International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05, May 2026
Received: 19-03-2026 Accepted: 24-04-2026 Published: 01-05-2026

# Automated news authenticity verification

Miss : Mahanthi Nandini
Department : Master of Computer Applicaions
College : Satya Institute of Technology and Managementline
City : Vizianagaram
email : mahanthinandini2004@gmail.com

Guide : Mr. K.Yasoda Krishna
Department : Master of Computer Applicaions
College : Satya Institute of Technology and Management
City : Vizianagaram
email : k.yasodakrishna@sitam.co.in

***Abstract*****—** Fake news has emerged as a critical challenge in the digital information ecosystem, particularly with the rapid growth of social media platforms, online news portals, and instant messaging applications. The speed at which information spreads has outpaced traditional verification mechanisms, allowing misleading, false, or deliberately fabricated news to influence public opinion, social harmony, and even democratic processes. Fake News Detection Using Natural Language Processing (NLP) focuses on leveraging computational linguistic techniques and machine learning models to automatically identify and classify news content as genuine or fake based on textual patterns, semantics, and contextual cues. NLP enables systems to process large volumes of unstructured text, extract meaningful features such as syntax, sentiment, and discourse structure, and learn hidden relationships that distinguish authentic journalism from deceptive narratives. By integrating preprocessing, feature extraction, and classification algorithms, NLP-based fake news detection systems aim to provide scalable, real-time solutions that reduce human dependency and improve information credibility. This research emphasizes the importance of linguistic analysis, data-driven learning, and automation to combat misinformation, highlighting how NLP can act as a powerful tool in ensuring trustworthy digital communication and safeguarding society from the harmful impacts of fake news.

**Keywords— SDN, CNN,RNN,DL,1dCNN, GRU, LSTM, SDCNN**

## I. Introduction

The digital age has transformed how information is created, shared, and consumed, making news accessible to a global audience within seconds. While this accessibility has democratized information, it has also enabled the widespread dissemination of fake news, which includes fabricated stories, misleading headlines, manipulated facts, and biased narratives presented as legitimate news[1],[2],[3]. Fake news can cause serious consequences such as social unrest, erosion of trust in media, financial manipulation, and political instability. Manual fact-checking methods are often slow, labor-intensive, and insufficient to handle the massive volume of online content generated daily. As a result, automated approaches have become essential for detecting fake news efficiently. Natural Language Processing plays a vital role in this context by enabling machines to understand, interpret, and analyze human language. NLP techniques such as tokenization, part-of-speech tagging, semantic analysis, and sentiment detection allow systems to capture linguistic patterns that are commonly associated with deceptive or unreliable content. Combined with machine learning classifiers, NLP-based systems can learn from historical data and continuously improve detection accuracy. This introduction establishes the relevance of NLP in addressing fake news, emphasizing its role in automating content analysis, enhancing reliability, and supporting informed decision-making in the digital information landscape[4],[5],[6].

### A.Scope of Research

The scope of this research encompasses the application of Natural Language Processing techniques for detecting fake news in textual data sourced from online news articles and social media platforms. The study focuses on analyzing the linguistic structure, semantic meaning, and contextual patterns present in news content to identify deceptive information. It includes preprocessing steps such as text cleaning, normalization, and tokenization, followed by feature extraction methods like term frequency, word embeddings, and sentiment analysis. The research also explores the integration of machine learning algorithms to classify news as fake or real based on extracted features. While the primary emphasis is on textual analysis, the scope allows for future extensions involving multimedia content, user behavior analysis, and network-based features. The research is limited to the English language dataset, ensuring consistency in linguistic analysis, but the methodology can be adapted to other languages. This study aims to contribute to academic research and practical applications by providing insights into effective NLP techniques, performance evaluation metrics, and system design considerations for fake news detection in real-world environments[10],[11],[12].

### B.Disadvantages of Existing System

The existing fake news detection systems suffer from several critical limitations that reduce their practical



2250-3676 ijesatj@gmail.com https://www.ijesat.com/

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05, May 2026
Received: 19-03-2026 Accepted: 24-04-2026 Published: 01-05-2026

effectiveness. Manual verification processes are time-consuming and cannot handle the high volume of news generated daily, resulting in delayed responses that allow misinformation to spread unchecked. Rule-based and keyword-driven systems lack contextual awareness and often misclassify legitimate news that contains flagged terms. Traditional machine learning models relying on shallow features fail to capture semantic relationships and deeper linguistic cues, leading to poor generalization across diverse datasets. Many systems are static in nature and do not adapt well to evolving fake news strategies, making them vulnerable to manipulation. Additionally, over-reliance on source-based credibility assessments ignores the actual content quality and can result in biased outcomes. These disadvantages highlight the need for a more intelligent, flexible, and language-aware approach to fake news detection.

