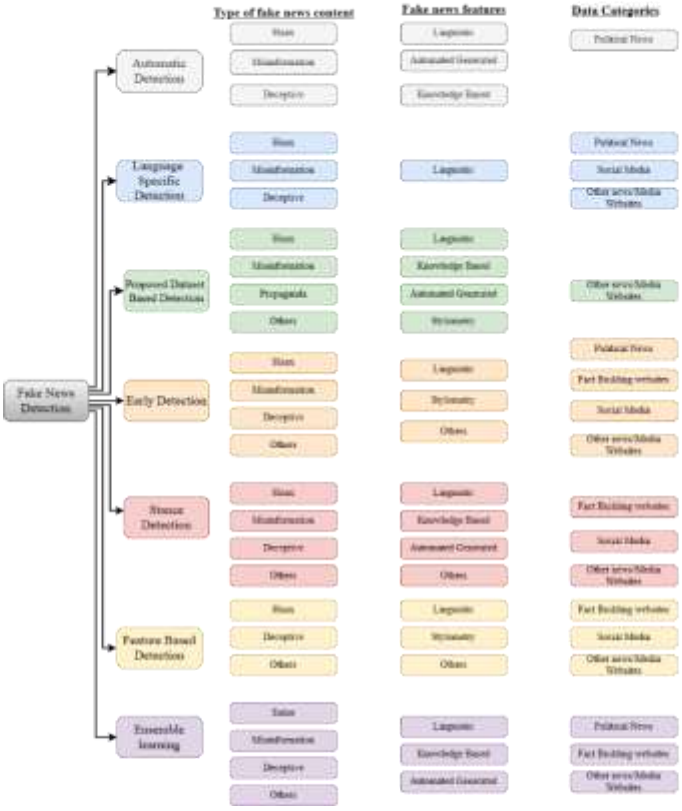

Figure 3Taxonomy of Fake news detection

The second tier categorizes researchers' work based on the types of fake news content. The third tier focuses on the identifying fake news features, while the fourth tier classifies the datasets used. The study highlights the significant emphasis placed by researchers on feature identification for detecting fake news.

Features have been crucial in fake news detection models, particularly because distinguishing between real and fake classes can be challenging due to their similar characteristics. Figure 4 illustrates the various aspects of features explored in this survey. Features can be examined from two perspectives, those that are difficult to imitate by malicious users (topographical features) and those that can be easily replicated.

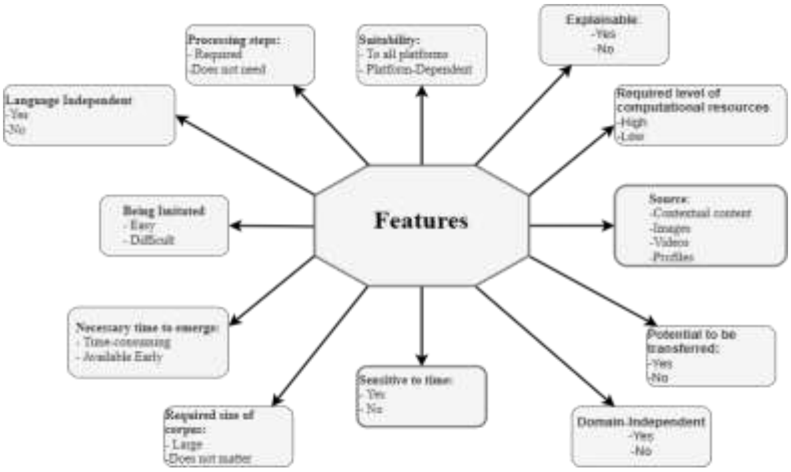

Figure 4Different aspects of Features.

Some features are not applicable to all platforms and situations. For Instance, [7] domain reputation features are not suitable for social network platforms. The emergence of certain features like user reactions to news articles requires a specific time. Unlike hand – crafted features, deep learning-based features lack interpretability and act as black boxes. Certain feature types, such as user profile – based features, require minimal pre-processing[8].

