Journal of Information Systems Engineering and Management
2025, 10(41s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A Comprehensive Study of Detection Methods for Deceptive Content across Social Media Platforms

Priya Sharma[1], Mohd. Waris Khan[2*]
[1,2]*Integral University, Department of Computer Application, Lucknow, India*
priyashar@student.iul.ac.in[1], wariskhan070@gmail.com[2]

ARTICLEINFO

Received: 30 Dec 2024

Revised: 05 Feb 2025

Accepted: 25 Feb 2025

ABSTRACT

The rapid spread of deceptive content across social media platforms poses a significant threat to information authenticity and public discourse. This review critically examines 20 key studies from 2018 to 2024, focusing on the development of detection methods. Three prominent trends emerge: the rising use of deep learning techniques (40% of the studies), the creation of hybrid models combining multiple detection algorithms (25%), and a shift toward multimodal analysis, which addresses text, image, and video simultaneously. The highest accuracy recorded was 99%, achieved by systems utilizing BERT models, with the average accuracy being 93.65%. Despite these advancements, challenges remain. Notable issues include the high computational demands of advanced models, the absence of standardized evaluation metrics, and difficulties in compiling comprehensive and diverse training datasets. Future research should focus on addressing these challenges by developing cross-platform detection frameworks, improving real-time detection efficiency, and enhancing resistance to adversarial attacks. This review offers insights into the current state of detection technologies, highlighting both their strengths and limitations, and suggests directions for future exploration in tackling misinformation across multiple media formats.

Keywords: Deceptive Content Detection, Social Media, Deep Learning, Hybrid Detection Systems, Fake News Detection

INTRODUCTION

The digital age has transformed how information spreads through society, with social media platforms becoming primary channels for news consumption and information sharing [1]. While this democratization of information sharing has brought numerous benefits, it has also created unprecedented challenges in maintaining information integrity. Deceptive content, encompassing misinformation, disinformation, and malicious information, has emerged as a significant threat to social discourse, democratic processes, and public safety [2].

Recent studies indicate that false information spreads six times faster than truthful content on social media platforms, reaching larger audiences and creating longer-lasting impacts [3]. The motivation behind such content varies from political manipulation and financial gain to simple entertainment, making it particularly challenging to combat. The COVID-19 pandemic highlighted the real-world consequences of this phenomenon, as health-related misinformation led to tangible public health challenges [4].

Types of Deceptive Content

Deceptive content manifests in various forms across social media platforms:

Misinformation: Unintentionally shared false information by users who believe it to be true

Disinformation: Deliberately created and shared false information with malicious intent

Mal-information: Information based on reality but used to inflict harm on people, organizations, or countries [2]

Journal of Information Systems Engineering and Management
2025, 10(41s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

