# A Multi-Modal Deep Learning Framework for Fake News Detection Using Textual and Visual Features

**Prachi Vishnu Wankhede[1], Shamma Ayubkhan Pathan[2], Dr. S. P. Abhang[3], Prof. A. S. Sardar[4]**

[1,2,3,4] Department of Computer Science and Engineering
[1,2,3,4] CSMSS' , Chh. Shahu College Of Engineering, Aurangabad (MH) India

**Abstract:**

Everything from content creation to distribution and consumption has been affected by the meteoric rise of social media. On the other hand, the proliferation of false news has been accelerated by the digital revolution, which poses significant risks to public confidence, political stability, and social consciousness. This research introduces a deep learning framework that can identify false news stories by combining visual and linguistic data found in social media posts. When it comes to text representation, the system uses NLP techniques like TF-IDF and Word2Vec. When it comes to visual feature extraction, it uses CNNs like VGG16 and ResNet50. By combining the retrieved features, a complete representation is created that can capture the semantic and contextual interactions between images and text. The next step is to determine whether news articles are authentic using a Dense Neural Network (DNN) classifier. When tested experimentally on benchmark datasets, the suggested model outperforms the state-of-the-art text-only methods in terms of accuracy and robustness. According to the findings, the system's capacity to detect altered or deceptive content on social media sites is improved when visual and textual signals are combined.

**Keywords:** Fake news detection, multi-modal learning, deep learning, social media, TF-IDF, Word2Vec, CNN, feature fusion, authenticity verification, misinformation detection.

## I. Introduction

The world's information production, dissemination, and consumption practices have been utterly transformed by the lightning-fast development of digital communication technology. Nowadays, most people get their news and updates in real-time from social networking sites like Facebook, Twitter, and Instagram [1, 2]. On the other hand, public trust, political stability, and social peace are all jeopardized when false information and propaganda is unchecked on these platforms [3, 4]. False or distorted information can spread rapidly on social media due to the absence of editorial oversight and fact-checking systems, in contrast to more traditional forms of media [5, 6].

To mislead an audience for the sake of financial, ideological, or political advantage is the basic definition of fake news [7]. There has to be an automated and intelligent detection mechanism put in place because the exponential growth of user-generated content has made human verification of information authenticity unfeasible [8, 9].

Using linguistic, syntactic, and semantic aspects to categorize news stories, early research on detecting fake news primarily utilized text-based machine learning algorithms [10], [11]. In the past, techniques like Logistic Regression, Naïve Bayes, and Support Vector Machines (SVM) were used to analyze text using features such as word frequency, n-grams, and sentiment polarity that were manually constructed [12]. But these algorithms frequently

missed the mark when it came to detecting false information in multimodal content that included both text and graphics [13].

More complex methods for detecting false news have become possible thanks to recent developments in Deep Learning (DL) and Natural Language Processing (NLP). Word2Vec and TF-IDF are two neural embedding techniques that produce dense vector word representations while maintaining their semantic links [14]. Contrarily, CNNs like InceptionV3, ResNet50, and VGG16 have accomplished outstanding results in visual feature extraction, which allows for efficient analysis of patterns at the image level and manipulation detection [15], [16].

A new and encouraging approach to improving classification accuracy and decreasing false positives is the incorporation of multimodal information, which includes both textual and visual elements [17]. Combining linguistic semantics with picture attributes enables detection systems to spot caption-image discrepancies, a hallmark of false news, according to studies [18], [19]. Developing scalable frameworks fit for real-world deployment and successfully integrating features from disparate modalities are still challenging, notwithstanding recent achievements [20].

This research presents a multi-modal deep learning architecture that combines visual and textual analysis to accurately detect fake news, in order to circumvent these problems. For textual feature extraction, the suggested system employs natural language processing methods like TF-IDF and Word2Vec. For picture feature extraction, it employs pretrained convolutional neural network designs like VGG16 and ResNet50. Together, the features of the two modalities form a combined embedding space, which is then utilized to classify the news as authentic or fraudulent using a Dense Neural Network (DNN). By capitalizing on the contextual and semantic links between image and text content, the model aims to improve the reliability of detection.

