

An Intelligent Real-Time Fake News Detection Framework Using Machine Learning and Textual Feature Engineering

Hamid Ghous

Department of Computer Science & Information Technology, University of Southern Punjab (USP) Multan, Punjab, Pakistan Email: hamidghous@usp.edu.pk

Syed Zulfiqar Haider Bukhari

Directorate of Information Technology, Islamia University Bahawalpur (IUB), Punjab, Pakistan Email: syedzulfiqar@iub.edu.pk

Muhammad Talah Zubair

Department of Computer NCBA&E Lahore Sub Campus Multan, Punjab, Pakistan Email: talhazubair009@gmail.com

Muhammad Akmal Shahzad

Department of Computer Science & Information Technology, University of Southern Punjab (USP) Multan, Punjab, Pakistan Email: akmalshahzadgurmani@gmail.com

Ghulam Muhy Ud Deen Raee*

Department of Computer Science & Information Technology, University of Southern Punjab (USP) Multan, Punjab, Pakistan Email: gmraees@gmail.com

Wajiha Qavi

Department of Computer Science & Information Technology, University of Southern Punjab (USP) Multan, Punjab, Pakistan Email: wajihaqavi00@gmail.com

ABSTRACT

The rapid dissemination of misinformation via digital platforms has created an urgent need for automated detection systems that can operate at scale. This paper presents an intelligent real-time fake news detection framework that integrates classical machine learning with optimized textual feature engineering. By focusing on efficient architectures, the proposed system addresses the limitations of manual fact-checking, which is increasingly inadequate in the face of modern misinformation campaigns.

The methodology involves a comprehensive text preprocessing pipeline—including conversion to lowercase, removal of URLs, numbers, and punctuation, as well as tokenization, stopword removal, and lemmatization. For feature representation, the system utilizes Term Frequency-Inverse Document Frequency with a bigram configuration, capturing up to 50,000 unique features to maintain contextual depth. The classification phase evaluates multiple models, specifically Logistic Regression, Support Vector Machines, Naive Bayes, Random Forest, and the SGD Classifier.

A key innovation of this framework is the application of probability calibration to

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

mitigate the issue of overconfident misclassifications commonly found in standard NLP models. By applying techniques such as Platt Scaling, the system transforms raw scores into reliable posterior probabilities. Experimental evaluations show that the calibrated Linear SVM is the most effective model, achieving a peak accuracy of 97.95% and an ROC-AUC of 99.75%, surpassing traditional uncalibrated benchmarks. The final system is deployed using a Streamlit-based web interface, providing real-time predictions with confidence scores and interactive visualizations powered by Plotly and Seaborn. The framework demonstrates that optimized traditional machine learning models can outperform more complex architectures in efficiency and reliability. These results provide a scalable solution for news platforms to verify content instantly, bridging the gap between theoretical research and practical application.

Keywords: Fake News Detection, Machine Learning, Probability Calibration, Real-Time Verification, Deep Text Analytics

Introduction

The exponential growth of online news platforms and social media has fundamentally altered how information is consumed, but it has also facilitated the viral spread of fake news [11]. Misinformation poses severe threats to public opinion, democratic processes, and global societal stability, making the identification of deceptive content a priority for modern information science [2], [3]. As malicious actors leverage digital networks to amplify fabricated narratives, the credibility of legitimate journalism is increasingly under pressure [3].

Manual verification of news content is no longer a viable solution due to the sheer volume and velocity of modern information flows [3], [12]. Human fact-checkers are often overwhelmed, creating a delay between the publication of a fake article and its eventual debunking [2]. Automated fake news detection systems, powered by Natural Language Processing and Machine Learning, are therefore essential for providing instantaneous, scalable verification that can match the speed of digital media [10], [12]. While recent research has shifted toward deep learning models, traditional machine learning algorithms remain highly effective when paired with rigorous feature engineering [1], [13]. Classical models such as SVM and Random Forest offer significant advantages in terms of computational efficiency and interpretability, allowing for deployment on lightweight, real-time architectures [1], [14]. This study emphasizes the optimization of these models through TF-IDF bigrams and comprehensive preprocessing to ensure high accuracy without the high overhead of transformer-based systems [5], [15].

