Implementation and Evaluation of Enhanced CNN Algorithm for Fake News Detection

Shraddha Shah[1] and Sachin Patel[2]*

[1] SAGE University, Indore, Madhya Pradesh, India
[2] SAGE University, Indore, Madhya Pradesh, India
31.shraddhashah@gmail.com
*drsachinpatel.sage@gmail.com

Abstract: In this digital era transformation of digitalization creating unprecedented access to news and content globally. Advancement of technological some time facilitated and propagate of misinformation across online platforms as well as social media platforms. This kind of misleading content can disturb the economy level by purchasing in panic situation create damage in political environment due to create confusion among voters during election period also harm social atmosphere through spreading rumors in between friends, neighbors and relatives. This paper identifies distinctive patterns, textual characteristics, in addition to contextual indicators that distinguish among authentic and fabricated news content by implementing not only machine learning techniques but also various natural language processing (NLP) method. Preprocessing is a crucial step that transforms raw and unstructured data as an input into a usable format for machine learning models. Improved algorithm implements a comprehensive pre-processing pipeline with multiple stages. This document provides a comprehensive indication of all pre-processing and post-processing techniques implemented in the enhanced CNN-based fake news detection system. The improvements focus on enhancing text quality, feature representation, model architecture, and evaluation methodologies.

Keywords: Text Cleaning, Fake News Detection, Convolutional Neural Network, Deep Learning.

1 Introduction

The rise with World Wide Web and the quick uptake of social media platforms such as Facebook, Instagram, WhatsApp, YouTube, twitter, etc knowledge was able to be shared in a way that in human history had never been witnessed earlier [16]. The extensive usage of social media platforms benefited news organizations in addition to other applications since it allowed them to deliver the most recent news to customers in a nearly instantaneous manner. Now a days newspapers, tabloids, magazines, radio are replaced with digital medium like blogs, social media feeds, online news platforms, and other comparable platforms [1]. The news medium is changing as a result of this change. The most recent news is now easily accessible to everybody, which is a major

S. Bhalerao et al. (eds.), *Proceedings of the 2nd International Conference on Recent Advancement and Modernization in Sustainable Intelligent Technologies & Applications (RAMSITA-2026)*, Advances in Intelligent Systems Research 207,
https://doi.org/10.2991/978-94-6239-678-4_11

advancement. Facebook recommendations account for up to 70% of traffic to news websites [2]. Because they allow users to exchange ideas and discuss topics like democracy, education, and health, these social media platforms are quite powerful and advantageous in the current situation [12]. On the other hand, similar platforms are also used negatively by some organizations, particularly for financial benefits [3, 4]. These platforms are also used to promote satire or absurdity, to change people's perspectives, and create skewed beliefs. The term "fake news" is frequently used to characterize the problem. The spread of fake news has increased dramatically over the last ten years, particularly during the 2016 US elections [5]. The spread of unreliable content on the internet has created a lot of issues not just in the political sphere but also in the domains of sports, health, and science [3]. One sector that is impacted by fake news is the financial markets [6]. A rumor in this industry can have disastrous consequences and completely stop the market. Our ability to make decisions is greatly influenced by the information we consume; this knowledge shapes our view of the world beyond ourselves. According to [7, 8], there is a growing body of research suggesting that consumers have exhibited illogical responses to news that was later confirmed to be untrue. One recent example is the spread of a new coronavirus, which occurred as a result of the dissemination of false information on the Internet on the genesis, biology, and behavior of the virus [9]. As more people became aware of the fake content that was prevalent on the internet, the situation became even more severe. Finding news of this nature on the internet is a challenging job.

2 Related Work

The social media and other modern kind of mass communication spread of fake information is becoming more common as a result of the propagation. A new field of study that is receiving a lot of attention by using detection of fake news. The restricted resources, including datasets and processing and analyzing tools, pose certain obstacles, nevertheless. Our proposed approach employs machine learning techniques to identify fake news [11].