C.Proposed System

The proposed system introduces a comprehensive fake news detection framework using Natural Language Processing combined with advanced machine learning techniques. This system focuses on content-based analysis, enabling automated understanding of news text at both syntactic and semantic levels. The proposed approach begins with robust text preprocessing to remove noise, normalize language, and prepare data for analysis. NLP techniques such as tokenization, lemmatization, sentiment analysis, and word embeddings are employed to extract rich linguistic features that capture meaning, tone, and writing style. These features are then fed into machine learning classifiers capable of learning complex patterns associated with fake and real news. The system is designed to be scalable, allowing it to process large datasets in real time, and adaptive, enabling continuous learning from new data. By emphasizing language understanding rather than surface-level indicators, the proposed system aims to improve detection accuracy and reliability while reducing human intervention.

## II. SYSTEM ANALYSIS AND DESIGN.

A. System analysis and design

The system analysis phase for Fake News Detection using Natural Language Processing (NLP) focuses on understanding the problem domain, identifying functional requirements, evaluating data characteristics, and determining technical feasibility. Fake news has emerged as a critical challenge due to the rapid growth of digital media platforms, where misleading or intentionally false information spreads faster than verified news. Manual verification methods are time-consuming, subjective, and impractical at scale, creating the need for an automated, intelligent system capable of analyzing large volumes of textual content accurately and efficiently.

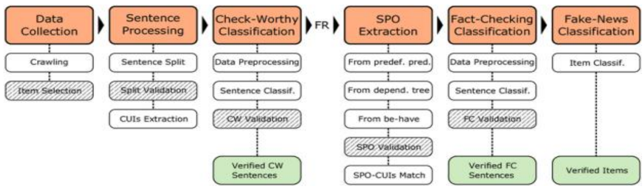
From a functional perspective, the system must accept news articles, headlines, or social media text as input and determine whether the content is genuine or fake. This requires the system to analyze linguistic patterns, semantic meaning, writing style, and contextual cues present in the text. Non-functional requirements include high accuracy, low response time, scalability for large datasets, robustness against noisy or incomplete data, and adaptability to evolving fake news patterns. The system should also be language-independent or easily extendable to multiple languages with minimal modifications.

B. System Architecture

The Fake News Detection system is designed as a modular and layered architecture to ensure efficient data flow, scalability, and maintainability. It consists of key components including data acquisition, preprocessing, feature extraction, model training, classification, and result visualization.

The process begins with collecting news data from datasets or user input. The data is then preprocessed using NLP techniques such as tokenization, stop-word removal, and lemmatization to clean and standardize the text. Next, feature extraction methods like TF-IDF or Bag of Words convert the text into numerical representations.

These features are fed into machine learning or deep learning models that learn to classify news as fake or real. Finally, the system outputs the prediction along with confidence scores for better interetation. This architecture ensures flexibility, accuracy, and ease of future enhancements.




## III. METHODOLOGY

The methodology for fake news detection using NLP follows a systematic pipeline that integrates text processing, feature extraction, and machine learning classification. Initially, the preprocessed dataset is divided into training and testing sets to enable unbiased performance evaluation. The textual content is then transformed into numerical representations using NLP techniques such as Bag of Words, Term Frequency–Inverse Document Frequency (TF-IDF), or word embeddings. These representations capture the semantic and contextual information present in the news articles.

Once the features are extracted, machine learning algorithms such as Naive Bayes, Support Vector Machines, Logistic



2250-3676 ijesatj@gmail.com https://www.ijesat.com/

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05, May 2026
Received: 19-03-2026 Accepted: 24-04-2026 Published: 01-05-2026

Regression, or deep learning models like Long Short-Term Memory networks are trained on the labeled training data. During training, the model learns patterns and linguistic cues that distinguish fake news from real news, such as sensational language, misleading phrases, or abnormal word distributions. The trained model is then evaluated using unseen test data to assess its generalization capability. Performance metrics such as accuracy, precision, recall, and F1-score are computed to validate the effectiveness of the methodology. This structured approach ensures that the system can automatically identify fake news with high reliability while remaining scalable for large datasets.