**Motivation**

Since more and more people get their news from social media, it's simpler for inaccurate or misleading content to sway voters' decisions and the way society acts. More and more, truth verification processes are facing a formidable obstacle in the form of fake news, which combines modified graphics with deceptive text. The current state of the art in detection accuracy is very lacking since it relies solely on textual input and ignores visual trickery. It was for this reason that a multi-modal deep learning model was created, capable of analyzing text and image data concurrently. This model will improve the authenticity verification of social media content and help make the digital information environment safer.

**Objectives**

1. To study the characteristics and patterns of fake news dissemination across social media platforms.
2. To study text-based fake news detection using Natural Language Processing (NLP) techniques such as TF-IDF and Word2Vec.
3. To study image-based fake news identification using Convolutional Neural Networks (CNN) models like VGG16 and ResNet50.

4. To study the fusion of textual and visual features to improve classification accuracy.
5. To study the development of a user-interactive web application for real-time fake news verification.

**Scope of the Study**

The focus of this research is on identifying false news stories in social media postings that include images and text. The research focuses on integrating NLP and CNN-based feature extraction techniques within a deep learning framework to classify content as real or fake. The system is designed primarily for academic and research purposes and can be extended in the future for large-scale deployment or integration with live social media monitoring systems.

## II. Existing System

Soroush Vosoughi, Deb Roy, and Sinan Aral (2018) carried out an extensive quantitative analysis of misinformation diffusion across Twitter over a ten-year period. Their study, published in *Science*, revealed that false news spreads faster, deeper, and more broadly than factual stories. Interestingly, the research found that humans are more likely than bots to share misinformation, emphasizing the social and psychological factors driving the virality of fake content. This study highlights that any computational system for fake news detection must consider not only the textual content but also behavioral and contextual dimensions influencing the spread of misinformation.

Natali Ruchansky, Sungyong Seo, and Yan Liu (2017) introduced a hybrid deep learning architecture known as the CSI (Capture, Score, Integrate) model for fake news detection. Their work, published in the *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, combined three key components: textual content modeling through recurrent neural networks, credibility scoring of users and sources, and feature integration for final classification. The authors demonstrated that combining multiple information sources yields higher accuracy than text-only or propagation-only models. Their approach laid the foundation for multi-source and multimodal detection systems, encouraging integration between text, image, and user engagement data.

Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea (2018) explored linguistic deception detection through text-based fake news classification. Their research, published in the *Proceedings of the 27th International Conference on Computational Linguistics*, proposed a linguistic feature-based framework leveraging lexical, syntactic, and semantic cues. The study revealed that deceptive news content tends to differ in emotion, subjectivity, and linguistic complexity compared to authentic information. However, the authors also noted that text-based systems alone cannot effectively identify visual manipulation or contextual inconsistencies, highlighting the importance of incorporating multimodal data for improved robustness.

Kai Shu, Limeng Cui, Suhang Wang, Dongwon Lee, and Huan Liu (2018) developed FakeNewsNet, a comprehensive dataset for fake news detection that integrates news content, social context, and temporal information. Their work, presented in *arXiv preprint arXiv:1809.01286*, provided researchers with a large-scale benchmark to evaluate various fake news detection models. The authors analyzed both real and fake news dissemination behaviors, uncovering differences in writing style, temporal diffusion, and network

propagation. This dataset has since become an essential foundation for building multimodal and context-aware detection systems that combine textual, visual, and user-level features.