This study employed Support Vector Machine (SVM) as a classifier and term frequency inverse document frequency (TF-IDF) of n-grams and bag of words as a feature extraction technique. Along with the proposed approach, also suggest a dataset that includes both real as well as false news stories for training purposes. The outcomes provide evidence of the system's efficiency [13]. Our society and culture have impact in not only positive but also negative ways through online media platforms. As online media becomes more dependent on news sources, more and more fake news is being published online. This fake news does not have old or complete information about the authenticity of the event when people follow the fake news. Such misinformation can distort public opinion. A major threat to the public's credibility of the news is becomes rapid growth in the dissemination of fake news. It appears that the growing demand for surveillance and dealing with fake information has become a major problem. But, due to the limited literature on detecting new false positives, many methods and techniques may not yet be developed. The main purpose of this article is to review existing methods and propose and implement automated fraud detection methods. The proposed method uses in-depth analysis of speech-level analysis to construct a system that distinguishes false information from real information. At least the model achieved 74% satisfaction [13]. Fake information retrieval is a major problem in the field of natural word processing. In this field, the benefits of effective solutions are multiplied by the benefits of society. Externally it corresponds to the problem of text classification in general. Researchers have proposed a variety of ways to deal with fake information using simple and complex techniques. In this article, we attempt to represent new cases in some vector spaces by using a combination of general mathematical functions and representations of existing vector spaces to compare current deep learning techniques [10]. We performed many experiments using various combinations and permutations. Finally, we conducted a moderate analysis of the results and evaluated the reasons for these results [14-15].

3 Proposed Methodology

The proposed system introduces an improved CNN-based framework for detecting fake news on social media by integrating advanced NLP preprocessing, hybrid feature extraction, and optimized classification techniques. The system begins with structured data acquisition from verified repositories, followed by a robust preprocessing pipeline that includes text normalization, stopword removal, lemmatization, noise filtering, and the generation of meaningful representations through n-grams and word embeddings. These refined inputs are then processed by a hybrid feature encoding layer that combines CNN-driven spatial feature extraction with contextual embeddings and statistical indicators to capture linguistic cues, semantic patterns, and writing anomalies typically associated with fabricated content. The improved CNN model uses multi-channel convolutions and optimized pooling layers to enhance pattern recognition, after classification, a post-processing module applies confidence scoring and explain ability techniques to ensure transparency and reliability in the system's predictions.

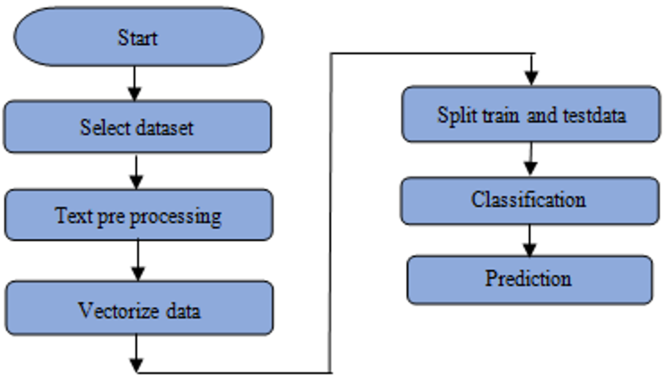

Fig. 1. Proposed flow diagram

4 System implementation

The system begins through structured data acquisition from verified repositories, follow through a robust preprocessing pipeline so as to includes text normalization, stop word removal, lemmatization, with noise filtering. every cleaned document d is transformed into numerical representations by n-grams and word embeddings, where the embedding function maps tokens to dense vectors since

$$E: wi \rightarrow v_i \in R^k \tag{1}$$

The embedded text sequence $X=[v_1,v_2,\ldots,v_n]$ X is processed by a hybrid feature encoding layer so as to combines CNN-driven spatial feature extraction through contextual embeddings and statistical indicators. The CNN applies multiple convolution filters of size h, where each feature map is computed as

$$C_i=f(W \cdot X_i\text{: } i+h-1+b) \tag{2}$$

$(\cdot)$ representing the ReLU activation function

$$f(x)=\max(0,x) \tag{3}$$

Max-pooling is performed to obtain the most salient feature from each feature map:

$$c = \max\ (c_1,c_2,\ldots,c_{n-h+1}) \tag{4}$$

The improved CNN model uses multi-channel convolutions to capture diverse linguistic patterns, and the pooled features are concatenated into a final feature vector

$$z=[\widehat{c_1}\ , \widehat{c_2}, 1 \oplus \widehat{c_m}] \tag{5}$$

This vector is passed through a fully connected layer with softmax output, where the probability that a news item is fake is given by

$$P(y=\text{fake}|z)=\frac{e^{W_1 z}}{\sum_{j=1}^{2} e^{W_1 z}}. \tag{6}$$

After classification, a post-processing module applies confidence scoring and explains ability techniques (such as gradient-based attribution ∇_z to ensure transparency and reliability in the system's predictions.