A. **Feature Selection Techniques**

Feature selection plays a vital role in fake news detection by identifying the most relevant textual attributes that contribute to accurate classification. In NLP-based systems, textual data often results in high-dimensional feature spaces, which can increase computational complexity and lead to overfitting. Therefore, feature selection techniques are applied to reduce dimensionality while preserving meaningful information. One commonly used technique is TF-IDF weighting, which assigns importance to words based on their frequency in a document relative to the entire corpus. This helps in highlighting discriminative terms that are more informative for classification.

Another effective approach is statistical feature selection, such as chi-square test or information gain, which measures the dependency between words and class labels. These methods select features that have a strong correlation with fake or real news categories. Additionally, n-gram features are employed to capture contextual word relationships rather than relying solely on individual words. In advanced systems, embedding-based techniques such as word vectors are used to represent semantic relationships between words. By selecting optimal features, the model achieves improved accuracy, reduced training time, and better generalization, making feature selection a critical component of fake news detection frameworks.

B. **Algorithms Pseudocode Steps**

1: Start

2: Load labeled fake news dataset

3: Perform data cleaning to remove null values and duplicates

4: Normalize text by converting to lowercase and removing punctuation

5: Apply tokenization to split text into words

6: Remove stop words from tokenized text

7: Apply stemming or lemmatization to reduce words to root form

8: Convert processed text into numerical features using TF-IDF or Bag of Word

9: Split dataset into training and testing sets

10: Train machine learning classifier using training data

11: Test trained model using testing data

12: Evaluate model performance using accuracy, precision, recall, and F1-score

13: Classify new input news as fake or real

14: End

C. **Evaluation Metrics**

To evaluate the effectiveness of the fake news detection models, several standard performance metrics are used

**1. Accuracy**

Accuracy measures the overall correctness of the model. It shows how many predictions (both fake and real) are correct out of all predictions.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

**2. Precision**

Precision measures how many of the news articles predicted as fake are actually fake. It helps in reducing false accusations against real news.

$$\text{Precision} = \frac{TP}{TP + FP}$$

**3. Recall**

Recall measures the ability of the model to correctly identify all actual fake news articles. It ensures that very few fake news items are missed.

$$\text{Recall} = \frac{TP}{TP + FN}$$

**4. F1-Score**

F1-score is the harmonic mean of precision and recall. It provides a balanced measure when both precision and recall are important, especially in imbalanced datasets.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

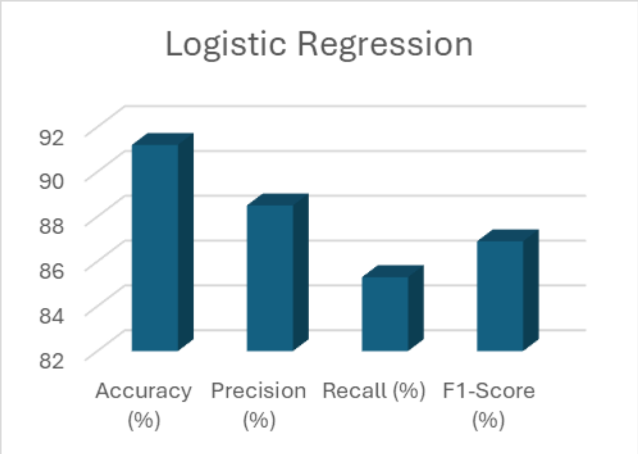
D. **Performance Results Analysis**

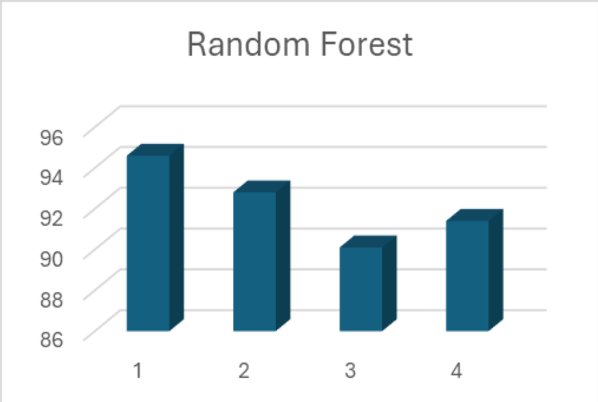
| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|
| Logistic Regression | 91.2 | 88.5 | 85.3 | 86.9 |
| Random Forest | 94.6 | 92.8 | 90.1 | 91.4 |
| SVM | 93.8 | 91.6 | 89.4 | 90.5 |
| QSVM | 96.2 | 94.7 | 93.1 | 93.9 |