A primary focus of this research is solving the problem of uncalibrated model outputs, which often lead to overconfident predictions in text classification. By integrating probability calibration, the framework ensures that the confidence scores provided to the user reflect the actual empirical likelihood of the news being fake. This contribution, combined with a user-friendly Streamlit deployment, provides a complete pipeline for reliable misinformation detection in real-world scenarios.

Research Gap

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

Despite the abundance of research in this field, several critical gaps remain: Most existing systems provide binary "Real" or "Fake" labels without reliable probability estimates, often leading to overconfident misclassifications [7], [16]. Many frameworks are tested on static datasets but fail to provide the lightweight architecture necessary for real-time web deployment [12]. There is a tendency to favor deep learning even when traditional ML models, if properly optimized with feature engineering, can achieve comparable or superior results with less overhead [1]. Many existing tools do not offer visual feedback or explainable confidence scores to the end-user.

Research Objectives

To address these gaps, this study pursues the following objectives:
To develop an end-to-end real-time fake news detection system
To implement a multi-stage preprocessing pipeline to handle noisy digital text.
To utilize TF-IDF with bigram representation to capture the stylistic and contextual signatures of deceptive writing [5]
To evaluate and compare the performance of multiple ML classifiers under identical conditions
To apply probability calibration to improve the accuracy of confidence scores [17], [18]
To build and deploy a functional web interface using Streamlit for real-world interaction [3], [9]

Literature Review

The detection of fake news has evolved into a critical multi-disciplinary challenge, necessitated by the viral spread of verifiably false information that undermines democratic stability and public trust [2], [11], [12], [19]. Surveys by Zhou and Zafarani [11] and Alghamdi et al. [19] highlight how social networks facilitate the viral transmission of misinformation, often outstripping the capacity of traditional editorial oversight [20], [21], [22]. Researchers have consistently noted that fake news thrives by exploiting psychological vulnerabilities and the "echo chamber" effect [20], [23], [24]. Consequently, automated systems have transitioned from simple heuristic tools to advanced frameworks that integrate linguistic analysis with social context to safeguard information integrity [2], [22], [24].

Effective detection frameworks rely on rigorous text preprocessing to minimize noise and standardize diverse textual data [25], [26]. Standard pipelines involve multi-stage cleaning, including the removal of URLs, punctuation, and numerical data to focus on core linguistic features [13], [27], [28]. Studies by de Oliveira et al. [21] and Hisham et al. [28] emphasize that lemmatization and stopword removal are essential for consolidating high-dimensional feature spaces [26], [29]. Beyond simple word counts, feature representation methods like TF-IDF combined with N-grams specifically bigrams and trigrams are crucial for capturing the unique rhetorical structures and contextual signatures found in deceptive writing [4], [5], [24], [30], [31].

Despite the emergence of complex architectures, traditional machine learning models remain highly effective and efficient for news classification [1], [13], [32]. Support

Vector Machines consistently demonstrate superior performance in high-dimensional text spaces, often reaching accuracies above 97% [5], [13], [33]. Ensemble methods like Random Forest are also favored for their ability to handle non-linear relationships, with some studies reporting accuracies as high as 99.95% on balanced datasets [4], [34], [35]. Meanwhile, models like Naive Bayes and Logistic Regression provide strong baselines and computational efficiency, which are vital for large-scale, real-time applications where low-latency inference is required [14], [15], [36], [37].

Advanced semantic analysis has been further enhanced by deep learning and emotional detection techniques [9], [38], [39]. Transformers like BERT offer deep bidirectional context understanding, while sequential models like LSTMs capture patterns in long-form narratives [9], [39], [40]. Research has also highlighted the importance of multilingual support, using cross-lingual transfer to address misinformation in low-resource languages [41], [42], [43], [44]. Furthermore, the emotional charge of an article serves as a powerful indicator of authenticity, as fake news is often designed to trigger high-arousal emotions [45], [46], [47]. Integrating sentiment and "dual emotion" analysis evaluating both the publisher's tone and the social response has been shown to significantly improve detection precision [46], [48], [49].

