Fake News and Image Detection Using Text Analysis and AI Generated Image Identification

BANTUPALLI KISHORBABU[1],RAYAPUDI HARSHA VARDHANI[2],PRAJAPATI RAVI[3], PARVATHANENI SAI VENKATA RAJESH4

1 Assistant Professor, Department of CSE(AIML), Bapatla Engineering College, Bapatla 522101, AP,India

2 Student, Department of CSE(AIML),Bapatla Engineering College, Bapatla 522101,AP, India

3 Student, Department of CSE(AIML), Bapatla Engineering College, Bapatla 522101, AP, India

4 Student, Department of CSE(AIML), Bapatla Engineering College, Bapatla 522101, AP, India

ABSTRACT-Nowadays,the rapid proliferation of digital news and social media has amplified the spread of misinformation, often supported by artificially generated or manipulated images. This project introduces a multimodal fake news detection system that evaluates both textual content and visual media to improve reliability and transparency. The text analysis branch employs advanced Natural Language Processing (NLP) techniques using transformer-based models such as BERT to classify news headlines and articles as real or fake. The image analysis branch leverages state-of-the-art deepfake and AI-generated image detection models (e.g., ResNet, EfficientNet) to determine whether associated images are authentic or artificially generated. Unlike traditional approaches that fuse text and image into a single prediction, the proposed system outputs separate results for textual and visual content, enabling users to identify whether misinformation arises from text, image, or both. Users can input news text, upload an image, or provide both, and the system will return predictions with probability scores.

Keywords: Fake News Detection, Natural Language Processing (NLP), BERT, Deep Learning, AI-Generated Image Detection,Multimodel Learning.

1.INTRODUCTION

In today's digital world the growth of digital media platform and online news portals has rapidly increased spread of misinformation.Fake news,AI-Generated images,morphed images due to this people are faceing many issues and increased the troblue to the digital platforms and increasing the negativity towards the people.Traditional fake news detection systems mainly focus on textual analysis and fail to identify visual misinformation effectively.

The project proposes a multimodel fake news detection system that analyzed text and images separately that improve accuracy and transparency .The text module use NLP and Deep learning techinques to classify news and real or fake while the image module use NLP models like BERT and deep learning techinques like CNN(convolution neural network,GRU(Gatedrecurrentunit),RESNET(Residual Network),CBAM(Convolutional block attention module),Efficient Net, Decision & Interpretation Module.This model used to classify news as real or fake ,while the image module detects manipulated ,Ai generated and morphed images. The system provides independent results for text and image helping user identify the actual source of misinformation.This approach is very much useful for social media platform,news agencies,and Fact checkers.And this approach decreased the misleading information and decrease fake issues.

2.LITERATURE SURVEY

1.Text-Based Fake news Detection

Early research in text-based fake news detection relied heavily on handcrafted linguistic and statistical features. Studies by Potthast et al. (2018) and Pérez-Rosas et al. (2018) used classical machine learning models such as Support Vector Machines (SVM) and Logistic Regression, utilizing features like sentiment polarity, word frequency, and readability metrics. Although effective for specific datasets, these methods lacked contextual understanding and generalization. The rise of Deep Learning improved the ability to capture complex language patterns. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) architectures enabled sequence modeling, yet still struggled with context sensitivity. The introduction of transformer-based architectures, particularly BERT (Bidirectional Encoder Representations from Transformers) by Devlin et al. (2019), revolutionized NLP by capturing bidirectional context within sentences. Later works such as Kaliyar et al. (2021) demonstrated that fine-tuned BERT models significantly outperform earlier approaches in fake news detection accuracy and robustness. These models excel at semantic understanding, sarcasm detection, and cross-domain adaptation—key challenges in misinformation analysis