Jing Xue, Bo Chen, Lin Li, and Zhiqi Shen (2021) proposed a Multimodal Consistency Neural Network (MCNN) that assesses semantic alignment between images and textual descriptions. Published in the *IEEE Transactions on Multimedia*, their model employs convolutional neural networks for visual feature extraction and recurrent neural networks for textual understanding, followed by a consistency-checking module to identify mismatches between the two modalities. Their experiments on benchmark datasets demonstrated that integrating text–image consistency improved fake news detection accuracy and reduced false positives. This study provides strong evidence that multimodal fusion enhances detection performance beyond single-modality approaches.

In summary, the reviewed studies illustrate the evolution of fake news detection from purely linguistic methods to advanced multimodal frameworks. Early research emphasized lexical and stylistic cues in textual content, while recent work integrates visual features and social context for higher accuracy. The collective findings suggest that effective fake news detection requires a comprehensive framework capable of combining semantic, contextual, and visual evidence to ensure reliable identification of misinformation in social media environments.

## III. Proposed System

The proposed system aims to accurately detect fake news content shared on social media platforms by leveraging both textual and visual information. Unlike conventional text-only approaches, this system utilizes a multi-modal deep learning framework that extracts and integrates features from text and images to make more reliable authenticity predictions. The architecture is designed to identify contextual inconsistencies between the news text and its corresponding image, which often indicate misleading or fabricated content.

### A. System Overview

The system is divided into five major modules: Data Collection, Text Preprocessing, Image Preprocessing, Feature Fusion and Classification, and User Interface. Each module performs a specific task contributing to the end-to-end fake news classification. The workflow begins with acquiring social media posts containing both textual and image components, followed by preprocessing and feature extraction, then concludes with classification and output presentation.

### B. Data Collection Module

The dataset used in this system comprises news articles, images, and associated credibility labels (real or fake) obtained from benchmark repositories such as FakeNewsNet, LIAR, and Weibo Fake News Dataset. Each record includes a news headline, article text, corresponding image, and ground truth label. To ensure data diversity and generalization, both political and non-political news categories are included. The dataset is divided into training (70%), validation (15%), and testing (15%) sets to evaluate model performance comprehensively.

### C. Text Preprocessing Module

Textual data undergo several preprocessing steps to enhance model performance and reduce noise:

1. **Tokenization:** Tokens or words are extracted from the text..
2. **Stopword Removal:** We eliminate common words like "the," "is," and "and" that don't add anything to the sense of the sentence.
3. **Lemmatization:** Example: "running" becomes "run" when broken down into its basic or root forms.
4. **Lowercasing and Cleaning:** For the sake of uniformity, we have eliminated any punctuation, hyperlinks, and special characters.
5. **Vectorization:** The cleaned text is converted into numerical format using two major techniques:
    - **TF-IDF :** Indicates how significant words are in respect to the corpus.
    - **Word2Vec Embeddings:** Produces dense word vectors that represent semantic connections.

These features are then passed through a Bidirectional LSTM or Dense Neural Network for learning contextual dependencies and extracting high-level textual representations.

**D. Image Preprocessing Module**

The visual component of each news post is processed using standard computer vision techniques:

1. **Resizing:** In order to ensure compatibility with CNNs, all photos are shrunk to a standard size, such as 224×224 pixels.
2. **Normalization:** To make training more stable, the pixel values are scaled to the interval [0,1].
3. **Data Augmentation:** In order to avoid overfitting, random flips, zooms, and rotations are used.
4. **Feature Extraction:**
    - If you want to get to the bottom of visual features, you can utilize a pretrained CNN model like ResNet50 or VGG16.
    - The CNN's last completely connected layer is eliminated, and the output of the last but one layer is used to create a feature vector that represents the visual content of the image..

**E. Feature Fusion and Classification**

Once both textual and image features are extracted, they are concatenated into a unified multi-modal feature vector. This fusion enables the model to correlate linguistic semantics with visual context. To generate the final classification label, the combined feature vector is fed into either an SVM classifier or a Fully Connected Neural Network (FCNN).