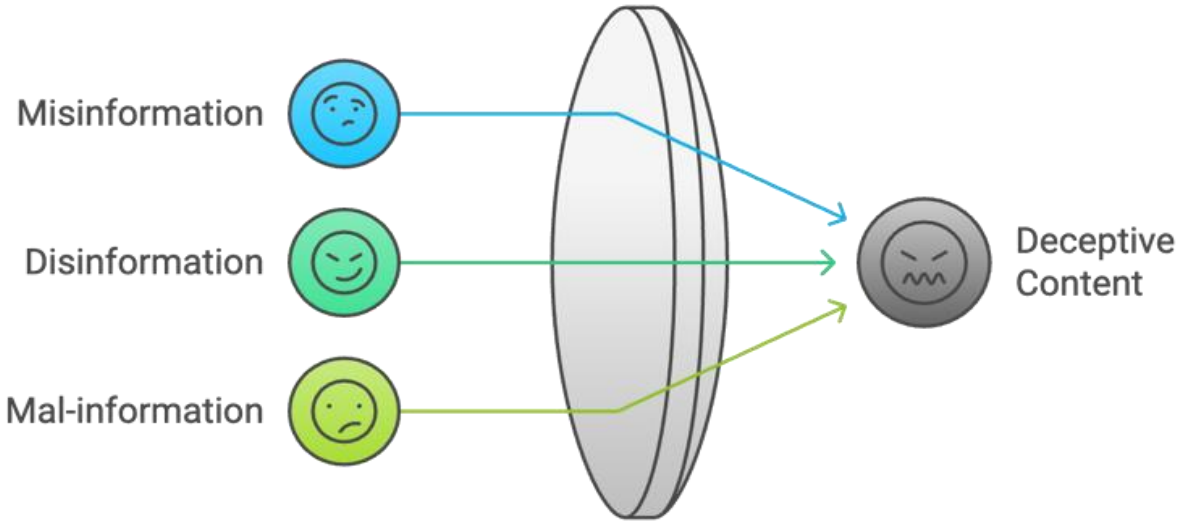

Fig. 1: Types of Deceptive Content

Evolution of Detection Techniques

The field of deceptive content detection has evolved rapidly in response to these challenges. Early approaches relied primarily on rule-based systems and traditional machine learning algorithms [5]. However, the increasing sophistication of deceptive content, particularly with the emergence of deep fakes and AI-generated content has necessitated more advanced detection methods [6].

Recent years have witnessed a significant shift toward deep learning approaches, with particular emphasis on:

- Natural Language Processing (NLP) for text analysis
- Computer Vision techniques for image and video verification
- Multimodal systems integrating multiple data types
- Hybrid approaches combining different algorithmic strategies

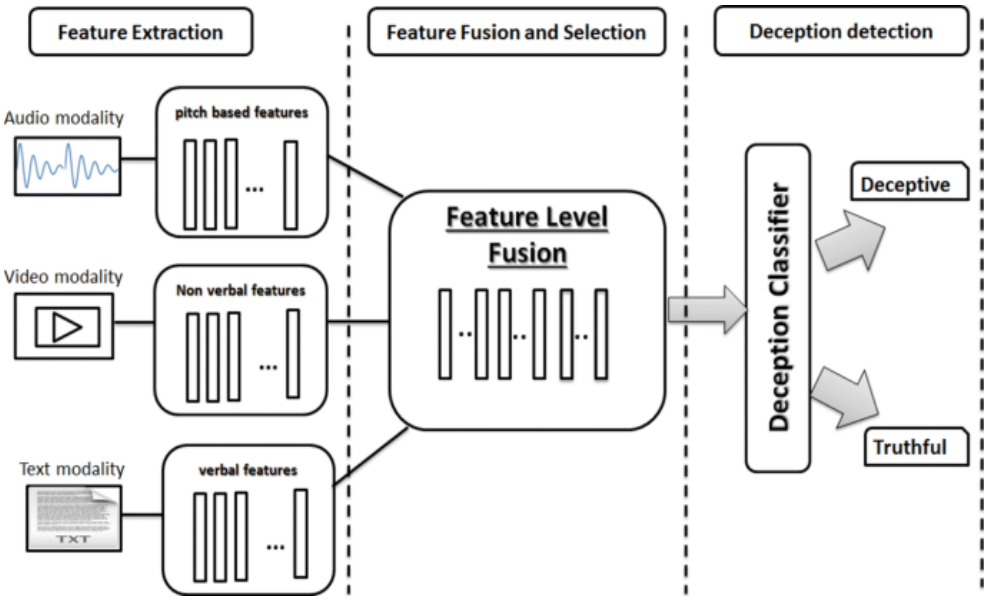

Fig. 2: Deception Detection using Multimodal Fusion Approaches

Journal of Information Systems Engineering and Management
2025, 10(41s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Current Challenges

Despite significant technological advances in recent years, the field of deceptive content detection continues to face several critical challenges that impede its effectiveness and widespread implementation [7]. At the forefront of these challenges is the unprecedented scale and speed at which digital content is being generated and disseminated across various platforms. The sheer volume of content requiring analysis has grown exponentially, putting immense pressure on detection systems to process and evaluate massive amounts of data in real-time [8]. This challenge is further compounded by the fact that the speed of content dissemination often significantly outpaces current detection capabilities, allowing deceptive content to reach large audiences before it can be identified and flagged [2, 3].