A critical yet often overlooked aspect of fake news detection is the reliability and security of model outputs [7], [8], [18]. Standard classifiers often produce uncalibrated probabilities, leading to overconfident misclassifications [6], [18], [50]. Techniques such as Platt Scaling and Bayesian Binning are essential for ensuring that confidence scores reflect true empirical likelihoods [6], [17], [51]. Additionally, models must be robust against adversarial attacks where text is subtly altered to bypass filters [52], [53], [54]. Studies show that adversarial training can reduce the success of such attacks significantly [55], [56]. To ensure transparency, Explainable AI tools like SHAP and LIME are being integrated to provide visual justifications for model decisions, although care must be taken to prevent these tools from being misused by adversaries [57], [58], [59], [60], [61].

The transition from academic theory to practical application requires lightweight, scalable architectures suitable for real-time verification [3], [9], [62]. Modern systems utilize frameworks like Streamlit and WebSocket to provide immediate feedback to users on platforms like WhatsApp or Twitter [3], [10], [62]. Low-latency inference, often as fast as 72ms, is a primary goal for these "human-in-the-loop" systems [2], [10]. As misinformation tactics continue to evolve, the integration of automated fact-checking pipelines in multiple languages remains a central focus for future development [12], [63], [64], [65]. This comprehensive body of literature underscores the necessity of a balanced approach that combines high accuracy with reliable confidence scoring and real-time accessibility.

Methodology

The proposed framework follows a comprehensive methodology for fake news detection, beginning with an extensive text preprocessing pipeline to enhance data quality. This includes lowercasing, noise removal, tokenization, stop word elimination, and lemmatization to standardize and refine textual input. Subsequently, feature extraction is performed using TF-IDF vectorization with bigram representation,

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

enabling the model to capture meaningful word patterns and contextual relationships. Multiple machine learning models, including Support Vector Machine, Random Forest, Logistic Regression, Naive Bayes variants, and SGD Classifier, are trained and evaluated to ensure robust performance. To improve prediction reliability, probability calibration using Plot Scaling is applied across all models. Finally, the system is deployed as a web-based application that supports real-time news classification, confidence scoring, visualization of results, and efficient model storage for fast and scalable usage.

Text Preprocessing Pipeline

To ensure high data quality, the framework performs several cleaning steps:

Lowercasing: Standardizing text to prevent the model from treating the same word differently based on case [27].

Noise Removal: Stripping URLs, HTML tags, punctuation, and numerical data that do not contribute to semantic meaning [13], [27].

Tokenization & Stopword Removal: Breaking text into individual words and removing common functional words (e.g., "is", "the") using the NLTK library [27].

Lemmatization: Reducing words to their dictionary base form (e.g., "better" becomes "good") to consolidate the feature space [27].

Feature Extraction via TF-IDF Bigrams

The system uses TF-IDF Vectorization to convert text into numerical vectors. By setting the maximum number of features to 50,000 and including **bigrams**, the model can identify significant two-word phrases that signal deceptive intent [5], [66].

Model Training and Calibration

Six distinct models were trained and evaluated:

Support Vector Machine: Utilized for its superior margin-based classification.

Random Forest: An ensemble of decision trees used to capture non-linear relationships [15].

Logistic Regression: Used as a highly interpretable probabilistic baseline.

Naive Bayes & Complement Naive Bayes: Probabilistic models optimized for text [1].

SGD Classifier: A linear model optimized for large-scale datasets.

Every model underwent **probability calibration** using Platt Scaling to ensure that the confidence scores provided to the user are statistically reliable [6], [18].

System Functionalities and Deployment

The system is deployed as a web application with the following functionalities:

Real-Time Prediction: Users can paste news articles and receive an instant classification [10].

Confidence Scoring: The interface displays a percentage score representing the model's certainty.

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

Visualization: Integration with **Plotly** and **Seaborn** to visualize classification metrics and feature importance.

Lightweight Persistence: Using **Joblib** for efficient model storage and fast loading times [3].