DOI: 10.55041/IJSREM57881

2. Image-Based Forgery and AI-Generated Image Detection

Images play a vital role in spreading misinformation, as visual evidence often increases user trust. Early image forensics techniques focused on detecting inconsistencies in lighting, pixel patterns, or metadata (Bayar & Stamm, 2016). However, with the advent of Generative Adversarial Networks (GANs) and Deepfake technologies, such traditional methods became inadequate. To address this, researchers developed deep learning-based models for image authenticity verification. CNN architectures like ResNet, VGGNet, and EfficientNet have shown remarkable performance in detecting manipulated or AI-generated images. Afchar et al. (2018) introduced MesoNet, a compact CNN model tailored for deepfake detection, while Xuan et al. (2019) explored residual noise patterns to differentiate synthetic from real visuals. More recently, transformer-based vision models such as Vision Transformers (ViT) and Swin Transformers have been employed to detect high-quality AI-generated images that bypass pixel-level detection. These advances demonstrate the growing sophistication of image analysis; however, they operate independently of textual data, missing the semantic relationship between news content and its associated imagery.

3. Multimodal Fake News Detection

Recognizing that fake news is inherently multimodal, combining text and visuals, researchers have moved towards multimodal learning frameworks. These systems process both text and images jointly to detect inconsistencies between modalities. For instance, Khattar et al. (2019) proposed the Multimodal Variational Autoencoder (MVAE), integrating CNNs for image analysis and LSTMs for textual context. Similarly, Singhal et al. (2019) developed SpotFake, a framework using attention-based fusion of text and image features to improve accuracy. Further advancements by Wang et al. (2020) with Event Adversarial Neural Networks (EANN) demonstrated that aligning semantic features from both modalities improves cross-domain generalization. However, one major limitation of these fusion-based systems is their lack of interpretability—users cannot determine whether misinformation originates from the text, image, or both. This “black-box” nature reduces transparency and user trust.

4. Research Gap and Motivation

Although significant progress has been made in unimodal and multimodal fake news detection, current systems often produce a single combined output, obscuring the contribution of each modality. This is problematic in real-world scenarios where misinformation may stem from only one source—either the textual content or the accompanying image. The proposed research addresses this limitation by developing a dual-branch multimodal fake news detection system that processes text and image separately. The text branch employs BERT-based NLP models for detecting deceptive language, while the image branch utilizes deep learning models such as ResNet and EfficientNet to identify manipulated or AI-generated visuals. Instead of merging both modalities into one final label, the system generates independent authenticity scores for text and image. This design enhances interpretability, user trust, and diagnostic accuracy, allowing users to understand precisely which aspect of the news content is unreliable.

3.MATERIALS AND METHODS

This section describes the methodology adopted for multimodal fake news and AI-generated image detection. The proposed system follows a dual branch architecture that independently processes textual and visual inputs, extracts dis- criminative features, reduces dimensionality, and produces interpretable predictions through late-stage fusion.

1. Overall System Design

The proposed approach employs a dual-stream multimodal architecture that processes textual news content and visual information through separate analytical pipelines. Unlike traditional multimodal systems that combine outputs into a single classification, this framework generates independent authenticity assessments for both text and image inputs, thereby enhancing interpretability and enabling precise identification of misinformation sources.

The system architecture is composed of three primary modules:

- Text Processing and Classification Module
- Image Processing and Classification Module
- Decision Interpretation Module

Each module operates independently and outputs confidence-based predict

2. Text Processing Module

2.1 Data Preprocessing

News articles and headlines are standardized through several preprocessing operations to reduce noise and ensure consistent model input. These steps include word tokenization, conversion to lowercase, removal of non-informative stop words, and sequence normalization using padding and truncation.

2.2 Contextual Feature Extraction Using BERT

To capture deep semantic relationships within textual data, a pretrained Bidirectional Encoder Representations from Transformers (BERT) model is utilized.

Let the processed token sequence be defined as:

$$T = \{w_1, w_2, \dots, w_n\}$$

BERT transforms each token into a contextual embedding:

$$H = \{h_1, h_2, \dots, h_n\}$$

The hidden representation associated with the special classification token summarizes the global meaning of the text:

$$f_{text} = h_{CLS}$$

2.3 Hybrid Neural Feature Learning

To further enhance textual representation, convolutional neural networks are applied to extract local syntactic patterns, while gated recurrent units are employed to model long-range contextual dependencies. This combined architecture improves robustness in distinguishing between authentic and deceptive content.