Large corpora are necessary features obtained from Linguistic Inquiry and Word Count (LIWC), News Evaluator for Linguistic Analysis (NELA) and different embeddings. Language – independent. Content – based and image – based features may not be large corpora are necessary features obtained from Linguistic Inquiry and Word Count (LIWC), News Evaluator for Linguistic Analysis (NELA) and different embeddings. Language – independent. Content – based and image – based features may not be universally applicable to all the domains. For instance, word embedding vectors require a domain – specific corpus with sufficient data.

3. PROPOSED SYSTEM

Neves et al[9] proposed an innovative approach called GANprintR(Generative Adversarial Network – fingerprint removal autoencoder) to deceive facial manipulation detection systems. The objective of this approach was to eliminate the traces left by Generative Adversarial Networks(GANs) without compromising the quality of the manipulated images. By successfully removing these fingerprints, the authors aimed to make the generated images indistinguishable from real ones, both to human observers and to machine – based detection systems. GANprintR was trained using real face images instead of synthetic ones with GAN fingerprints. The underlying strategy assumed that training
the model with authentic face images would enable it to learn the essential structure of real faces, thus enhancing the quality of existing fake images.

To evaluate the effectiveness of GANprintR, the authors employed three states of the art manipulation detection approaches: XceptionNet[10], Steganalysis[11], and Local Artifacts[12]. These methods were applied across three different scenarios: controlled scenarios, in – the – wild scenarios, and GAN – fingerprint removal.

In the pre – processing step, the authors removed all background information from all the images while retaining the facial regions whenever possible. To ensure unbiased results, only frontal face images were included in the analysis. As a result, the input for the detection systems consisted of images of a fixed size (224 * 224 pixels).

The study found that in controlled scenarios, where the same samples were used for both the development and evaluation of the detection models, XceptionNet exhibited excellent manipulation detection accuracies, achieving Equal Error Rate(EER) values of less than 0.5%. Steganalysis also performed well, while the Local Artifacts showed poor accuracy with an average EER of 35.5%.

Overall, Neves et al. introduced GANprintR as an effective method for evading facial manipulation detection systems by eliminating GAN fingerprints while preserving image quality. The study highlighted the varying performance of different detection approaches across different scenarios with XceptionNet demonstrating the highest accuracy in controlled settings. These findings contribute to the ongoing research and development of techniques to combat image manipulation and enhance the security of visual media.

The authors proposed a framework called ProSOUL (Propaganda Spotting in Online Urdu Language) to address the lack of data sets and LIWC (Linguistic Inquiry and Word Count) dictionaries in Urdu. To overcome this limitation, they translate the English QCRI’s propaganda data set(Qprop) and LIWC dictionary into Urdu, creating a labelled data set. The paper’s main contribution lies in the analysis of various feature sets. After preprocessing, they extract uni – gram, bi – gram, and tri – gram word and character forms. They also utilized the NELA(News Landscape) features to analyze news articles from stylistic and psycho – linguistic perspectives. Additionally, Word2Vec and BERT (Bidirectional Encoder Representations from Transformers) – Base embeddings are employed. The effectiveness of these feature sets is evaluated through several experiments. Logistic regression (LogReg) classifier is trained and assessed using n – gram (C – ngram and W – ngram), NELA, and Word2Vec features. Furthermore, a CNN(Convolutional Neural Network) model is trained using BERT embeddings. In the first experiment, the performance of different combinations of NELA, C – ngram, and W – ngram stylistic features is evaluated. In the context of Urdu text classification, the W – ngram feature achieves significantly higher accuracy, precision, recall, and F – measure values ranging from 0.90 to 0.91 compared to various combinations of NELA and C-ngram features. However, when applied to English text, the classifier’s accuracy using the top 10,000 n – grams drop to 0.65, indicating poorer results. This discrepancy can be attributed to language variations, which are simplified through machine translation. In the subsequent analysis focusing on NELA features, it is observed that the accuracy for Type – Token Ratio(TTR) is 0.70, demonstrating notable improvement compared to previous results.