Results and Evaluation

The models were tested on comprehensive datasets such as ISOT and LIAR to ensure generalizability [15], [67].

Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC
Linear SVM	97.95%	97.95%	97.95%	97.95%	99.75%
Random Forest	97.69%	97.69%	97.69%	97.69%	99.77%
Logistic Regression	97.41%	97.41%	97.41%	97.41%	99.69%
SGD Classifier	96.87%	96.87%	96.87%	96.87%	99.54%
Naive Bayes	92.78%	92.79%	92.78%	92.78%	97.76%

The Linear SVM emerged as the top performer, achieving an exceptional accuracy of 97.95%. The ROC-AUC of 99.75% indicates nearly perfect separation between the real and fake news classes.

Discussion

The high performance of the Linear SVM highlights the effectiveness of margin-based classification in sparse, high-dimensional text data [1], [5]. The use of TF-IDF with bigrams was critical; bigrams captured sequences like "source says" or "official reports" which vary significantly between factual and fabricated content [13].

Furthermore, the integration of probability calibration addressed the primary weakness of many NLP classifiers overconfidence. By providing a well-calibrated confidence score, the system allows users to interpret the results with appropriate skepticism [16], [18]. The Streamlit deployment ensures that these complex backend processes are presented in a simple, user-friendly format, facilitating real-time interaction without requiring technical expertise [3], [9].

Conclusion

This research successfully developed an intelligent framework for real-time fake news detection. By optimizing traditional machine learning models and implementing probability calibration, we achieved a highly accurate (97.95%) and reliable system. The framework demonstrates that classical ML, when paired with thoughtful feature engineering and user-centric deployment, remains a powerful and efficient solution for combating digital misinformation.

Future work will focus on integrating Explainable AI to highlight specific words or phrases that led to a classification and expanding the system to handle multilingual datasets to address global misinformation challenges.

References

- H. K. and Dr. V. K. Verma, "Fake News Detection Using Classical Machine Learning Models: A Comparative Analysis," *International Journal of Emerging Technologies and Innovative Research*, pp. 306–306, Dec. 2025, doi: 10.48175/ijetir-9244.
- U. Twaha and Y. Arfin, "An AI-Driven Framework for Real-Time Fake News Detection: Developing a Machine Learning-Based Filter for News Platforms in the United States," *International Journal of Future Engineering Innovations*, vol. 2, no. 4, pp. 158–169, Jan. 2025, doi: 10.54660/ijfei.2025.2.4.158-169.
- V. Gunasekaran and D. Karthikeyan, "Real Time Fake news Detection Web App Enhanced by Machine Learning Algorithms," vol. 8, pp. 1–7, Jun. 2024, doi: 10.1109/conit61985.2024.10627483.
- M. A. Pradhan, "Fake News Detection Methods: Machine Learning Approach," *International Journal for Research in Applied Science and Engineering Technology*, vol. 8, no. 7, pp. 971–975, Jul. 2020, doi: 10.22214/ijraset.2020.29630.
- R. Rizal, A. Faturahman, A. Impron, I. Darmawan, E. Haerani, and A. Rahmatulloh, "Unveiling the Truth: Detecting Fake News Using SVM and TF-IDF," pp. 1–6, Feb. 2025, doi: 10.1109/icadeis65852.2025.10933324.
- M. P. Naeini, G. Cooper, and M. Hauskrecht, "Obtaining Well Calibrated Probabilities Using Bayesian Binning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Association for the Advancement of Artificial Intelligence, Feb. 2015. doi: 10.1609/aaai.v29i1.9602.
- H. Guo, R. Pasunuru, and M. Bansal, "An Overview of Uncertainty Calibration for Text Classification and the Role of Distillation," pp. 289–306, Jan. 2021, doi: 10.18653/v1/2021.repl4nlp-1.29.
- R. Singh and S. Goshtasbpour, "Platt-Bin: Efficient Posterior Calibrated Training for NLP Classifiers," in *Findings of the Association for Computational Linguistics: ACL 2022*, Jan. 2022, pp. 3673–3684. doi: 10.18653/v1/2022.findings-acl.290.
- A. Chabukswar, B. Vishwa, D. S. P, and K. R. Venugopal, "An Integrated Framework for Misinformation Detection Using Streamlit Application," *Journal of Machine and Computing*, pp. 266–279, Oct. 2025, doi: 10.53759/7669/jmc202606020.
- B. Jeyarajan and V. J. Pugazhenth, "SmartFactCheckBot: A Dual-Interface AI Platform for Real-Time Misinformation Detection with Explainable Predictions," pp. 1–5, Dec. 2025, doi: 10.1109/icca66035.2025.11431031.
- X. Zhou and R. Zafarani, "A Survey of Fake News," *ACM Computing Surveys*, vol. 53, no. 5. Association for Computing Machinery, pp. 1–40, Jul. 07, 2020. doi: 10.1145/3395046.
- S. Jain, A. Gupta, A. P. Tripathi, and A. Mishra, "Fake News Detection Web