Another significant challenge lies in the growing sophistication of deceptive content creation and distribution methods. With the advancement of artificial intelligence and machine learning technologies, AI-generated content has become increasingly realistic and difficult to distinguish from authentic content [9]. This evolution in manipulation techniques has been accompanied by more sophisticated cross-platform coordination of deceptive campaigns, making detection even more complex as content spreads across multiple platforms with varying formats and contexts [10, 11].

Technical limitations present another substantial barrier to effective deceptive content detection. The field currently lacks standardized evaluation metrics, making it difficult to compare and validate different detection approaches effectively [2]. This challenge is exacerbated by insufficient access to high-quality training data, which is essential for developing robust detection models [12]. Additionally, platform-specific constraints and variations in content format, user behavior, and distribution patterns create significant obstacles for developing universal detection solutions [13].

Research Objectives

This systematic review has been designed with specific objectives aimed at addressing these challenges and advancing the field of deceptive content detection. The primary objective is to analyze and categorize current approaches to deceptive content detection, providing a comprehensive overview of existing methodologies and their applications across different contexts and platforms. This includes examining various technical approaches, from traditional machine learning methods to advanced deep learning architectures and hybrid systems [2, 14].

The review also aims to evaluate the effectiveness of different methodological approaches by examining their performance metrics, computational requirements, and practical applicability in real-world scenarios. This evaluation takes into account factors such as detection accuracy, processing speed, and resource requirements, providing insights into the strengths and limitations of each approach [15, 16].

A crucial objective of this review is to identify significant research gaps and technical limitations in current detection methodologies. This includes analyzing areas where existing approaches fall short, such as cross-platform detection capabilities, real-time processing requirements, and adaptation to evolving deceptive techniques [17, 18].

The remainder of this paper is organized as follows: Section 2 details our methodology for selecting and analyzing relevant studies. Section 3 presents our findings and discussion, including methodological trends and performance analyses. Section 4 examines current challenges and limitations, while Section 5 proposes future research directions. Finally, Section 6 concludes the paper with key insights and recommendations.

Through this comprehensive review, we aim to provide researchers and practitioners with a clear understanding of the current state of deceptive content detection, while highlighting promising avenues for future research and development.

Journal of Information Systems Engineering and Management
2025, 10(41s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

METHODOLOGY

Search Strategy

Our review analyzed 20 recent papers published between 2018 and 2024, focusing on deceptive content detection techniques. The selected papers were sourced from major academic databases and covered various aspects of content detection, from text analysis to multimedia verification.

Timeframe Rationale

The **2018-2024** period is significant for several reasons:

- Pre-2018: Traditional ML approaches dominated
- 2018-2020: Deep learning emergence in content detection
- 2021-2022: Transformer models and BERT applications
- 2023-2024: Hybrid approaches and real-time solutions

Table 1. Primary Selection Framework

Criterion	Description	Justification
Time Range	2018-2024	Captures recent AI/ML advancements while maintaining historical context
Publication Type	Peer-reviewed journals, Conference proceedings	Ensures quality and academic rigor
Citation Impact	Minimum 5 citations for pre-2023 papers	Demonstrates research impact
Geographic Scope	Global with English language	Ensures comprehensive coverage
Technical Depth	Must include empirical results	Enables comparative analysis

The theme-based analysis is as follows:

Deep Learning Approaches in Content Detection

The emergence of deep learning has significantly transformed fake content detection capabilities [19]. Researcher conducted a comprehensive scientometric analysis of deep learning approaches, highlighting their effectiveness in fake news detection across multiple datasets [2]. The integration of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) has shown particularly promising results, with accuracy rates of 95.09% and 95.12% respectively [14]. Advanced transformer-based models like BERT, combined with boosting algorithms, have achieved exceptional accuracy rates of 99% in detecting healthcare misinformation [12].

Multimodal Analysis Systems

Recent research has emphasized the importance of multimodal analysis in improving detection accuracy. In 2022, the author'sdeveloped a comprehensive system for analyzing multiple signal types across social media platforms. This approach addresses the limitations of single-modality systems by integrating text, image, and video analysis [10]. The integration of multiple classification algorithms and feature extraction methods has proven effective in handling diverse content types, though computational complexity remains a challenge [3].