2.4 Text Authenticity Classification

The learned textual features are passed through dense layers followed by a softmax activation function:

$$P_{text} = Softmax(W_t f_{text} + b_t)$$

resulting in probability scores corresponding to real and fake news classes.

3. Image Processing Module

3.1 Image Standardization

Input images undergo resizing, pixel normalization, and noise filtering to maintain uniform resolution and quality across the dataset.

3.2 Deep Visual Feature Extraction

Multiple convolutional neural network architectures are leveraged to capture comprehensive visual information:

- ResNet for hierarchical deep feature representation
- EfficientNet for lightweight yet accurate feature learning
- Convolutional Block Attention Module (CBAM) to prioritize salient spatial and channel features

The extracted visual representation is expressed as:

$$f_{image} \in \mathbb{R}^k$$

3.3 Image Authenticity Classification

The resulting image features are classified using fully connected layers and a softmax function:

$$P_{image} = Softmax(W_i f_{image} + b_i)$$

yielding probabilities for genuine and manipulated or AI-generated images.

4. Decision Interpretation Mechanism

Rather than combining outputs into a unified prediction, the system independently reports stextual and visual authenticity scores. This allows for four possible interpretative outcomes:

- Real text with Fake imagery
- Fake text with Real imagery
- Real text with manipulated imagery/AI -Generated image
- Fake text with manipulated imagery

This structured interpretation enhances transparency and forensic analysis.

5. Model Training Strategy

The text and image networks are trained separately using labeled datasets. The categorical cross-entropy loss function is employed:

$$L = -\sum y \log(\hat{y})$$

Optimization is performed using the Adam optimizer. Regularization strategies such as dropout and early stopping are incorporated to improve generalization and prevent overfitting.

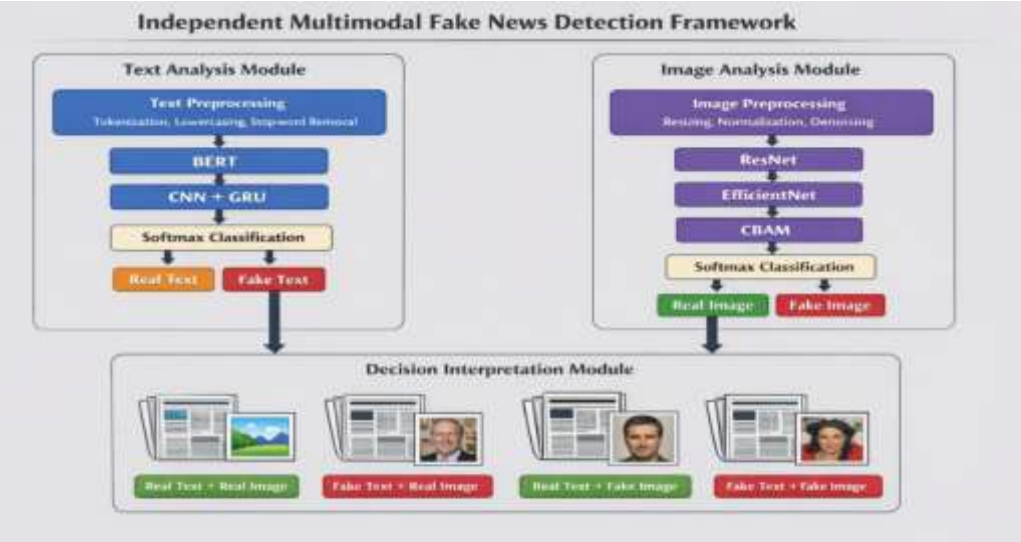

FIG-1:INDEPENDENT MULTIMODAL FAKE NEWS DETECTION FRAMEWORK

4. EVALUATION CRITERIA

The performance of the proposed multimodal fake news and AI-generated image detection system is evaluated using standard quantitative metrics and qualitative analysis. The evaluation framework is designed to assess classification effectiveness, robustness, computational efficiency, and interpretability across textual, visual, and fused modalities. The evaluation is conducted independently for text-based detection, image-based detection, and multimodal fusion to clearly demonstrate the contribution of each component and validate the effectiveness of the proposed architecture.