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

- Application: A Unified Deep Learning Framework for Real-Time Fake News Detection,” *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT* , vol. 10, no. 2, pp. 1–9, Feb. 2026, doi: 10.55041/ijssrem56718.
- K. Poddar, G. B. A. D, and K. S. Umadevi, “Comparison of Various Machine Learning Models for Accurate Detection of Fake News,” *2019 Innovations in Power and Advanced Computing Technologies (i-PACT)* , pp. 1–5, Mar. 2019, doi: 10.1109/i-pact44901.2019.8960044.
- Z. Sadok, “An Adaptive and Energy-Efficient Edge Framework for Real-Time Fake News Detection,” pp. 1440–1447, May 2025, doi: 10.1109/iwcmc65282.2025.11059582.
- R. Mogre, “Fake News Detection Using Machine Learning Models: A Comparative Study of ISOT and LIAR Datasets,” *American journal of student research.* , pp. 519–524, Jan. 2026, doi: 10.70251/hyjr2348.41519524.
- “Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers),” Jan. 2022, doi: 10.18653/v1/2022.acl-short.
- A. Mudinas, D. Zhang, and M. Levene, “Bootstrap Domain-Specific Sentiment Classifiers from Unlabeled Corpora,” *Transactions of the Association for Computational Linguistics* , vol. 6, pp. 269–285, Dec. 2018, doi: 10.1162/tacl_a_00020.
- J.-C. Klie, B. Webber, and I. Gurevych, “Annotation Error Detection: Analyzing the Past and Present for a More Coherent Future,” *Computational Linguistics* , vol. 49, no. 1, pp. 157–198, Oct. 2022, doi: 10.1162/coli_a_00464.
- J. Alghamdi, S. Luo, and Y. Lin, “A comprehensive survey on machine learning approaches for fake news detection,” *Multimedia Tools and Applications* , vol. 83, no. 17, pp. 51009–51067, Nov. 2023, doi: 10.1007/s11042-023-17470-8.
- Ștefan E. REPEDE, “Researching disinformation using artificial intelligence techniques: challenges,” *BULLETIN OF CAROL I NATIONAL DEFENCE UNIVERSITY* , vol. 12, no. 2, pp. 69–85, Jul. 2023, doi: 10.53477/2284-9378-23-21.
- N. R. de Oliveira, P. S. Pisa, M. A. Lopez, D. S. V. Medeiros, and D. M. F. Mattos, “Identifying Fake News on Social Networks Based on Natural Language Processing: Trends and Challenges,” *Information* , vol. 12, no. 1, pp. 38–38, Jan. 2021, doi: 10.3390/info12010038.
- F. C. D. da Silva, R. J. Vieira, and A. C. B. García, “Can Machines Learn to Detect Fake News? A Survey Focused on Social Media,” in *Proceedings of the ... Annual Hawaii International Conference on System Sciences/Proceedings of the Annual Hawaii International Conference on System Sciences* , Jan. 2019. doi: 10.24251/hicss.2019.332.
- A. Bondielli and F. Marcelloni, “A survey on fake news and rumour detection techniques,” *Information Sciences* , vol. 497, pp. 38–55, May 2019, doi: 10.1016/j.ins.2019.05.035.
- M. Choraś *et al.* , “Advanced Machine Learning techniques for fake news (online disinformation) detection: A systematic mapping study,” *Applied Soft*