The analysis reveals that the TTR feature achieves an accuracy of 0.70, demonstrating its superior performance compared to other features. This emphasizes the significance of lexical diversity in identifying propaganda content in Urdu. Regarding word embedding techniques, Word2Vec exhibits better performance in the Urdu language with an accuracy of 0.87, whereas in English text classification, BERT outperforms Word2Vec with an accuracy of 0.95. The relatively poor performance of BERT in Urdu text classification can be attributed to the limited vocabulary coverage of the multilingual BERT model of Urdu. Additionally, when evaluating ProSOUL on unseen instances, the combination of NELA, W – ngram, and C – ngram features exhibits significantly improved performance, achieving an F – measure value of 0.84 compared to other feature combinations. Finally, in another experiment involving real – world data, ProSOUL demonstrates superior classification accuracy for online news compared to general web content.

The analysis reveals that the TTR feature achieves an accuracy of 0.70, demonstrating its superior performance compared to other features. This emphasizes the significance of lexical diversity in identifying propaganda content in Urdu. Regarding word embedding techniques, Word2Vec exhibits better performance in the Urdu language with an accuracy of 0.87, whereas in English text classification, BERT outperforms Word2Vec with an accuracy of 0.95. The relatively poor performance of BERT in Urdu text classification can be attributed to the limited vocabulary coverage of the multilingual BERT model of Urdu. Additionally, when evaluating ProSOUL on unseen instances, the combination of NELA, W – ngram, and C – ngram features exhibits significantly improved performance, achieving an F – measure value of 0.84 compared to other feature combinations. Finally, in another experiment involving real – world data, ProSOUL demonstrates superior classification accuracy for online news compared to general web content.

4. FUTURE RESEARCH DIRECTIONS

Based on the survey findings, it is evident that there are specific areas that can be targeted for further enhancement of Fake News.

4.1 Blockchain Fake News Detection:

The utilization of blockchain technology in fake news detection has the potential to revolutionize the field. By leveraging the transparency and immutability of blockchain[13], it becomes possible to verify the authenticity of information, trace its sources, and establish trust in news displayed on the internet. Furthermore, a blockchain enabled platform can provide users with reliable ways to verify content, its sources, and the credibility of registered profiles, ensuring genuine connections with news agencies and journalism.

4.2 Exploring Techniques for Detecting Deep Fakes:

The field of deep fake detection presents an intriguing area for future research[14]. Current approaches rely on manually created visual. The field of deep fake detection presents an intriguing area for future research. Current approaches rely on manually created visual forensics features, which may not be suitable for detecting manipulations in real images associated with fake news. Moreover, these hand – crafted features are time – consuming to develop and often fail to capture complex patterns, resulting in limited generalization performance. Leveraging deep learning methods in_computer vision offers promise in detecting manipulated and deceptive content in multimedia data. The utilization of transfer learning and pre – trained models enhance detection performance, even when working with smaller datasets.

4.3 Generating Synthetic Training Data for Fake News Detection:

In Machine Learning, particularly deep learning, the quality of training data is crucial for algorithmic success. It should not only be sufficiently large but also accurately represent

real – world scenarios. Synthetic data generation serves as a valuable approach to enhance existing data sets, addressing challenges such as privacy concerns and imbalanced class distributions. By generating
In Machine Learning, particularly deep learning, the quality of training data is crucial for algorithmic success. It should not only be sufficiently large but also accurately represent real – world scenarios. Synthetic data generation serves as a valuable approach to enhance existing data sets, addressing challenges such as privacy concerns and imbalanced class distributions. By generating synthetic data that closely resembles the target data, researchers aim to improve prediction outcomes. Exploring user – friendly techniques to generate synthetic data, including images and videos, is an ongoing research area. Leveraging technologies like Generative Adversarial Networks (GANs) offers innovative ways to generate synthetic data for fake news detection and other applications.