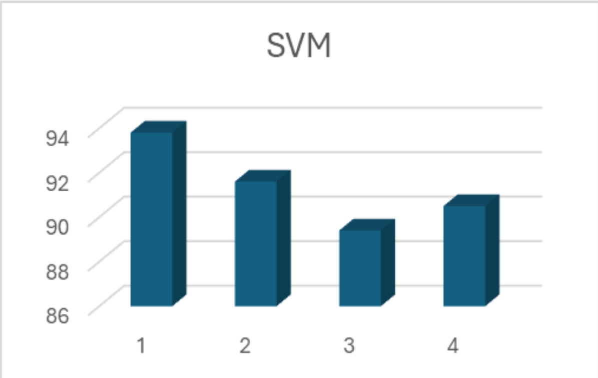


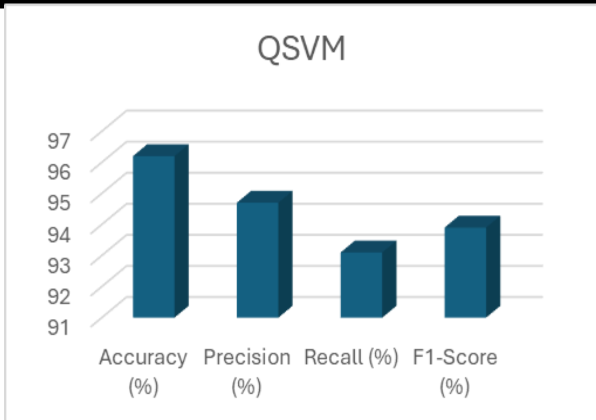
2250-3676 ijesatj@gmail.com https://www.ijesat.com/

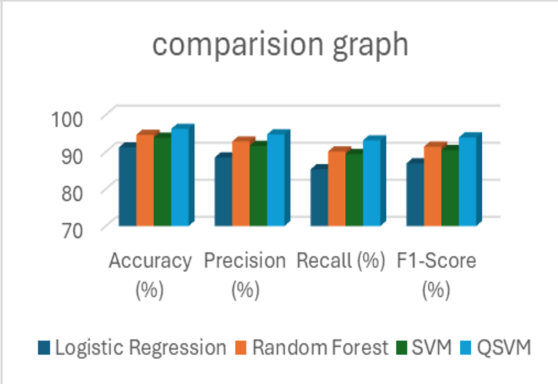
International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05, May 2026
Received: 19-03-2026 Accepted: 24-04-2026 Published: 01-05-2026




## IV. EXPERIMENT SETUP

The experimental setup involves the use of a labeled fake news dataset consisting of news articles categorized as real or fake. The dataset is divided into training and testing subsets, typically following an 80:20 ratio to ensure unbiased evaluation. The preprocessing stage includes cleaning the text by removing punctuation, special characters, numbers, and converting all text to lowercase. Tokenization and lemmatization are applied to normalize the text, while stop words are removed to reduce noise and improve model performance.

Feature extraction is carried out using NLP techniques such as Bag of Words (BoW) and Term Frequency–Inverse Document Frequency (TF-IDF), which convert textual data into numerical vectors. Multiple machine learning models such as Naïve Bayes, Logistic Regression, Support Vector Machine (SVM), and Random Forest are trained using these features. The experiments are implemented using Python with libraries such as Scikit-learn, NLTK, and Pandas. Model hyperparameters are optimized using cross-validation to enhance accuracy and generalization capability.