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

- Computing* , vol. 101,
pp. 107050–107050, Dec. 2020, doi: 10.1016/j.asoc.2020.107050.
- R. Oshikawa, J. Qian, and W. Y. Wang, “A Survey on Natural Language Processing for Fake News Detection,” *arXiv (Cornell University)* , Mar. 2022, doi: 10.48550/arxiv.1811.00770.
- P. Dhiman and A. Kaur, “Text pre-processing techniques in addressing fake news,” in *Computational Methods in Science and Technology* , 2024, pp. 142–146. doi: 10.1201/9781003501244-24.
- Ö. Özgöbek, B. Kille, A. R. From, and I. U. Netland, “Fake News Detection by Weakly Supervised Learning Based on Content Features,” in *Communications in computer and information science* , Springer Science+Business Media, 2022, pp. 52–64. doi: 10.1007/978-3-031-17030-0_5.
- M. S. F. M. Hisham, R. Hasan, and S. Hussain, “An Innovative Approach for Fake News Detection using Machine Learning,” *Sir Syed University Research Journal of Engineering & Technology* , vol. 13, no. 1, pp. 115–124, Jun. 2023, doi: 10.33317/ssurj.565.
- A. Mahabub, “A robust technique of fake news detection using Ensemble Voting Classifier and comparison with other classifiers,” *SN Applied Sciences* , vol. 2, no. 4, Mar. 2020, doi: 10.1007/s42452-020-2326-y.
- P. Bharadwaj and Z. Shao, “Fake News Detection with Semantic Features and Text Mining,” *International Journal on Natural Language Computing* , vol. 8, no. 3, pp. 17–22, Jun. 2019, doi: 10.5121/ijnlc.2019.8302.
- S. N. Alsubari, “Data Analytics for the Identification of Fake Reviews Using Supervised Learning,” *Computers, materials & continua/Computers, materials & continua (Print)* , vol. 70, no. 2, pp. 3189–3204, Sep. 2021, doi: 10.32604/cmc.2022.019625.
- H. Khare and V. K. Verma, “Fake News Detection Using Classical Machine Learning Models: A Comparative Review,” *International Journal of Advanced Research in Science Communication and Technology* , pp. 141–145, Nov. 2025, doi: 10.48175/ijarsct-29621.
- J. S. Wibowo, E. N. Wahyudi, and H. Listiyono, “Performance Comparison of SVM, Naive Bayes, and Random Forest Models in Fake News Classification,” vol. 9, no. 8, Aug. 2024, doi: 10.47191/etj/v9i08.27.
- I. Jahan *et al.* , “Advanced machine learning techniques for fake news detection: A comprehensive analysis,” *Magna Scientia Advanced Research and Reviews* , vol. 12, no. 2,
pp. 203–212, Dec. 2024, doi: 10.30574/msarr.2024.12.2.0198.
- C. S. Krishna and G. Divya, “Accuracy measure for identification of fake blog articles in social website using random forest over support vector machine,” *AIP conference proceedings*
, vol. 2729, pp. 60008–60008, Jan. 2024, doi: 10.1063/5.0189438.
- A. Albahr and M. A. Albahr, “An Empirical Comparison of Fake News Detection using different Machine Learning Algorithms,” *International Journal of Advanced Computer Science and Applications* , vol. 11, no. 9, Jan. 2020, doi:

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

- 10.14569/ijacsa.2020.0110917.
- S. R. Pappu, N. Sangeetha, K. Maniteja, Y. R. Dilip, and A. G. S. Abhishek, "Analysis of Classifiers for Fake News Detection: A Machine Learning Perspective," *International Journal for Research in Applied Science and Engineering Technology*, vol. 10, no. 7, pp. 1573–1580, Jul. 2022, doi: 10.22214/ijraset.2022.44925.
- M. K. M and G. David, "Real-Time Fake News Detection Using Machine Learning," *International Journal of Advanced Research in Science Communication and Technology*, pp. 274–281, Oct. 2025, doi: 10.48175/ijarsct-29336.
- 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, pp. 100007–100007, Jan. 2021, doi: 10.1016/j.jjime.2020.100007.
- M. H. Goldani, S. Momtazi, and R. Safabakhsh, "Detecting fake news with capsule neural networks," *Applied Soft Computing*, vol. 101, pp. 106991–106991, Dec. 2020, doi: 10.1016/j.asoc.2020.106991.
- L. Lan, T. Huang, Y. Li, and Y. Song, "A Survey of Cross-Lingual Text Classification and Its Applications on Fake News Detection," *World Scientific Annual Review of Artificial Intelligence*, vol. 1, Jan. 2023, doi: 10.1142/s2811032323500030.
- A. De, D. Bandyopadhyay, B. Gain, and A. Ekbal, "A Transformer-Based Approach to Multilingual Fake News Detection in Low-Resource Languages," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 21, no. 1, pp. 1–20, Nov. 2021, doi: 10.1145/3472619.
- M. Ghayoomi and M. Mousavian, "Deep transfer learning for COVID-19 fake news detection in Persian," *Expert Systems*, vol. 39, no. 8, Apr. 2022, doi: 10.1111/exsy.13008.
- D. Telman *et al.*, "Cross-Lingual and Multilingual Approaches to Fake News Detection in the Kazakh Language," *Procedia Computer Science*, vol. 275, pp. 708–715, Jan. 2026, doi: 10.1016/j.procs.2026.01.082.
- Z. Liu, T. Zhang, K. Yang, P. M. Thompson, Z. Yu, and S. Ananiadou, "Emotion detection for misinformation: A review," *Information Fusion*, vol. 107, pp. 102300–102300, Feb. 2024, doi: 10.1016/j.inffus.2024.102300.
- M. Á. Alonso, D. Vilares, C. Gómez-Rodríguez, and J. Vilares, "Sentiment Analysis for Fake News Detection," *Electronics*, vol. 10, no. 11, pp. 1348–1348, Jun. 2021, doi: 10.3390/electronics10111348.
- A. Giachanou, P. Rosso, and F. Crestani, "Leveraging Emotional Signals for Credibility Detection," pp. 877–880, Jul. 2019, doi: 10.1145/3331184.3331285.
- X. Zhang, J. Cao, X. Li, Q. Sheng, L. Zhong, and K. Shu, "Mining Dual Emotion for Fake News Detection," pp. 3465–3476, Apr. 2021, doi: 10.1145/3442381.3450004.
- S. Kh. Hamed, M. J. A. Aziz, and M. R. Yaakub, "Fake News Detection Model on Social Media by Leveraging Sentiment Analysis of News Content and Emotion Analysis of Users' Comments," *Sensors*, vol. 23, no. 4, pp. 1748–1748, Feb. 2023, doi: 10.3390/s23041748.
- Z. Jiang, A. Liu, and B. V. Durme, "Addressing the Binning Problem in Calibration