4 User profile-based feature:

An essential feature for detecting fake news is the user profile – based feature[15], which provides valuable insights into profiles that spread false information. Factors such as the number of posts, account age, and follower count help identify suspicious posts. Approachability is crucial when considering user – based features, as privacy concerns restrict readily available user data and protect user interactions. Previous studies highlight the presence of bots in propagating fake news, emphasizing the importance of bot detection techniques to differentiate between bot accounts and normal users. Incorporating user profile features an essential feature for detecting fake news is the user profile – based feature, which provides valuable insights into profiles that spread false information. Factors such as the number of an essential feature for detecting fake news is the user profile – based feature, which provides valuable insights into profiles that spread false information. Factors such as the number of posts, account age, and follower count help identify suspicious posts.

4.5 Identifying and Monitoring Online Religious Content

Social media platforms face a deluge of fabricated religious content[16], strategically aimed at instilling panic among users. Various endeavors have been made to modify religious content and disseminate false information by circulating the manipulated verses through diverse media channels. Consequently, detecting counterfeit religious content, which may exist in multiple languages like Arabic, Portuguese, Persian, and Chinese, etc., emerges as a significant area for future research demands attention.

4.6 Enhancing Transparency in Multi – Model Credibility Analysis Systems through AI Explanation:

As social media systems have progressed, the distinction between fake news and accurate and verifiable information has become increasingly blurred. Consequently, an important area of research lies in the development of AI explainable credibility analysis systems specifically designed to combat fake news on social media. Traditional manual censor boards are insufficient and lack the capacity to handle the vast scale of news content, making it necessary to explore automated approaches. An existing example of such automated framework focuses on health blogs, providing credibility scores for the author, text, and image of the blog, along with user – friendly explanations. These framework systems can be further expanded to encompass other domains such as politics, enabling a comprehensive approach to combat misinformation[17].

5. Conclusion

With the rapid development of the digital realm, the prevalence of online counterfeit content has seen a significant surge. With the rapid development of the digital realm, the prevalence of online counterfeit content has seen a significant surge. This proliferation of fake information easily accessible to the public poses a substantial threat to journalism and democracy by misleading people. False content, spreading rapidly and leaving a profound impact, has the power to sway public media users lack awareness on specific subjects, making them susceptible to deception through counterfeit content. Additionally, people's reliance on online media platforms and the captivating nature of fake news contribute to its With the rapid development of the digital realm, the prevalence of online counterfeit content has seen a significant surge. With the rapid development of the digital realm, the prevalence of online counterfeit content has seen a significant surge. This proliferation of fake information easily accessible to the public poses a substantial threat to journalism and democracy by misleading people. False content, spreading rapidly and leaving a profound impact, has the power to sway public media users lack awareness on specific subjects, making them susceptible to deception through counterfeit content. Additionally, people's reliance on online media platforms and the captivating nature of fake news contribute to its widespread dissemination. While ongoing research endeavors to detect fake news have made progress, various methodologies from diverse fields such as artificial intelligence, linguistic analysis, and knowledge engineering have been employed. However, no ideal approach has been devised to accurately classify real and fake news. The primary challenge lies in the overwhelming volume of social media content, which grows exponentially every day. The unfiltered nature of content on social networking sites, including subjective opinions and unchecked information, presents a daunting task in developing effective fake news detection algorithms. This paper conducts an extensive examination of existing techniques for detecting fake news, presents a new taxonomy, discusses the major challenges in this field, and provides recommendations for future enhancements.