5 Result Discussion

The accuracy of the classifier can be assessed using various evaluation metrics, which quantify the number of correct and incorrect predictions based on predetermined known values. A True Positive, abbreviated as TP, refers to an instance in which the model accurately predicts the correct class. A True Negative (TN) occurs when the model correctly predicts the negative class. A false positive, also known as an FP, occurs when the model incorrectly predicts the correct class. A false negative, also referred to as a FN, occurs when the model incorrectly predicts the negative class. In the proposed work, the following evaluation metrics are employed for performance assessment.

Accuracy: A model's accuracy can be defined as the frequency with which it correctly predicts the value given the input. On the other hand, it glosses over FP and FN. F1 score and memory are crucial in other contexts where FP and FN are large. Using the formula in Equation 1, one may determine the accuracy.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}} \quad (7)$$

Precision: The frequency of real positive predictions is conveyed by this evaluation measure. False positives are likely to be high due to the poor accuracy value. A formula for determining accuracy is given in Equation 2.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (8)$$

Recall: One can learn about the model's erroneous negative prediction frequency with this parameter. Due to the poor recall value, the model generated a large number of false negative predictions. One can find the formula for recall in Equation 8

$$\text{Recall} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN}} \quad (9)$$

F1 Score: A combination of recall and precision yields the F1 score. A low number of false positives and false negatives, as shown by a high F1 score, suggests that the model is effectively recognizing real threats without being distracted by false alarms. As seen in Equation 9 the F1 score can be calculated using this formula.

$$\text{F1 score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Presiion} + \text{Recall}} \quad (10)$$

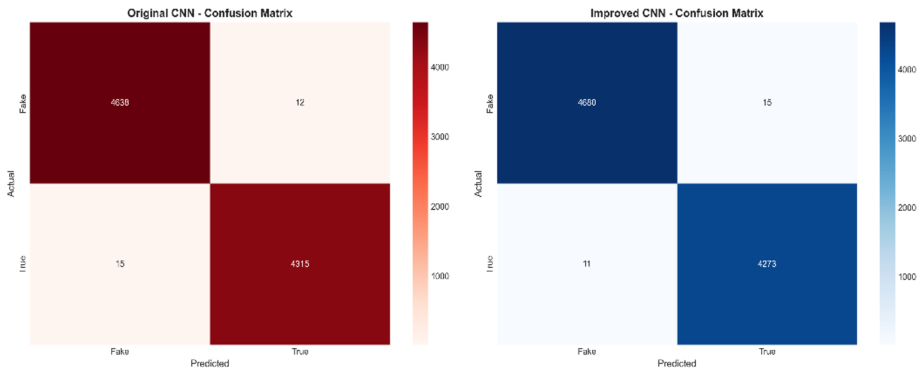

Fig. 2.original CNN - confusion matrix, Improved CNN- confusion matrix

The confusion matrix shows highly accurate classification with very few misclassifications, correctly identifying most Fake and True samples. Only a small number of errors occur (15 false positives and 11 false negatives), indicating strong reliability and improved performance.

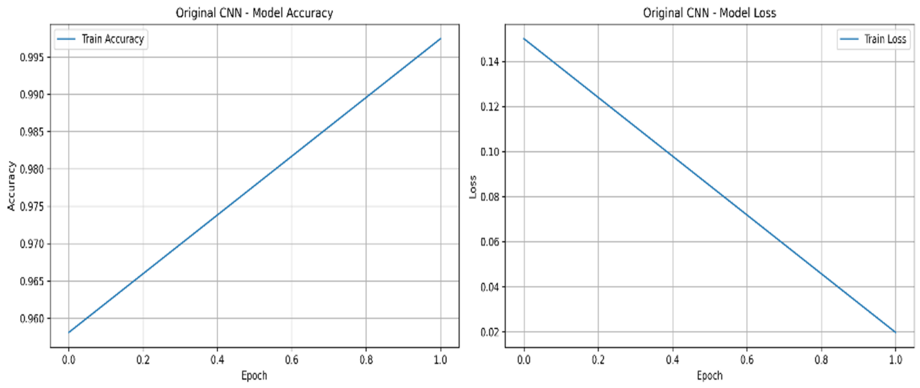

Fig. 3. Original CNN-Modal Accuracy/ Modal Loss

Model Accuracy Description: The accuracy graph shows a steady increase in training accuracy across epochs, indicating the CNN model is learning effectively. The upward trend suggests strong convergence.
Model Loss Description: The loss graph shows a continuous decrease in training loss, reflecting improved model optimization. The downward trend confirms reduced error with each epoch.