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

- Assessment through Scalar Annotations,” *Transactions of the Association for Computational Linguistics* , vol. 12, pp. 120–136, Jan. 2024, doi: 10.1162/tacl_a_00636.
- W. Kweon, S. Kang, and H. Yu, “Obtaining Calibrated Probabilities with Personalized Ranking Models,” in *Proceedings of the AAAI Conference on Artificial Intelligence* , Association for the Advancement of Artificial Intelligence, Jun. 2022, pp. 4083–4091. doi: 10.1609/aaai.v36i4.20326.
- P. Kumar, S. Arora, and S. Arora, “Defending Against Adversarial Attacks in Text-based Fake News Detection: A Machine Learning Survey,” *Zenodo (CERN European Organization for Nuclear Research)* , May 2023, doi: 10.5281/zenodo.8129094.
- I. Alsmadi *et al.* , “Adversarial NLP for Social Network Applications: Attacks, Defenses, and Research Directions,” *IEEE Transactions on Computational Social Systems* , vol. 10, no. 6, pp. 3089–3108, Nov. 2022, doi: 10.1109/tcss.2022.3218743.
- A. Koren, C. Underwood, E. Serra, and F. Spezzano, “Evaluating the robustness of fake news detectors to adversarial attacks with real user comments,” *International Journal of Data Science and Analytics* , vol. 20, no. 6, pp. 5689–5702, May 2025, doi: 10.1007/s41060-025-00790-3.
- C. Underwood, M. A. Islam, E. Serra, and F. Spezzano, “Generating Realistic Adversarial User Comment Attacks to Evaluate the Robustness of Fake News Detectors,” *IEEE Transactions on Computational Social Systems* , pp. 1–18, Jan. 2026, doi: 10.1109/tcss.2026.3653161.
- H. Ahmed, I. Traoré, S. Saad, and M. Mamun, “Effect of Text Augmentation and Adversarial Training on Fake News Detection,” *NPARC* , vol. 11, no. 4, pp. 4775–4789, Jan. 2024, doi: 10.1109/tcss.2023.3344597.
- S. Bansal, H. P. S. Panwar, and G. Munjal, “Detecting Fake News with XAI: A Look at SHAP for Making Machine Learning Models More Explainable,” pp. 1–6, Jul. 2025, doi: 10.1109/i2itcon65200.2025.11210605.
- K. Siwek, D. A. Stankowski, and M. Stodolski, “Trust Oriented Explainable AI for Fake News Detection,” Mar. 12, 2026, *Cornell University* . doi: 10.48550/arxiv.2603.11778.
- G. George and T. Abdullahi, “A SHAP-Based XAI Approach to Evaluating Machine Learning Classification of Fake News,” *Nile Journal of Engineering and Applied Science* , vol. 2, no. 2, pp. 1–1, Jan. 2025, doi: 10.5455/njeas.214016.
- K. Siwek, D. A. Stankowski, and M. Stodolski, “Trust Oriented Explainable AI for Fake News Detection,” *ArXiv.org* , Mar. 2026, Accessed: Mar. 2026. [Online]. Available: <http://arxiv.org/abs/2603.11778>
- R. Kozik, M. Ficco, A. Pawlicka, M. Pawlicki, F. Palmieri, and M. Choraś, “When explainability turns into a threat - using xAI to fool a fake news detection method,” *Computers & Security* , vol. 137, pp. 103599–103599, Nov. 2023, doi: 10.1016/j.cose.2023.103599.
- B. Thiagarajan, B. R. A. V. S, E. E. A. D, M. G, and S. Pavithra, “Real-Time Fake News Detection using NLP and WebSocket Frameworks,” pp. 1568–1574, Dec. 2025, doi: 10.1109/icacrs67045.2025.11324029.

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

- M. Hosseini, A. J. Sabet, S. He, and D. Aguiar, “Interpretable fake news detection with topic and deep variational models,” *Online Social Networks and Media* , vol. 36, pp. 100249–100249, May 2023, doi: 10.1016/j.osnem.2023.100249.
- S. Munir and H. Munir, “NLP-based approach to multilingual fake news detection through social media in low-resource languages: A review,” Sep. 09, 2025. doi: 10.20944/preprints202509.0710.v1.
- J. Drchal, H. Ullrich, T. Mlynář, and V. Moravec, “Pipeline and dataset generation for automated fact-checking in almost any language,” *Neural Computing and Applications* , vol. 36, no. 30, pp. 19023–19054, Aug. 2024, doi: 10.1007/s00521-024-10113-5.
- S. Abro, S. Shaikh, Z. Hussain, Z. Ali, S. U. Khan, and G. Mujtaba, “Automatic Hate Speech Detection using Machine Learning: A Comparative Study,” *International Journal of Advanced Computer Science and Applications* , vol. 11, no. 8, Jan. 2020, doi: 10.14569/ijacsa.2020.0110861.
- C. Blackledge and A. Atapour-Abarghouei, “Transforming Fake News: Robust Generalisable News Classification Using Transformers,” in *2021 IEEE International Conference on Big Data (Big Data)* , Dec. 2021, pp. 3960–3968. doi: 10.1109/bigdata52589.2021.9671970.