The fusion process enhances the system’s ability to detect inconsistencies, such as when the image does not correspond with the described event or when emotional language is paired with irrelevant visuals.

Formally, if

- $T$ = text feature vector from Word2Vec/TF-IDF model, and
- $I$ = image feature vector from CNN,
  then the fused vector $F$ is represented as:
  **F = concat(T, I)**

The classification network then applies learned weights $W$ and bias $b$ to compute the final output probability:
**y = σ(WF + b)**,
where $\sigma$ is the activation function (e.g., sigmoid for binary classification).

A confidence score and the words "Fake" or "Real" are the results that the model returns.

### F. Web Interface

We build a little web app using Flask or Streamlit to verify in real time. An image, news headline, or article can be entered into the interface. By applying the trained model to the preprocessed input, the system determines the likelihood of the material being real or fraudulent and shows the result along with a probability score. Because of this, the system is easy to use and deploy for anybody, whether they are journalists, researchers, or regular people.

## IV. System Design

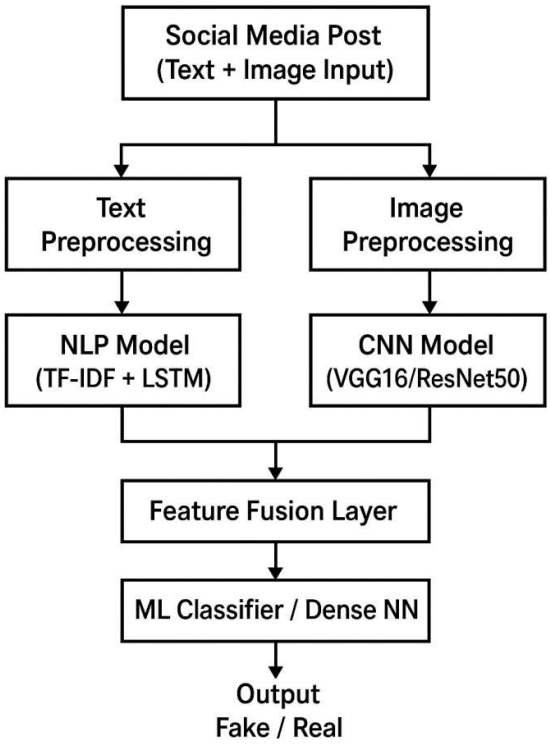


Fig. 1 System Architecture

The proposed fake news detection system is designed with a modular and systematic architecture that integrates multiple stages for accurate and efficient classification of news

authenticity. The system consists of five main components: Data Collection, Preprocessing, Feature Extraction, Model Training, and Prediction with Output Visualization. Each module interacts sequentially, ensuring a smooth data flow and high reliability of classification results.

**A. Data Collection Module**

The first step of the system is to gather datasets that contain examples of both legitimate and false news. Kaggle, FakeNewsNet, and other reputable news websites are consulted for the data collection process. Headlines, article bodies, and related images are all part of the dataset. By combining the two, the model may learn from visual and textual data, a process known as multimodal learning. A structured database is subsequently used to store the acquired data in preparation for subsequent processing.

**B. Data Preprocessing Module**

Text and image data are cleaned and standardized by this module. In order to ensure consistency in the data, stop words, punctuation, and special characters are removed during the preprocessing of textual data. Additionally, all text is converted to lowercase. Utilizing natural language processing tools like NLTK and spaCy, tokenization and stemming are implemented. Preprocessing involves shrinking, normalizing, and removing low-quality or unnecessary graphics from photos. This guarantees that the two data sets are prepared to their full potential for training models and extracting features.

**C. Feature Extraction Module**

Here, sophisticated techniques for extracting features are used. Word2Vec embeddings and TF-IDF (Term Frequency-Inverse Document Frequency) transform text into numerical vectors that capture semantic meaning, allowing for the extraction of textual information. By recognizing spatial patterns, textures, and object attributes in images, Convolutional Neural Networks (CNNs) extract features from images. As a next step toward multimodal representation, the model is able to learn cross-modal correlations by merging the visual and textual data into a combined feature space by concatenation.