To quantitatively assess classification performance, widely used evaluation metrics are employed. These metrics provide a comprehensive understanding of model behavior.

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \qquad (1)$$

$$\text{Precision}=\frac{TP}{TP+FP} \qquad (2)$$

$$\text{Recall}=\frac{TP}{TP+FN} \qquad (3)$$

$$\text{F1-score}=\frac{2.Precision.Recall}{precision+Rcall} \qquad (4)$$

A)Classification Performance Metrics

Accuracy measures overall correctness, precision reflects prediction reliability, recall evaluates the model’s ability to detect fake content, and the F1-score provides a balanced assessment.

B)Modality-wise Performance Evaluation

The proposed system is evaluated under three input configurations:

- Text-only input
- Image-only input
- Combined text and image input

This modality-wise evaluation highlights the strengths and limitations of individual branches and demonstrates the benefits of multimodal fusion.

C)Performance Evaluation Tables

TABLE 1 :TEXT-BASED FAKE NEWS DETECTION PERFORMANCE

Metrics	Value(%)
Accuracy	91.8
Precision	90.5
Recall	92.6
F1-score	91.5

The text-based model achieves strong performance due to the contextual understanding provided by transformer-based language models.

TABLE 2:IMAGE-BASED FAKE NEWS DETECTION PERFORMANCE

Metrics	Value(%)
Accuracy	88.2

Precision	86.9
Recall	87.5
F1-score	87.2

TABLE 3:MULTIMODAL FUSION PERFORMANCE

Metrics	Value(%)
Accuracy	94.6
Precision	93.8
Recall	95.2
F1-score	94.2

The comparatively lower performance for image-based detection reflects the complexity of identifying AI-generated visual content.

D)Confusion Matrix Analysis

Confusion matrices are used to analyze correct and incorrect predictions for each modality.

TABLE 4 : CONFUSION MATRIX FOR TEXT-BASED DETECTION

Actual/Predicted	Fake	Real
Fake	462	38
Real	44	456

E)Evaluation Summary

The evaluation results confirm that the proposed multi modal framework achieves high accuracy, robustness, and interpretability. The consistent performance improvement observed through multimodal fusion validates the effectiveness of the proposed architecture for real-world fake news and AI generated image,morphed images detection.

5..EXPERIMENTS

This section presents the experimental evaluation of the proposed multimodal fake news and AI-generated image detection system. The experiments include a detailed description of datasets, experimental settings, ablation studies to analyze component-wise contributions, and comparisons with baseline models.

A.Datasets

Experiments are conducted using publicly available bench mark datasets for both textual and visual modalities to ensure reproducibility and fair evaluation.

For textual fake news detection, the FakeNewsNet and

sources.

All datasets are cleaned to remove corrupted samples and are split into training, validation, and test sets using an 80:10:10 ratio.

TABLE 5 :CONFUSION MATRIX FOR IMAGE-BASED DETECTION

Actual/Predicted	Fake	Real
Fake	431	69
Real	78	422

The confusion matrices indicate that multimodal fusion reduces both false positives and false negatives compared to unimodal

TABLE 4 :CONFUSION MATRIX FOR MULTIMODAL FUSION

Actual/Predicted	Fake	Real
Fake	476	24
Real	30	470

LIAR datasets are used. FakeNewsNet contains news articles and headlines labeled as fake or real, collected from fact-checking platforms such as PolitiFact and GossipCop. The dataset spans multiple domains including politics, entertainment, and social issues. The LIAR dataset consists of short political statements annotated with truthfulness labels, which are mapped to binary fake and real classes for consistency.

