Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 02 | Feb-2026
ISSN (O) 3107-6696

An Advanced Multimodal Artificial Intelligence System for Real-Time Fake Information Detection

Sabthikaroselin.M[1], Praveen.G[2], Roshini. I[3]

[1,2]Student, [3]Assistant Professor

Department of Computer science

Sree Sakthi Engineering College, Tamil nadu, India

Abstract - The rapid proliferation of digital media has democratized information sharing, but it has simultaneously catalyzed the spread of fake news, misinformation, and disinformation. The velocity and volume at which false narratives propagate pose significant threats to public discourse, social stability, and democratic institutions. This paper proposes a novel Multimodal Context-Aware Architecture (MCAA) driven by Artificial Intelligence (AI) to detect fake information in real-time. By integrating advanced Natural Language Processing (NLP) techniques for textual analysis with Computer Vision for media verification, the proposed system evaluates the veracity of news articles and social media posts. We employ a hybrid deep learning model combining a pre-trained Transformer network (RoBERTa) for text and a Convolutional Neural Network (ResNet) for visual feature extraction, alongside a metadata analyzer. Experimental evaluations on benchmark datasets (such as FakeNewsNet and LIAR) demonstrate that our multimodal approach achieves an accuracy of 94.2%, outperforming unimodal baselines. This paper details the system architecture, methodology, implementation, and future directions for robust misinformation detection.

Keywords: Fake News Detection, Artificial Intelligence, Natural Language Processing, Deep Learning, Multimodal Neural Networks, Misinformation.

1. Introduction

In the contemporary digital era, information is generated and consumed at an unprecedented scale. Social media platforms, news aggregators, and blogs have become primary sources of news for a global audience. However, the lack of rigorous editorial oversight on these platforms has created a fertile ground for "fake news"—intentionally fabricated information designed to deceive, manipulate, or generate advertising revenue.

The consequences of unmitigated fake information are severe, ranging from swaying political elections to inducing public panic during global health crises. Manual fact-checking by human experts is highly accurate but fundamentally unscalable given the sheer volume of daily digital content. Consequently, there is an urgent need for automated, intelligent systems capable of identifying and flagging malicious content in real-time.

This paper presents an AI-driven framework that moves beyond traditional text-only analysis. Because modern misinformation often pairs misleading text with manipulated or out-of-context images, we propose a multimodal AI system.

The primary contributions of this paper are:

1. The development of a Multimodal Context-Aware Architecture (MCAA) that simultaneously analyzes text, embedded media, and user-context metadata.
2. A comprehensive evaluation of the proposed model against standard benchmark datasets, demonstrating superior performance over existing unimodal systems.
3. An analysis of feature importance to understand how different data modalities contribute to the detection of deceptive content.

2. Literature Review

The detection of fake information has evolved significantly from early rule-based systems to complex machine learning frameworks.

- **Linguistic and Text-Based Approaches:** Early AI systems relied heavily on traditional Natural Language Processing (NLP). Researchers utilized Support Vector Machines (SVM) and Naive Bayes classifiers trained on term frequency-inverse document frequency (TF-IDF) features. These models detected stylistic cues like excessive use of emotional language or hyperbole. However, they struggled with context and nuanced linguistic manipulation.
- **Deep Learning and NLP:** The introduction of Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) networks, allowed systems to understand sequential

dependencies in text. More recently, Transformer-based architectures like BERT (Bidirectional Encoder Representations from Transformers) have established new state-of-the-art benchmarks by capturing deep bidirectional contexts.

- **Multimodal Detection:** Recent literature emphasizes the necessity of multimodal analysis. Misinformation often relies on the mismatch between text and images (e.g., a real photo from 2015 used to describe an event in 2026). Models that fuse text and visual features have shown marked improvements in detecting sophisticated fake news campaigns.

3. Proposed Methodology

The proposed Multimodal Context-Aware Architecture (MCAA) is designed to process an input data object (e.g., a news article or a social media post) containing text, images, and metadata (source credibility, timestamp, author engagement).

The architecture consists of three primary extraction pipelines and a centralized fusion module:

3.1 Textual Feature Extraction

For textual analysis, the system utilizes **RoBERTa (Robustly Optimized BERT Pretraining Approach)**. RoBERTa dynamically masks tokens during training, allowing it to better understand the contextual relationships in the article's headline and body.

- **Input:** Tokenized textual content.
- **Output:** A high-dimensional dense vector representing the semantic meaning, emotional tone, and syntactic structure of the text.