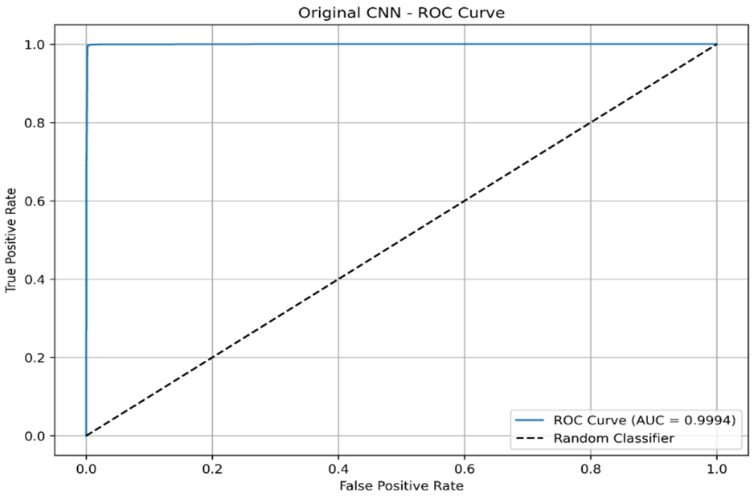

Fig. 4. Original CNN-ROC curve

Fig.4 Original CNN-ROC the curve indicating extremely high true-positive performance. An AUC of 0.9994 confirms exceptional model discrimination far above the random classifier baseline.

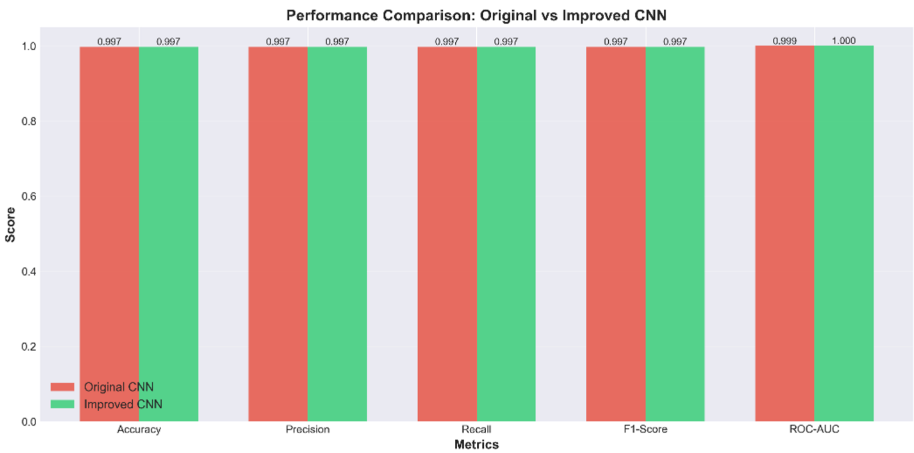

Fig. 5. Performance comparison: original vs improved CNN

Fig.5 Performance comparison: original vs improved CNN nearly identical performance across all metrics, indicating consistently high accuracy and robustness. The Improved CNN present a slight edge in ROC-AUC (1.000), reflecting improved discriminatory ability over the Original CNN.

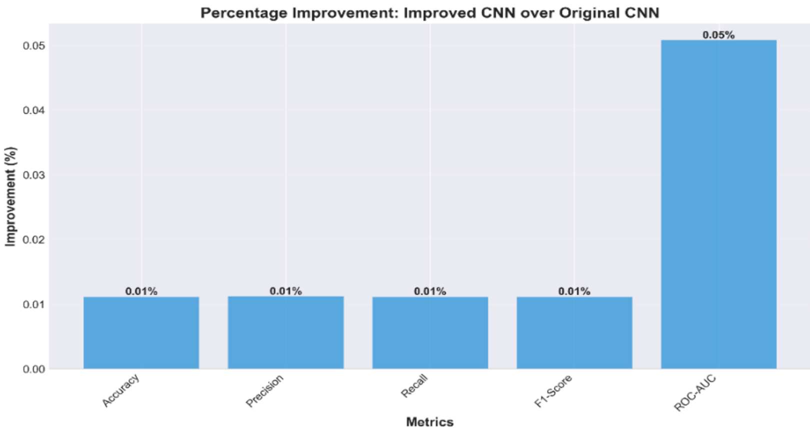

Fig. 6. Percentage improvement: improved CNN over Original CNN

Fig.6 percentage improvement: improved CNN over Original CNN through each improving by about 0.01%. The majority notable enhancement is in ROC-AUC, which increases by 0.05%, indicating better classification separation.