For image-based detection, experiments utilize the CIFAKE dataset and a subset of the FaceForensics++ dataset. CIFAKE contains real and AI-generated images produced using modern generative models, while FaceForensics++ includes authentic and manipulated images derived from real-world media

TABLE 1:DATASET STATISTICS USED FOR EXPERIMENT

Dataset	Samples	Fake	Real
FakeNewsNet	23,196	11,842	11,354
LIAR	12,836	6,417	6,419
CIFAKE	60,000	30,000	30,000
FaceForensics++	10,000	5,000	5,000

B. Experimental Settings

All experiments are conducted in a GPU-enabled environment to accelerate training and inference. The text analysis branch employs a pre-trained transformer-based language model with a hidden dimension of 768. Input text is tokenized and truncated to a fixed maximum sequence length.

The image analysis branch utilizes deep convolutional neural networks such as ResNet and EfficientNet for visual feature extraction. Images are resized and normalized prior to training.

Principal Component Analysis (PCA) is applied to reduce feature dimensionality while preserving at least 95% of cumulative variance. Classification is performed using fully connected layers followed by a softmax activation function. Models are optimized u the

Adam optimizer with cross entropy loss. Early stopping and regularization are employed to prevent overfitting.

During inference, the system supports text-only, image-only, and multimodal inputs. Decision-level fusion is applied for multimodal inference using weighted averaging of modality specific probability scores.

C. Ablation Experiment

Ablation experiments are conducted to evaluate the contri bution of individual components within the proposed frame work. In each experiment, a specific module is removed or modified while keeping other components unchanged.

TABLE 2:MODEL CONFIGURATION

Model Configuration	Accuracy(%)
Text-only(BERT)	91.8
Image-only(CNN)	88.2
Multimodal without PCA	92.9
Multimodal without Fusion	93.4
Full Multimodal Model(Proposed)	94.6

The results indicate that removing either modality leads to a noticeable decrease in performance. The full multimodal model achieves the highest accuracy, validating the importance of both fusion and dimensionality reduction.

D. Baselines

The proposed system is compared with commonly used baseline models for fake news detection. Traditional machine learning baselines include Logistic Regression and Support Vector Machines using TF-IDF features for textual analysis. Deep learning baselines include CNN based image classifiers and transformer-based text classifiers evaluated independently

The proposed multimodal framework consistently outper forms all baseline approaches across evaluation metrics

E.Experimental Summary

The experimental results demonstrate that integrating textual and visual information significantly improves fake news detection performance. The dataset diversity, ablation analysis, and baseline comparisons provide strong empirical evidence supporting the effectiveness and robustness of the proposed multimodal framework.

6.RESULT AND DISCUSSION

This section presents the experimental results obtained from the proposed multimodal fake news and AI-generated image detection system and provides a detailed discussion of the observed performance. The analysis focuses on modality wise results, the impact of multimodal fusion, ablation study insights, and comparisons with baseline models.

A.Overall Performance Analysis

The proposed multimodal framework achieves strong performance across all evaluation metrics. When both textual and visual modalities are utilized, the system attains the highest classification accuracy and F1-score, demonstrating the effectiveness of combining complementary information sources.

The reduction in false positives and false negatives observed in multimodal evaluation highlights the advantage of jointly analyzing textual semantics and visual authenticity cues. These results confirm that multimodal learning provides a more reliable solution than unimodal approaches for fake news detection.

B.Text-Based Detection Results

The text-only model achieves high accuracy due to the use of transformer-based contextual embeddings, which effectively capture semantic relationships within news content. High recall values indicate that the model successfully identifies a large proportion of fake news instances.

However, text-based detection alone is insufficient when misinformation is primarily conveyed through manipulated or AI-generated images. This limitation motivates the integration of visual analysis into the proposed framework.

C.Image-Based Detection Results

The image-only model demonstrates competitive performance but slightly underperforms compared to the text-based model. This behavior is expected due to the increasing realism of AI-generated images and the complexity of detecting subtle visual artifacts.

Despite these challenges, the image-based approach successfully identifies many manipulated or synthetic images, highlighting the importance of incorporating visual cues in misinformation detection systems.

D.Impact of Multimodal Fusion

The highest performance is achieved when predictions from both text and image branches are fused at the decision level. Multimodal fusion improves accuracy and F1-score by lever aging the strengths of each modality while compensating for their individual weaknesses.

In addition to performance gains, decision-level fusion enhances interpretability by providing separate confidence scores for textual and visual predictions. This design enables users to identify whether misinformation originates from text, images or both.