3.2 Visual Feature Extraction

Visual content is processed using **ResNet-50** (Residual Networks). The model analyzes the accompanying images to extract deep visual features. Furthermore, an Error Level Analysis (ELA) filter is applied prior to the CNN to highlight regions of the image that may have been digitally altered.

- **Input:** Standardized image matrices.
- **Output:** Feature vectors representing image semantics and potential manipulation artifacts.

3.3 Metadata and Context Analysis

Information such as the domain age, author's historical reliability, and engagement metrics (likes, shares, comment velocity) are processed through a structured Multi-Layer Perceptron (MLP). This provides the system with context regarding *how* the information is spreading.

3.4 Multimodal Fusion and Classification

The feature vectors from the textual, visual, and metadata pipelines are concatenated into a unified representation space. A dense feed-forward neural network with a Softmax output layer is then employed to calculate the probability of the input belonging to one of two classes: **Real** or **Fake**.

4. Implementation and Evaluation Setup

To rigorously test the proposed MCAA system, we conducted experiments using established academic datasets.

4.1 Datasets

- **FakeNewsNet:** A comprehensive dataset containing news content, social context, and spatiotemporal information derived from PolitiFact and GossipCop.
- **LIAR Dataset:** A benchmark dataset for fake news detection consisting of 12.8K human-labeled short statements in various contexts.

4.2 Data Preprocessing

- **Text:** Lowercasing, removal of special characters, and subword tokenization using Byte-Pair Encoding (BPE).
- **Images:** Resized to 224x224 pixels and normalized to match the ImageNet pre-training distributions.
- **Balancing:** Synthetic Minority Over-sampling Technique (SMOTE) was used to address class imbalances in the datasets to prevent model bias.

4.3 Evaluation Metrics

The performance of the system is evaluated using standard machine learning metrics:

- **Accuracy:** Overall correctness of the model.
- **Precision:** The ratio of correctly predicted positive observations to the total predicted positives.
- **Recall:** The ratio of correctly predicted positive observations to the all observations in actual class.
- **F1-Score:** The weighted average of Precision and Recall.

5. Results and Discussion

The proposed MCAA model was compared against several baseline models: a Text-Only Support Vector Machine (SVM), a Text-Only BERT model, and an Image-Only CNN.

5.1 Performance Comparison

Model Architecture	Modality	Accuracy	Precision	Recall	F1-Score
Baseline SVM	Text (TF-IDF)	78.4%	77.1%	76.5%	76.8%
Fine-tuned BERT	Text Only	88.6%	88.2%	89.1%	88.6%
ResNet-50	Image Only	65.2%	63.4%	61.0%	62.1%
Proposed MCAA	**Text + Image + Meta**	**94.2%**	**93.8%**	**94.5%**	**94.1%**

5.2 Discussion

The results clearly indicate the superiority of the multimodal approach. While text-only deep learning models (BERT) perform admirably, they fail to identify deceptive posts where the text is factually neutral but the accompanying image is manipulated or historically inaccurate. The Image-Only model performed poorly on its own, which is expected as images without context rarely convey "fake news" natively.

By fusing the modalities, the MCAA system successfully correlates discrepancies between text and imagery. For instance, the system effectively flagged instances where an article discussed a current political protest, but the visual feature extraction recognized the image as belonging to a different geographic location from years prior.

6. Conclusion and Future Work

This paper presented an advanced Artificial Intelligence system for the detection of fake information, utilizing a Multimodal Context-Aware Architecture. By integrating RoBERTa for deep text understanding, ResNet for visual verification, and an MLP for metadata context, the system achieved a highly robust accuracy of 94.2% on standard benchmarks. The results validate the hypothesis that a holistic, multi-faceted approach to data evaluation is strictly necessary to combat modern, sophisticated disinformation campaigns.

Future Work:

Future research will focus on improving the computational efficiency of the MCAA system to allow for lightweight deployment directly on edge devices and browser extensions. Additionally, extending the model to evaluate video-based misinformation, particularly the rising threat of "Deepfakes," represents a critical next step in safeguarding digital information ecosystems.

References

1. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). "Fake News Detection on Social Media: A Data Mining Perspective." *ACM SIGKDD Explorations Newsletter*, 19(1), 22-36.
2. Wang, W. Y. (2017). "Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (Vol. 2: Short Papers), 422-426.
3. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *arXiv preprint arXiv:1810.04805*.
4. Khattar, D., Goud, J. S., Gupta, M., & Varma, V. (2019). "MVAE: Multimodal Variational Autoencoder for Fake News Detection." *The World Wide Web Conference*, 2915-2921.
5. He, K., Zhang, X., Ren, S., & Sun, J. (2016). "Deep Residual Learning for Image Recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778.