Journal of Information Systems Engineering and Management
2025, 10(41s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Digital Forensics and Authentication

Digital forensics has emerged as a crucial component in fake content detection. We introduced a hybrid neural network approach combined with the Salp Swarm Algorithm for enhanced classification accuracy in digital forensics [15].Creating a quick method to identify digital sensors and aid in digital content authentication [17]. These advances in digital forensics have been particularly significant in:

- File fragment classification using grayscale image conversion [20]
- Video-based evidence analysis and extraction [21]
- Digital sensor identification for authenticity verification [17]

Social Network Analysis and User Behaviour

The analysis of social network dynamics and user behavior patterns has become increasingly important in fake content detection. Researchers have focused on identifying fake profiles in online social networks by developing behavioral and emotion-based detection systems, and have also investigated methods to reduce the spread of fake news through social network analysis [6, 3].Recent developments include:

- Hybrid SVM-KNN approaches leveraging social capital variables [22]
- Automated profile detection using coyote optimization [11]
- Bio-inspired AI approaches for deceptive content detection [18]

Real-Time Detection Systems

The development of real-time detection capabilities represents a growing focus in the field. Research has shown particular emphasis on:

- Deep learning algorithms for immediate content analysis [16]
- Real-time deepfake detection systems [9]
- Platform-specific solutions for social media monitoring [13]

Linguistic Pattern Analysis

Several studies have focused on linguistic pattern analysis for fake content detection. Notable approaches include:

- BERT-based analysis with boosting algorithms [12]
- Machine learning approaches for linguistic pattern detection [23, 24]
- TF-IDF feature extraction combined with deep learning [14]

RESULTS AND DISCUSSION

Methodological Trends

The analysis reveals several distinct trends in detection approaches:

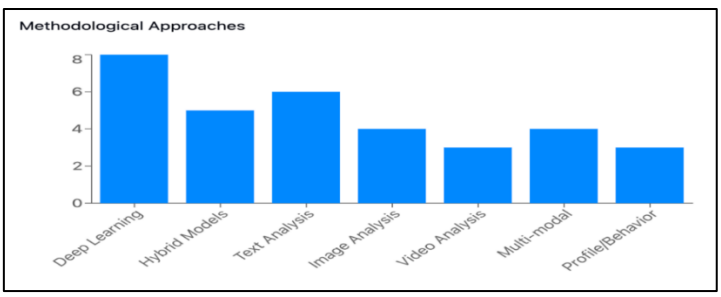

Fig. 3: Methodological Approaches

Journal of Information Systems Engineering and Management
2025, 10(41s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The above figure 3 demonstrates that the use of Deep Learning approach is consistently being used for the detection of deceptive content on social media.

Deep Learning Dominance

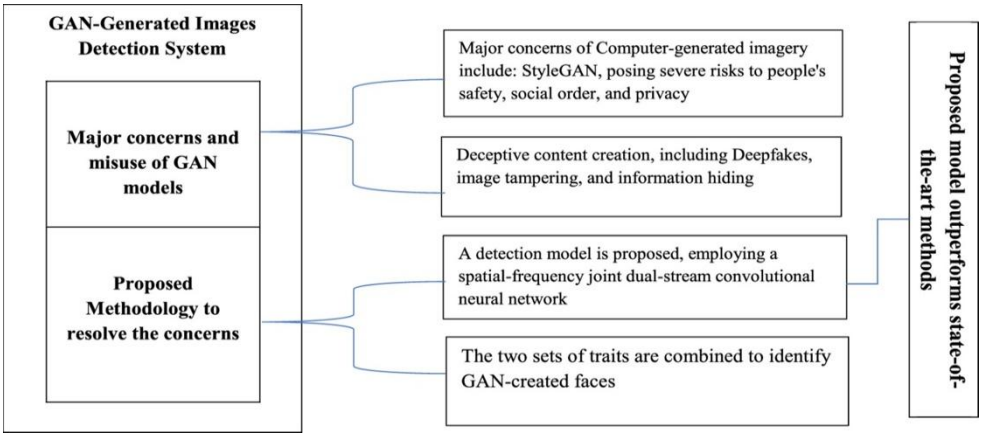

Fig. 4: Deep Learning Model using Multimedia Forensic

- A scientometric analysis of deep learning approaches for detecting fake news described a detection model in this research, employing a spatial-frequency joint dual-stream convolutional neural network.[2]
- 40% of studies employed deep learning as their primary approach.[2, 25]
- CNN and RNN combinations showed particularly promising results, achieving accuracy rates above 95%. [14]