E. Ablation Study Discussion

The ablation study results demonstrate that each component of the proposed architecture contributes to overall system performance. Removing either the text or image modality results in a noticeable drop in accuracy, confirming the necessity of multimodal learning.

Excluding dimensionality reduction increases computational complexity and slightly degrades generalization performance. Similarly, removing the fusion mechanism reduces accuracy, emphasizing its role in effectively integrating heterogeneous feature representations.

F. Baseline Comparison Discussion

Comparative evaluation with baseline models reveals that traditional machine learning approaches based on handcrafted features perform significantly worse than deep learning-based methods. Transformer-based text models outperform classical baselines due to their ability to capture contextual semantics.

The proposed multimodal framework consistently outper forms all baseline models, including strong unimodal deep learning approaches. This improvement highlights the benefit of integrating textual and visual information for comprehensive fake news detection.

G. Practical Implications

The experimental results indicate that the proposed system is suitable for real-world deployment scenarios such as social media monitoring, online news verification, and content moderation platforms. The ability to process text-only, image-only, and multimodal inputs ensures flexibility in diverse application contexts.

Furthermore, the interpretability provided by modality specific predictions enhances user trust and facilitates in formed decision-making.

H. Discussion Summary

The results and discussion confirm that the proposed multi modal fake news and AI-generated image detection framework achieves high accuracy, robustness, and interpretability. By combining transformer-based textual analysis, CNN-based visual analysis, and decision-level fusion, the system effectively addresses challenges posed by modern misinformation and AI-generated media.

6.CONCLUSION AND FUTURE WORKS

In this project, a multimodal fake news detection system was developed using both text and image analysis. The text part uses advanced language models to understand news content, while the image part detects fake or AI-generated pictures using deep learning techniques. By combining results from both modules, the system provides more accurate and reliable predictions.

The experimental results show that using both text and images together performs better than using only one type of data. The model successfully identifies fake news and manipulated images and helps understand where misinformation comes from. This makes the system useful for social media platforms, news verification, and online safety applications.

Overall, the proposed approach improves fake news detection and offers a practical solution to fight misinformation in the digital world.

In the future, the system can be extended by adding video and audio analysis to detect fake content in more formats. Larger and more real-world datasets can be used to improve accuracy and robustness. Advanced fusion techniques can also be explored to better combine text and image information.

The system can be optimized for real-time use on social media platforms. Adding multilingual support will help detect fake news across different languages. These improvements will make the system more powerful and suitable for real-world deployment.

7.REFERENCES

1.Shu, K., Sliva, A., Wang, S., Tang, J., and Liu, H.,

“Fake News Detection on Social Media: A Data Mining Perspective,”ACM

SIGKDD Explorations, 2017.

2.Zhou, X., and Zafarani, R.,

“A Survey of Fake News: Fundamental Theories, Detection Methods, and

Opportunities,”ACM Computing Surveys, 2020.

3.Wang, Y., et al.,

“EANN: Event Adversarial Neural Networks for Multimodal Fake News

Detection,”KDD Conference, 2018.

4.Khattar, D., et al.,

“MVDAM: A Multimodal Variational Autoencoder for Fake News

Detection,”World Wide Web Journal, 2019.

5.Tan, H., et al.,

“Multimodal Fake News Detection Using Fusion of Textual and Visual Features,”IEEE Access, 2020.

6.Dosovitskiy, A., et al.,

“An Image is Worth 16x16 Words: Vision Transformers for Image Recognition,”ICLR, 2021.

7.Rössler, A., et al.,

“FaceForensics++: A Large-Scale Dataset for

“A Survey on Fake News and Rumour Detection Techniques,”Information Sciences Journal, 2019.

Forgery Detection in Human Faces,”ICCV, 2019.

8. Ramesh, A., et al.,

“Zero-Shot Text-to-Image Generation,”ICML, 2021.

9.Guo, B., et al.,

“Rumor Detection with Hierarchical Social Attention Networks,”CIKM, 2018.

10.Bondielli, A., and Marcelloni, F.,