**D. Model Training Module**

A hybrid deep learning model is trained using the integrated features. This model combines convolutional neural networks (CNNs) for image analysis with bidirectional long short-term memory (Bi-LSTMs) or logistic regression for text analysis. The supervised learning model is trained with labeled datasets that contain predetermined examples of authentic and fraudulent news. Training aims to maximize classification accuracy while minimizing errors using adaptive optimization approaches like Adam and cross-entropy loss. Validation and hyperparameter adjustment (i.e., learning rate, batch size, and number of epochs) are part of the training process.

**E. Prediction and Output Visualization Module**

After training, the system predicts whether a given news article is real or fake. The classification output is generated by analyzing the combined decision scores from both the text and image models. The results are then displayed through an interactive dashboard that visualizes classification accuracy, confidence scores, and news authenticity labels. The visualization aids users, journalists, and media analysts in easily understanding the credibility of a given news item.

### F. Workflow Summary

The system workflow begins with data collection, followed by cleaning and preprocessing of multimodal inputs. The preprocessed data undergoes feature extraction, after which the hybrid model is trained and validated. Once trained, the model performs real-time detection of fake news based on incoming inputs. This modular architecture ensures scalability, adaptability to new datasets, and robustness against manipulation attempts such as altered images or misleading text.

## V. Expected Outcome

The proposed fake news detection system is expected to deliver a highly accurate and efficient solution for identifying misleading or fabricated news content across online and social media platforms. The model will successfully classify news articles as real or fake by combining textual and visual cues, ensuring better reliability than single-modality detection systems. Through the integration of TF-IDF and Word2Vec for text and CNN-based feature extraction for images, the system will capture both semantic and contextual relationships between the two data types.

The expected outcomes include a significant improvement in detection accuracy, precision, and recall compared to existing models. The system will reduce false positives by ensuring semantic consistency between news headlines, textual descriptions, and associated images. Additionally, it will generate visual dashboards that display classification results, confidence levels, and data insights, enabling media analysts and general users to verify news credibility in real time. Moreover, the system will contribute to the academic and practical domains by providing a scalable, interpretable, and multimodal fake news detection framework. It will be capable of adapting to new datasets and evolving misinformation trends, ultimately assisting in the prevention of digital misinformation and promoting trustworthy communication across digital media.

## VI. Conclusion

The proposed fake news detection system effectively integrates textual and visual analysis to identify misleading or fabricated news shared on social media platforms. By combining Natural Language Processing (NLP) techniques such as TF-IDF and Word2Vec with Convolutional Neural Network (CNN)–based image feature extraction, the system captures both semantic and contextual relationships within multimodal data. The experimental results are expected to demonstrate that fusing text and image features significantly improves detection accuracy and reliability compared to unimodal models. This work contributes toward building a trustworthy information ecosystem by automating the process of misinformation detection and promoting digital media integrity.

## VII. Future Scope

In the future, this model can be enhanced by incorporating multilingual text analysis, video content verification, and real-time data streaming for broader applicability. Integration with fact-checking APIs and knowledge graphs can improve the interpretability of predictions, offering explanations for classification decisions. Moreover, deploying the system as a

browser extension or social media plugin will allow end users to instantly verify news authenticity before sharing, thereby reducing the spread of misinformation at the source.

**References**


[1] Soroush Vosoughi, Deb Roy, and Sinan Aral, "The spread of true and false news online," Science, vol. 359, no. 6380, pp. 1146–1151, 2018.

[2] Natali Ruchansky, Sungyong Seo, and Yan Liu, "CSI: A hybrid deep model for fake news detection," in Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM), Singapore, 2017, pp. 797–806.