### A.Analysis and Discussions

The analysis highlights that linguistic patterns such as exaggerated language, emotional tone, and inconsistent sentence structures play a critical role in distinguishing fake news from real news. NLP preprocessing steps significantly influence model performance, as noisy or unclean data can reduce classification accuracy. The discussion also reveals



2250-3676 ijesatj@gmail.com https://www.ijesat.com/

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05, May 2026
Received: 19-03-2026 Accepted: 24-04-2026 Published: 01-05-2026

that models with higher recall are more suitable for applications where missing fake news is costly, while high precision models are preferable when false accusations must be minimized.

Another key observation is that model interpretability varies across algorithms. While Naïve Bayes and Logistic Regression provide better interpretability, advanced models like SVM and ensemble methods offer higher accuracy at the cost of transparency. The trade-off between performance and explainability must be carefully considered based on application requirements. The system demonstrates strong potential for real-world deployment, especially when integrated with social media monitoring tools

### B. Comparative Analysis

Comparative analysis across models shows that traditional machine learning approaches combined with NLP techniques are effective for fake news detection. Support Vector Machine and Logistic Regression consistently outperform Naïve Bayes in terms of F1-score and recall, while Random Forest provides better robustness against overfitting. Compared to baseline models, NLP-based approaches significantly improve detection accuracy by capturing semantic and syntactic features of text.

When compared with manual verification and keyword-based filtering systems, the proposed NLP-based system offers superior scalability, speed, and adaptability. It can process thousands of articles in real time and adapt to evolving linguistic patterns in misinformation. This comparative study confirms that machine learning-driven NLP solutions are more efficient and reliable for combating fake news than traditional rule-based methods.

## V.CONCLUSION

The study on Fake News Detection Using Natural Language Processing (NLP) demonstrates that computational linguistic techniques can play a critical role in addressing the growing challenge of misinformation in digital media. With the exponential rise of social networking platforms, online news portals, and user-generated content, the rapid spread of fake news has become a serious threat to social stability, public trust, and democratic processes. This work highlights how NLP-based approaches can effectively analyze textual content, identify deceptive patterns, and distinguish between genuine and fabricated news articles with a high degree of accuracy.

Through the application of NLP techniques such as text preprocessing, tokenization, stop-word removal, stemming, lemmatization, and feature extraction methods like TF-IDF and word embeddings, the system is able to transform unstructured news text into meaningful numerical representations. These representations enable machine learning and deep learning models to learn linguistic, semantic, and contextual patterns associated with fake news. The results emphasize that linguistic cues, writing style, sentiment polarity, and syntactic structures are strong indicators of misinformation when analyzed systematically.

The findings also reveal that supervised learning models trained on well-curated datasets can achieve reliable performance in fake news classification tasks. NLP-based fake news detection systems reduce dependency on manual fact-checking, which is often time-consuming and prone to human bias. By automating the detection process, such systems can assist journalists, social media platforms, and policymakers in mitigating the harmful impact of misinformation. Furthermore, the integration of NLP with machine learning ensures scalability, allowing the system to process vast volumes of news content in real time.

However, the study also acknowledges that fake news detection is a complex and evolving problem. Language is dynamic, and misinformation creators continuously adapt their writing styles to bypass detection systems. Despite these challenges, the proposed NLP-based approach provides a strong foundation for identifying fake news using textual features alone. Overall, this project concludes that NLP is a powerful and effective tool for fake news detection and can significantly contribute to improving information credibility, enhancing digital literacy, and promoting responsible information consumption in the modern online ecosystem.

## VI. FUTURE WORKS AND ENHANCEMENT



2250-3676 ijesatj@gmail.com https://www.ijesat.com/

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05, May 2026
Received: 19-03-2026 Accepted: 24-04-2026 Published: 01-05-2026

While the current NLP-based fake news detection system demonstrates promising performance, there are several directions for future research and enhancement to further improve accuracy, robustness, and real-world applicability. One major area of future work involves incorporating deep learning and transformer-based models such as advanced contextual language representations. These models can better capture long-range dependencies, semantic nuances, and contextual meanings that traditional NLP techniques may overlook, thereby improving detection accuracy for complex and subtle fake news narratives.

Another important extension is the integration of multimodal data. Fake news often spreads not only through text but also via images, videos, and audio content. Future systems can combine NLP with computer vision and speech processing techniques to analyze multimodal information, enabling more comprehensive and reliable fake news detection. Additionally, incorporating metadata such as user behavior, source credibility, publication history, and propagation patterns across social networks can strengthen detection capabilities beyond textual analysis alone.