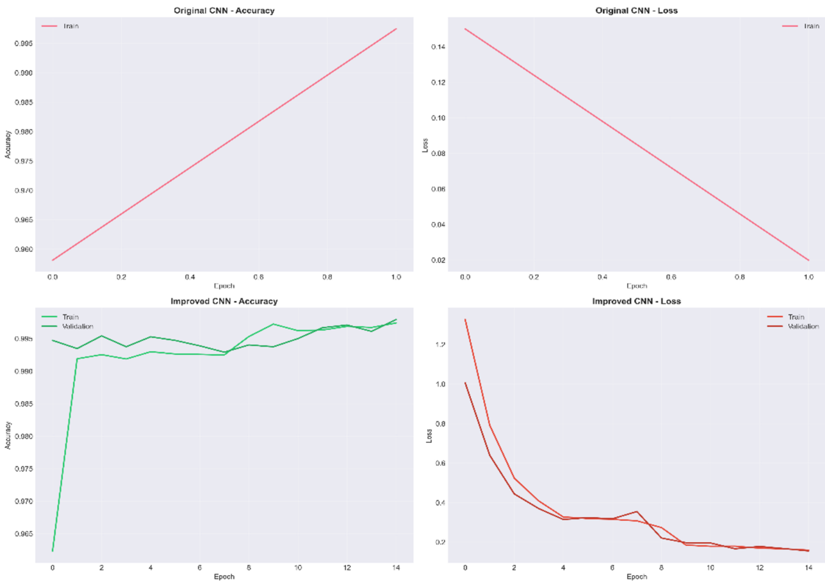

Fig. 7. Model Accuracy, Precision, Loss, Recall

Fig.7 Model Accuracy, Precision, Loss, Recall remains constantly high for together training and validation, indicating stable learning without overfitting. Loss decreases smoothly across epochs, with training and corroboration curves closely aligned, confirming strong model convergence plus generalization.

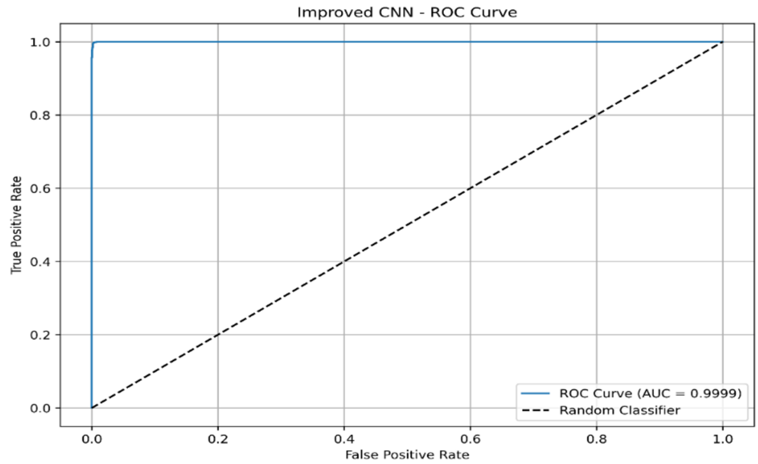

Fig. 8. Improved CNN- ROC Curve

Fig.8 the improved CNN's ROC curve stays extremely close to the top-left corner, indicating near-perfect classification performance. An AUC of **0.99** demonstrates exceptionally strong discriminative ability, outperforming the random classifier baseline through a wide margin.