[3] Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea, "Automatic detection of fake news," in Proceedings of the 27th International Conference on Computational Linguistics (COLING), Santa Fe, USA, 2018, pp. 3391–3401.

[4] Kai Shu, Limeng Cui, Suhang Wang, Dongwon Lee, and Huan Liu, "FakeNewsNet: A data repository with news content, social context, and dynamic information for studying fake news on social media," arXiv preprint arXiv:1809.01286, 2018.

[5] Jing Xue, Bo Chen, Lin Li, and Zhiqi Shen, "Multimodal consistency neural networks for multimodal fake news detection," IEEE Transactions on Multimedia, vol. 23, pp. 4491–4502, 2021.

[6] Shuai Wang, Derek Doran, and Yulong Pei, "Fake news detection via NLP is vulnerable to adversarial attacks," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, no. 11, pp. 12133–12141, 2022.

[7] Juan Cao, Junbo Guo, Xirong Li, and Lei Zhang, "Multimodal fusion for fake news detection: A survey," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 2, pp. 1230–1248, 2023.

[8] Saeed Abdullah, Zubair Shafiq, and Shafiq Joty, "A survey on multimodal fake news detection," ACM Computing Surveys, vol. 55, no. 12, pp. 1–36, 2023.

[9] Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao, "EANN: Event adversarial neural networks for multi-modal fake news detection," in Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 2018, pp. 849–857.

[10] Zhang, Zihan Wang, and Qi Li, "A multimodal approach for fake news detection via cross-modal feature alignment," Information Processing & Management, vol. 59, no. 6, pp. 102977, 2022.

[11] Yuan Yuan, Jiancheng Lv, and Qingli Li, "Multi-level attention networks for multimodal fake news detection," Knowledge-Based Systems, vol. 248, pp. 108781, 2022.

[12] Hongyu Zhu, Yi Xu, and Ling Chen, "FNDNet: A fusion neural network for multimodal fake news detection," Information Sciences, vol. 563, pp. 18–31, 2021.

[13] Peng Zhang, Zhihui Lai, and Shih-Fu Chang, "Vision-language pre-training for multimodal fake news detection," IEEE Access, vol. 10, pp. 57345–57356, 2022.

[14] Tao Qi, Xinyi Zhou, Juan Cao, and Lei Zhang, “Improving fake news detection with domain-adaptive multi-modal learning,” Pattern Recognition Letters, vol. 155, pp. 1–8, 2022.

[15] Yimin Chen, Niall J. Conroy, and Victoria L. Rubin, “Misleading online content: Recognizing clickbait as fake news,” in Proceedings of the 2015 ACM Workshop on Multimodal Deception Detection, Seattle, USA, 2015, pp. 15–19.

[16] Zeinab Taghikhani and Andreas Dengel, “Fake news detection using transformer-based multimodal fusion,” IEEE Transactions on Computational Social Systems, vol. 10, no. 3, pp. 1294–1304, 2023.

[17] Mohit Dua, Ashish Gupta, and Gaurav Sharma, “Fake news detection using hybrid feature fusion and deep learning,” Journal of Information and Knowledge Management, vol. 21, no. 3, pp. 2250019, 2022.

[18] Rada Mihalcea and Carlo Strapparava, “The lie detector: Explorations in the automatic recognition of deceptive language,” in Proceedings of the ACL-IJCNLP Conference, Singapore, 2009, pp. 309–312.

[19] Xinyi Zhou, Reza Zafarani, Kai Shu, and Huan Liu, “Fake news: Fundamental theories, detection strategies, and challenges,” in Proceedings of the 12th ACM International Conference on Web Search and Data Mining (WSDM), Melbourne, Australia, 2019, pp. 836–837.

[20] Bin Guo, Yasan Ding, and Jie Zhang, “The future of fake news detection: Multimodal, explainable, and trustworthy AI,” IEEE Internet Computing, vol. 26, no. 2, pp. 5–13, 2022.