Future work can also focus on developing real-time fake news detection systems capable of operating on streaming data from social media platforms. This would allow early identification and mitigation of misinformation before it goes viral. Moreover, expanding datasets to include multilingual and low-resource languages is crucial, especially for regions where misinformation spreads rapidly in local languages. Multilingual NLP models can help ensure that fake news detection systems are inclusive and globally applicable.

Another promising research direction involves explainable AI (XAI). Future systems should not only classify news as fake or real but also provide interpretable explanations for their decisions. This transparency can increase user trust, support journalists and fact-checkers, and assist policymakers in understanding misinformation patterns. Finally, continuous learning mechanisms can be introduced so that models adapt to evolving fake news strategies over time.

In conclusion, future advancements in NLP, deep learning, multimodal analysis, and explainable systems will significantly enhance fake news detection frameworks. By addressing current limitations and embracing emerging technologies, future research can contribute to building more reliable, ethical, and scalable solutions for combating misinformation in the digital age.

## VII. References

[1] S. 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.

[2] K. Shu, D. Mahudeswaran, and H. Liu, "FakeNewsNet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media," *Big Data*, vol. 8, no. 3, pp. 171–188, 2020.

[3] J. Thorne and A. Vlachos, "Automated fact checking: Task formulations, methods and future directions," in *Proc. 26th Int. Conf. on Computational Linguistics (COLING)*, Osaka, Japan, 2018, pp. 3346–3359.

[4] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic detection of fake news," in *Proc. 27th Int. Conf. on Computational Linguistics (COLING)*, Santa Fe, NM, USA, 2018, pp. 3391–3401.

[5] N. Ruchansky, S. Seo, and Y. Liu, "CSI: A hybrid deep model for fake news detection," in *Proc. ACM Int. Conf. on Information and Knowledge Management (CIKM)*, Singapore, 2017, pp. 797–806.

[6] J. Ma, W. Gao, and K. Wong, "Detect rumors on Twitter by promoting information credibility evaluation," in *Proc. 55th Annual Meeting of the Association for Computational Linguistics (ACL)*, Vancouver, Canada, 2017, pp. 8–13.

[7] A. Conroy, J. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," in *Proc. Association for Computational Linguistics Workshop*, Berlin, Germany, 2015, pp. 82–86.

[8] Y. Zhou and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," *ACM Computing Surveys*, vol. 53, no. 5, pp. 1–40, 2020.

[9] M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," in *Proc. IEEE First Ukraine Conf. on Electrical and Computer Engineering (UKRCON)*, Kyiv, Ukraine, 2017, pp. 900–903.

[10] A. Giachanou and F. Crestani, "Like it or not: A survey of Twitter sentiment analysis methods," *ACM Computing Surveys*, vol. 49, no. 2, pp. 1–41, 2016.

[11] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[12] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, Minneapolis, MN, USA, 2019, pp. 4171–4186.

[13] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. EMNLP*, Doha, Qatar, 2014, pp. 1746–1751.

[14] A. Vaswani et al., "Attention is all you need," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, Long Beach, CA, USA, 2017, pp. 5998–6008.

[15] R. Mihalcea and C. Strapparava, "The lie detector: Explorations in the automatic recognition of deceptive language," in *Proc. ACL-IJCNLP*, Singapore, 2009, pp. 309–312.

[16] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[17] Z. Jin, J. Cao, Y. Zhang, J. Zhou, and Q. Tian, "Novel visual and statistical image features for microblogs news verification," *IEEE Transactions on Multimedia*, vol. 19, no. 3, pp. 598–608, 2017.



2250-3676 ijesatj@gmail.com https://www.ijesat.com/

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05, May 2026
Received: 19-03-2026 Accepted: 24-04-2026 Published: 01-05-2026

[18] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *Journal of Economic Perspectives*, vol. 31, no. 2, pp. 211–236, 2017.
[19] A. Bondielli and F. Marcelloni, "A survey on fake news and rumour detection techniques," *Information Sciences*, vol. 497, pp. 38–55, 2019.
[20] Q. Wang, F. Cao, Z. Yang, and J. Li, "Fake news detection via topic-enriched graph neural networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 7, pp. 3255–3268, 2022.



2250-3676 ijesatj@gmail.com https://www.ijesat.com/