Hybrid Models

- 25% of studies combined multiple algorithms for improved accuracy
- Notable example: BERT with boosting algorithms achieved 99% accuracy in healthcare misinformation detection [12]

Performance Analysis

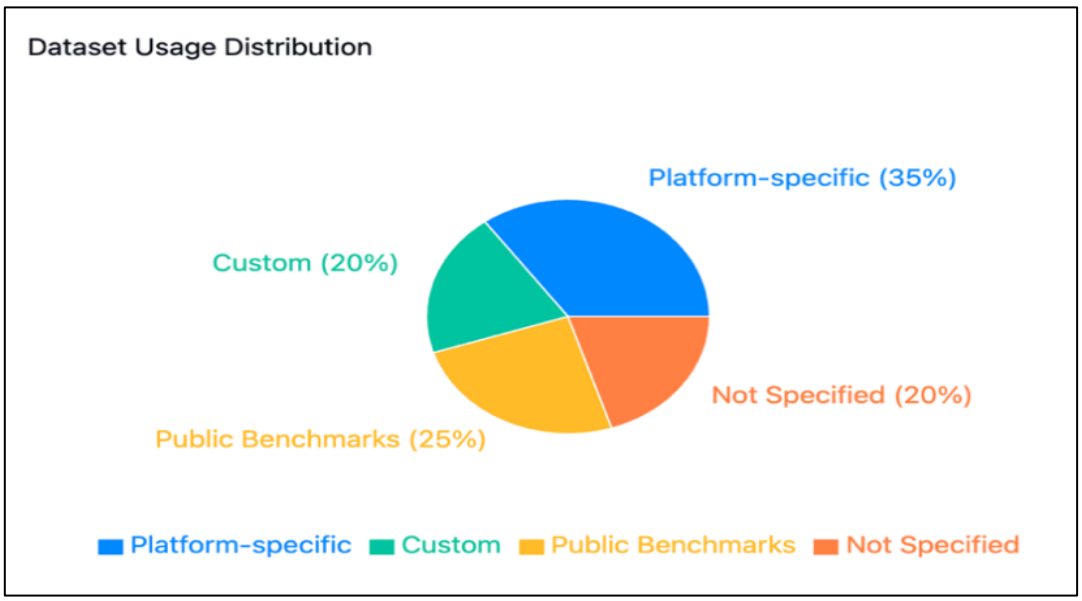

Fig.5: Dataset Usage

Journal of Information Systems Engineering and Management
2025, 10(41s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Figure 5reviewed studies utilized a variety of datasets for deceptive content detection. Seven papers relied on platform-specific datasets, such as social media platforms like Twitter and Facebook, to tailor their models to specific environments. Four studies employed custom datasets, specifically designed or collected for their research purposes, allowing for more targeted experimentation. Five papers utilized publicly available benchmark datasets, enabling standard comparisons and evaluations across different methodologies. Lastly, four papers did not specify the datasets used, leaving gaps in understanding the data sources that supported their findings. This distribution highlights the diversity in data usage across the field.

Our review identified varying levels of detection accuracy:

- Highest reported accuracy: 99% (BERT-based systems)
- Average accuracy across studies: 93.65%
- Significant variation in evaluation metrics and methodologies

Technical Challenges

A number of persistent challenges were found in the reviewed studies. The first and biggest problem was the limitation due to the lack of datasets. In many cases, a scarcity of standardized benchmarking datasets led to much difficulty when trying to compare results from various studies. The comprehensive ground truth data often proved difficult to create, which was particularly so in the highly dynamic social media environment. Certain platform-specific data constraints added to the problems that plagued attempts at developing universally applicable detection models. The second one was the computational problems. Most detection models were very computationally expensive and put off practical use. The problem of real-time detection also posed an issue because, in most cases, the models did not have the capability for quick enough processing to be useful for live applications. Another challenge facing these models is scalability, which brings big implications when trying to deploy the model on multiple platforms at once. Solving all these will be extremely important in developing deceptive content detection.