REFERENCES

1) A. Waqar, A. Raza, H. Abbas, and M. K. Khan, " A framework for preservation of cloud users' data privacy using dynamic reconstruction of metadata," *Journal of Network and Computer Applications*, Vol. 36, no. 1, pp. 235 – 248, 2013.
2) F. Wang, D. Jiang, H. Wen, and S, Qi; 'Security level protection for intelligent terminals based on different privacy, " *Telecommunication systems*, vol. 74, no. 4, pp. 425 – 435, 2020.
3) S. Kumari, M. K. Khan, and R. Kumar, "Cryptanalysis and improvement of ' a privacy enhanced scheme for telecare medical information systems'," *Journal of* S. Kumari, M. K. Khan, and R. Kumar, "Cryptanalysis and improvement of ' a privacy enhanced scheme for telecare medical information systems'," *Journal of medical systems*, vol. 37, no. 4, pp. 1 – 11, 2013
4) K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, " Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22 – 36, 2017.

5) Fake news and its impact on the economy,[online]. Available: https://priorityconsultants.com/blog/fake-news-and-its-impact-on-the-economy/

6) N. Grinberg, K. Joseph, L. Friedland, B. Swire Thompson, and D. Lazer, "Fake news on twitter during the 2016 US presidential election," *Science*, vol. 363, no. 6425, pp. 374 – 378, 2019.

7) S. Kausar, B. Tahir, and M. A. Mehmood, "ProSOUL: A framework to identify propaganda from online Urdu content," *IEEE Access,* Vol. 8, PP. 186 039 – 186 054,

8) Z. Zhao, J. Zhao, Y. Sano, O. Levy, H. Takayasu, M. Takayasu, D. Li , J. Wy, and S. Havlin , "Fake news propagates differently from real news even at early stages of spreading," *EPJ Data Science*, Vol. 9, no. 1, p. 7, 2020.

9) J. Neves, R. Tolosana, R. Vera-Rodriguez, V. Lopes, H. Proenca, and J. Fierrez, "GANprintR: Improved fakes and evaluation of the state of the art in face manipulation detection," pp. 1038 – 1047, 04 2020.

10) F. Chollet, "Xception: Deep learning with depth wise separable convolutions," 07 2017, pp. 1800 – 1807.

11) L. Nataraj, T. M. Mohammed, B. Manjunath, S. Chandrasekaran, A. Flenner, M. J. Bappy, and A. Roy Chowdhury, " Detecting gan generated fake images using co-occurrence matrices," *Electronic Imaging,* vol. 2019, pp. 532 – 1, 01 2019.

12) F. Matern, C. Riess, and M. Stamminger, "Exploiting visual artifacts to expose deepfakes and face manipulations," 01 2019, pp. 83 – 92.

13) S. Paul, J. I. joy, S. Sarker, A. . A. . H. Shakib, S. Ahmed, and A. K. Das, "Fake news detection in social media using blockchain," in 2019 *7th International conference on Smart Computing Communications (ICSCC),* 2019, 99. 1 – 5.

14) A. Zervopoulos, A. G. Alvanou, K. Bezas, A. Papamichail, M. Maragoudakis, and K. kermanidis, " Deep learning for fake news detection on twitter regarding the 2019 Hong Kong protests," *Neural Computing and Applications,* vol. 34, no. 2, pp. 969 – 982, 2022.

15) S. Agarwal and A. Samavedhi, "Profiling fake news: Learning the semantics and characterization of misinformation," in *International Conference on Advanced Data Mining and Applications.* Springer, 2011, pp, 203 – 216.

16) S. Hakak, A. Kamsin, O. Tayan, M. Y. I. Idris, A. Gani, and S. Zerdoumi, "Preserving content integrity of digital Holy Quran: Survey and open challenges," *IEEE Access,* vol. 5, pp. 7305 – 7325, 2017.

17) V. Wagle, K. Kaur, P. Kamat, S. Patil, and K. Kotecha, "Explainable AI for multi – model credibility analysis: Case study of online beauty health (mis) – information," *IEEE Access*, vol. 9, pp. 127 985 – 128 022, 2021.

18) RN Kumar, V. Bapuji, A. Govardhan, S. Sharma, "Soft Computing and Artificial Intelligence Techniques for Intrusion Detecting Systems", Network and Complex systems, 2012.