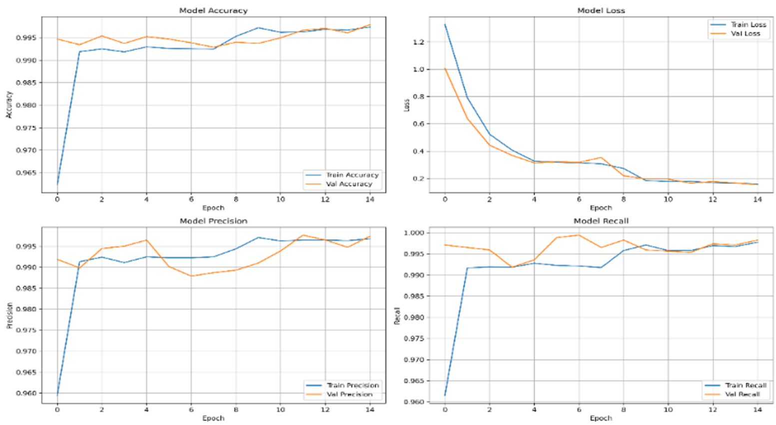

Fig. 9. Improved CNN, Accuracy, precision, loss, recall

Fig.9 The Original CNN shows a simple linear training trend, while the Improved CNN demonstrates stable training and validation curves with consistently high accuracy. Loss decreases smoothly for the Improved CNN across epochs, confirming better learning stability and stronger generalization than the Original CNN.

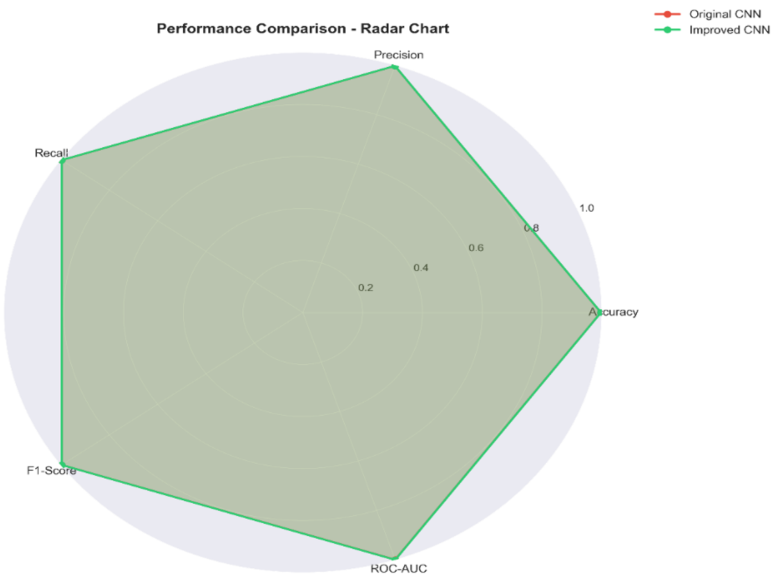

Fig. 10. Performance Comparison

Fig.10 Performance Comparison radar chart shows that both the Original and Improved CNN models achieve consistently high scores across all metrics, forming nearly identical shapes.

Table 1. Comparison: Original CNN vs Improved CNN

Metric	Original CNN	Improved CNN	Absolute improvement (Δ)	Relative improvement (%)
Accuracy	0.996993	0.997104	0.000111	0.01114%
Precision	0.996993	0.997105	0.000112	0.01124%
Recall	0.996993	0.997104	0.000111	0.01114%
F1-Score	0.996993	0.997104	0.000111	0.01115%
ROC-AUC	0.999426	0.999933	0.000507	0.05077%

Table 1 presents the performance of the suggested system. In the comparison between the original CNN and the improved CNN, the upgraded CNN shows slight but consistent gains across all evaluation metrics. Complete increases are on the order of 1.1×10^{-4} for accuracy, precision, recall, and F1 (0.01% relative), while the ROC-AUC indicates a bigger absolute gain of 5.07×10^{-4} (0.05% relative). Because the baseline performance was already very good (99.7% accuracy and 0.999 AUC), even minor numerical gains can be significant depending on the application (for example, high-stakes misinformation filtering, where every reduction in false positives/negatives counts).