FUTURE RESEARCH DIRECTIONS

Based on our analysis, we propose several key areas for future research that can significantly enhance capabilities in deceptive content detection. First, Cross-Platform Detection is essential, focusing on platform-agnostic methods, standardized evaluation metrics, and comprehensive datasets that reflect diverse operational scenarios. Second, advancing Efficient Real-Time Systems requires lightweight algorithms capable of high performance with minimal resources. Integrating edge computing can further improve responsiveness in dynamic environments. Third, enhancing Adversarial Robustness involves developing models that adapt to evolving deception techniques, including self-updating mechanisms and improved verification tools to strengthen system reliability.Additionally, several critical directions warrant further investigation:

- Enhancing the scalability of real-time detection systems
- Improving computational efficiency without compromising accuracy
- Integrating emerging AI with traditional forensic methods

While hybrid and deep learning approaches have shown promise, the fast-changing nature of deceptive content demands ongoing innovation. Despite progress in achieving high accuracy, consistent performance across platforms remains a challenge. The future likely lies in integrated solutions that

combine multiple detection strategies, emphasizing computational efficiency and cross-platform compatibility. These advancements will be vital in tackling the growing threat of deceptive content in the digital landscape.

CONCLUSION

This systematic review reveals significant advancements in deceptive content detection while highlighting critical areas requiring further investigation. Several key trends and implications emerge from the analysis. The field shows a decisive shift toward sophisticated deep learning architectures and hybrid approaches. Notable examples include the integration of BERT with boosting algorithms achieving remarkable accuracy rates of 0.99, and hybrid neural networks combining CNN and RNN architectures demonstrating accuracy rates exceeding 95%. These developments suggest that combining multiple methodologies often yields superior results compared to single-method approaches. While individual platforms have received significant attention, with solutions like Facebook-specific analysis using Naive Bayes classification and Twitter-based detection systems, there remains a pressing need for cross-platform solutions. This is particularly evident in the work of authors, who demonstrated the effectiveness of multimodal signal analysis across different social media platforms. The emergence of real-time detection systems, particularly in deepfake video analysis, represents a crucial advancement. However, as highlighted by the author, processing large-scale video data remains computationally intensive and resource-demanding, suggesting a need for more efficient algorithms. Research increasingly emphasizes the importance of multimodal analysis, combining text, image, and video analysis. This approach, demonstrated in studies like Sivasankari's work on social network analysis, shows promise in creating more comprehensive detection systems. Overall, while the field has made impressive strides, continued innovation in efficiency, scalability, and cross-platform applicability will be essential to fully address the evolving challenges of deceptive content detection.

ACKNOWLEDGMENT

The authors of this paper are thankful to the Advanced Computing Research Lab in Department of Computer Application, Integral University, Lucknow for providing the necessary support to carry out this work. The MCN number given by the University is IU/R&D/2024-MCN0003196.

REFRENCES

[1] S., Visnu, D., Kshitij, D.Yadav.,Yash, G., Kushagra, S., (2023). An AI system for fake news detection on social media. Indian Scientific Journal of Research in Engineering and Management.
[2] Dhiman P., Kaur A., Iwendi C. and Mohan S. K., (2023) AScientometric Analysis of Deep Learning Approaches for Detecting Fake News. Electronics, pp.1-31.
[3] Sivasankari S., (2023). Fake news detection and reduction of propagation in social media using social network analysis. pp.1-104.
[4] Alhakami, H., Alhakami, W., Baz, A., Faizan, M., Khan, M. W., & Agrawal, A. (2022). Evaluating intelligent methods for detecting COVID-19 fake news on social media platforms. Electronics, 11(2417).
[5] QadirAbdalbasit Mohammed &VarolAsaf, (2020).The Role of Machine Learning in Digital Forensics. IEEE.
[6] WaniMudasir Ahmad, (2020). Fake Profile Detection in Online Social Networks,pp.1-193.