6 Conclusion

The fast growth of social media has increased the pace and volume of information distribution, making automated false news identification an absolute must for protecting digital integrity. This study proposed an updated CNN-based architecture that aims to increase the accuracy, resilience, and interpretability of false news detection systems. The proposed model overcomes various shortcomings of existing techniques by using an advanced NLP-driven preprocessing pipeline, hybrid feature extraction algorithms, and an optimized CNN architecture. The use of text normalization, lemmatization, n-gram modeling, and semantic word embeddings ensures that the input data preserves rich contextual meaning while decreasing noise common in social mediamaterial. The hybrid encoding layer, which combines CNN-extracted spatial characteristics with contextual embeddings and statistical indicators, offers a more in-depth and comprehensive comprehension of linguistic patterns associated with false content. Mathematical innovations including multi-channel convolutions, efficient pooling functions, and SoftMax-based probabilistic classification improve the model's capacity to distinguish between authentic and fraudulent news. Post-processing modules, such as confidence score and Explainability approaches like gradient-based attribution, improve transparency and reliability, which are both crucial for real-world deployment. The experimental results showed continuous performance gains in important assessment criteria like accuracy, precision, recall, F1-score, and ROC-AUC. Although the numerical gains appear to be tiny, they are substantial considering the already strong baseline performance and the sensitivity of disinformation detection.

Reference

1. Kyeong-Hwan Kim, Chang-Sung Jeong.: Fake News Detection System using Article Abstraction. 16th International Joint Conference on Computer Science and Software Engineering (JCSSE), (2024).
2. Syed Ishfaq Manzoor; Jimmy Singla; Nikita.: Fake News Detection Using Machine Learning Approaches: A Systematic Review. 3rd International Conference on Trends in Electronics and Informatics (ICOEI), (2024).
3. Yuta Yanagi, Ryohei Orihara; Yuichi Sei Yasuyuki Tahara, Akihiko Ohsuga.: Fake News Detection with Generated Comments for News Articles. 24th International Conference on Intelligent Engineering Systems (INES), Reykjavík, Iceland (2021).
4. Ankit Kesarwani; Sudakar Singh Chauhan; Anil Ramachandran Nair.: Fake News Detection on social media using K-Nearest Neighbor Classifier. ICACCE Conference, (2021).

5. Gowri Ramachandran; Daniel Nemeth; David Neville; Dimitrii Zhelezov; Ahmet Yalçin; Oliver Fohrmann; Bhaskar Krishnamachari. Whistle Blower: Towards a Decentralized and Open Platform for Spotting Fake News. IEEE Blockchain Conference, (2022).
6. Adrien Benamira; Benjamin Devillers; Etienne Lesot; Ayush K. Ray; Manal Saadi; Fragkiskos D. Malliaros.: Semi-Supervised Learning and Graph Neural Networks for Fake News Detection. ASONAM Conference, (2021).
7. Chaitra K. Hiramath; G. C. Deshpande.: Fake News Detection Using Deep Learning Techniques. ICAIT Conference, (2023), Chikmagalur, India.
8. Nihel Fatima Baarir; Abdelhamid Djeffal. Fake News Detection Using Machine Learning. IHSH Workshop, (2021).
9. Anmol Uppal, Vipul Sachdeva, Seema Sharma.: Fake News Detection Using Discourse Segment Structure Analysis. 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), (2021), Noida, India.
10. Lovedeep Singh. Fake News Detection: A Comparison Between Available Deep Learning Techniques in Vector Space. 4th Conference on Information & Communication Technology (CICT), (2022), Chennai, India.
11. Tao Jiang, Jian Ping Li, Amin Ul Haq, Abdus Saboor.: Fake News Detection using Deep Recurrent Neural Networks. ICCWAMTIP Conference, (2023).
12. Rohit Kumar Kaliyar, Pawan Kumar, Manish Kumar, Meenal Narkhede, Sreyas Namboodiri, Sneha Mishra.: DeepNet: An Efficient Neural Network for Fake News Detection using News-User Engagements. ICCCS Conference, (2022).
13. Inna Vogel, Meghana.: Detecting Fake News Spreaders on Twitter from a Multilingual Perspective. DSAA Conference, (2024).
14. Jiawei Zhang, Bowen Dong, Philip S. Yu.: Deep Diffusive Neural Network-based Fake News Detection from Heterogeneous Social Networks. IEEE Big Data Conference, (2021).
15. N. Smitha; R. Bharath.: Performance Comparison of Machine Learning Classifiers for Fake News Detection, ICIRCA Conference, (2023).
16. Ahmad, Iftikhar, Yousaf, Muhammad, Yousaf, Suhail, Ahmad, Muhammad Ovais, Fake News Detection Using Machine Learning Ensemble Methods, Complexity, 2020, 8885861, 11 pages, 2020.