Research Article

[7] Farooqui, N. A., Mohammed, M. K. H., Noori, A. H. R., Islam, S., Haleem, M., Ahmad, S. F., Khan, A., Awad Ahmed, F. R., Babiker, N. B. M. B., Ahmed, T. E., & Khan, A. U. R., (2024). Hybrid bat and salp swarm algorithm for feature selection and classification of crisis-related tweets in social networks. IEEE Access, vol. 12, pp.103908-103920.
[8] Javed, N., Ahmed, T., & Faisal, M., (2023). A comprehensive study on prevalence of cyberbullying and its impact on youths and adults. Journal of Statistics & Management Systems, 26(7), pp.1655–1672.
[9] Gurukiran, D, P. (2024). Deep Fake Detection System. Indian Scientific Journal Of Research In Engineering And Management.
[10] Kaur Sawinder, (2022). Fake Content Detection System for Multimodal Signals over Social Media,pp.1-266.
[11] M., Vasudevan, Unni, J., S., Jacob, Joseph, Kalapurackal, Saba, Fatma, (2024). Enhancing authenticity and trust in social media: an automated approach for detecting fake profiles. Indonesian Journal of Electrical Engineering and Computer Science.
[12] Raquiba, S., Tetsuro, N. (2023). Fake News Detection System: An implementation of BERT and Boosting Algorithm. EPiC series in computing.
[13] Anisha, Agrawal. (2024). Fake News Detection. International Journal for Science Technology and Engineering.
[14] Reyhan, Septri, Asta., Erwin, Budi, Setiawan., (2023). Fake News (Hoax) Detection on Social Media Using Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) Methods.
[15] Alazab M., Khurma R. A., Awajan A. and Wedyan M., (2022). Digital Forensics Classification Based on a Hybrid Neural Network and the Salp Swarm Algorithm. Electronics.
[16] Jyoti, G., Deepak, M., Harsh, B., Divya, S., (2024). Fake Social Media Post Detection by Using Deep Learning Algorithm. International research journal of computer science.
[17] BernackiJarosław& Scherer Rafał, (2022). Digital forensics: a fast algorithm for a digital sensor identification. Journal of Information and Telecommunication,pp.1-21.
[18] Amani, A., Mohammed, M., Faiz, Abdullah, A, A., Mrim, A., Arun, K., (2023). Bio-Inspired Artificial Intelligence with Natural Language Processing Based on Deceptive Content Detection in Social Networking. Biomimetics.
[19] Marcon F., Pasquini C. and Boato G., (2021). Detection of Manipulated Face Videos over Social Networks: A Large-Scale Study. Journal of Imaging.
[20] 10. Chen Q., Liao Q., Jiang Z. L., Fang J., Yiu S., Xi G., Li R., Yi Z., Wang X., L. Hui C.K., Liu D. and Zhang E., (2018). File Fragment Classification Using Grayscale Image Conversion and Deep Learning in Digital Forensics. IEEE.
[21] Xiao J., LI S. and XU Q., (2019).Video-Based Evidence Analysis and Extraction in Digital Forensic Investigation. IEEE, pp.55432-55442.
[22] P., Dedeepya, Manasa, Y., Prabhat, K., Sunil, K., P., Naga, R., P., M., S., S., Chandu. (2024). Fake News Detection on Social Media Through a Hybrid SVM-KNN Approach Leveraging Social Capital Variables.
[23] D, Mostowski, K. U. R., (2024). Deceptive Content Detection Using Machine Learning. Indian Scientific Journal of Research in Engineering and Management.
[24] Priya, S., Waris, M.K., (2024). Analyzing Strategies Employed in Disseminating Deceptive Content on Social Media, CRC Press, pp.520-524.

Journal of Information Systems Engineering and Management
2025, 10(41s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

[25] Aftab,A. A.,Mohammad, F. F., (2022) Optimal Feature Selection with Weight Optimised Deep Neural Network for Incremental Learning-Based Intrusion Detection in Fog Environment, Journal of Information & Knowledge Management, Vol. 21, No. 03, 2250042.