A Deep Learning Multimodal Framework for Fake News Detection

Shweta Kumari
Department of Computer Science and Engineering, National Institute of Technology Patna, Patna, Bihar, India
shwetak.phd18.cs@nitp.ac.in (corresponding author)

Maheshwari Prasad Singh
Department of Computer Science and Engineering, National Institute of Technology Patna, Patna, Bihar, India
mps@nitp.ac.in

Received: 20 June 2024 | Revised: 8 July 2024 | Accepted: 11 July 2024

 DOI: https://doi.org/10.48084/etasr.8170

ABSTRACT

The swift spread of fake news on social media platforms presents significant challenges to the society, necessitating the development of a more efficient model for fake news detection. Existing fake news detection methods primarily focus on linguistic and compositional characteristics, which may not be able to differentiate between various forms of fake news and impede effective detection. This paper proposes an innovative approach to address such challenges. It introduces a comprehensive framework for detecting fake news, leveraging advanced multimodal techniques to analyze multilingual text and visual data. The proposed framework employs Natural Language Processing (NLP) for text preprocessing, the DeepL translator for language consistency, and vectorization for feature extraction. For the detection models, Long Short-Term Memory (LSTM) networks are used for sequential text analysis, while the Contrastive Language-Image Pretraining (CLIP) model is utilized for image analysis to be performed. The combined features are then processed through a decision-making layer for the news to be classified as real or fake. Results demonstrate the model's high efficacy, with an accuracy of 99.22% for text and 93.12% for text and images, outperforming the existing state-of-the-art techniques.

Keywords-sentiment analysis; machine learning; fake news; multilingual; multimodal

I. INTRODUCTION

The internet has revolutionized communication, with social media platforms like Facebook, Instagram, and X (Twitter) enabling real-time information sharing and news updates [1]. Over 50% of people use social media for news [2], but the lack of verification systems has made it easy to spread false information. Social media's rapid information diffusion contributes to the quick spread of fake news [3], impacting user perceptions and society. This has been evident in events like the 2016 US presidential election and the selection of a new Air Marshal in India [4].

Fake news can significantly affect mental health and societal stability, especially during crises like that of COVID-19 [5-6]. The ease of generating realistic fake news with tools like ChatGPT further complicates distinguishing it from genuine journalism [7]. Fake news can be in various forms, as shown in Figure 1. The proposed model effectively distinguishes between real and fake news.

This framework involves data collection and preprocessing with advanced NLP methods and the DeepL translator for language consistency. The framework performs feature extraction using vectorization for textual data and the Contrastive Language-Image Pretraining (CLIP) model for image analysis, optimizing performance with an LSTM network for sequential text analysis. The outputs from the LSTM and CLIP models are integrated into a cohesive multimodal analysis framework. The main contributions of this study are:

- To develop a multilingual-multimodal fake news detection framework integrating NLP, deep learning, and CLIP for image-text embedding.
- To evaluate multilingual fake news detection systems using diverse datasets, assessing metrics, including accuracy, precision, recall, and F1-score for comprehensive performance analysis.
- To investigate how linguistic diversity and cultural context affect fake news detection efficacy by analyzing the impact of language structure, and semantics on detection accuracy and false positives.

Fig. 1. Fake news may encompass a variety of forms.

II. LITERATURE SURVEY

Social media platforms like Facebook, Instagram, YouTube, and X (Twitter) offer powerful venues for news and entertainment, but they also pose risks due to the ease of disseminating fake information. Authors in [8] proposed a method to identify fake news by examining its numerous characteristics.

Authors in [9] presented a sparse and graph-regularized CANDECOMP/PARAFAC (SGCP) optimization approach, which demonstrated to be effective through experiments on real-world datasets. Authors in [10] introduced the Indian Fake News Dataset (IFND), highlighting its contribution to fostering research in fake news detection with improved prediction models. Authors in [11] presented the Social Engagement-based News Authenticity Detection SENAD model incorporating an authenticity score and user engagement metrics, achieving 76% accuracy. Authors in [12] proposed NLP techniques for governments to improve policy analysis, regulatory compliance, and feedback analysis, benefiting societal governance.

Authors in [13] introduced a linguistic inquiry and word count method using psycholinguistic approaches for news categorization, while authors in [14] conducted sentiment analysis on Amazon Fine Food Reviews using LSTM, ALBERT, and RF classifiers, all achieving 96% accuracy. Authors in [15] proposed a machine learning approach considering user social capital, news content, and social networks, using XGBoost to determine feature importance. The model achieved up to 94% accuracy with RF classifiers, contrasting with Neural Networks (NNET's) 92.1% performance.

Authors in [16] employed a decentralized spark cluster to develop a stacked ensemble model incorporating N-grams, Hashing, TF-IDF, and count vectorizer for feature extraction, achieving 92.45% accuracy. Authors in [17] proposed a multiscale transformer model for detecting fake news across mixed languages, achieving 86.86% accuracy. Authors in [18] addressed fake news detection in Turkish and English, achieving accuracy rates from 87.14% to 92.48%, using language-specific algorithms. Authors in [19] developed a multimodal fake news detection framework, achieving 91.94% accuracy by leveraging diverse data types. Authors in [20] introduced the Clip-GCN model for emergent news detection, combining text and image semantic features to achieve 88.7% accuracy on Chinese and English social media datasets.

TABLE I COMPARISON CHART OF LITERATURE REVIEW

Reference (published year)	Dataset	Model	Accuracy
[9] (2023)	Monolingual Text	RM, SVM, DT, K-NN, LDA	81.38%
[10] (2023)	Monolingual Text	IFND	94%
[11] (2022)	Monolingual Text, images	SENAD, CNN	76%
[15] (2023)	Monolingual Text	XGBoost, CRT, RF, NN, SVM, LR	94%
[16] (2023)	Monolingual Text	EC, DT, RF, LR	92.45%
[17] (2023)	Bilingual Text	CNN, LSTM, BERT	86.86%
[18] (2023)	Bilingual Text	CNN, RNN-LSTM	92.48%
[19] (2022)	Monolingual Text, image	TL, LSTM, CNN	91.94%
[20] (2024)	Bilingual Text, images	CLIP, GCN	88.7%

The literature highlights a gap in fake news detection methods evaluated predominantly in English texts, emphasizing the scarcity of evaluations across linguistic boundaries. Table I provides a comparison of existing studies, stressing the need for robust multilingual systems capable of accurately identifying and mitigating fake news. Multilingualism and multimodal approaches pose challenges such as linguistic nuances and cultural differences, necessitating the development of effective detection systems capable of addressing diverse linguistic landscapes.

III. PROPOSED METHODOLOGY

This section outlines a comprehensive framework for detecting fake news, encompassing data analysis, feature extraction, multilingual conversion to English, image analysis, and final detection. Figure. 2 illustrates the proposed framework, integrating advanced techniques across three main parts: data preprocessing, deep learning plus CLIP, and decision-making processes. The proposed framework for detecting fake news begins with data collection, followed by preprocessing using NLTK for tokenization and stemming to normalize text data. DeepL is utilized to translate the text from eleven languages to English, ensuring linguistic uniformity. Text feature extraction involves text vectorization, while visual features are extracted using the CLIP model, generating multimodal representations for integrated analysis. These transforms processed data into numerical vectors, facilitating machine learning model analysis and learning. The proposed model will work for two different datasets, namely:

- Only text data
- Mixed data (text and image)

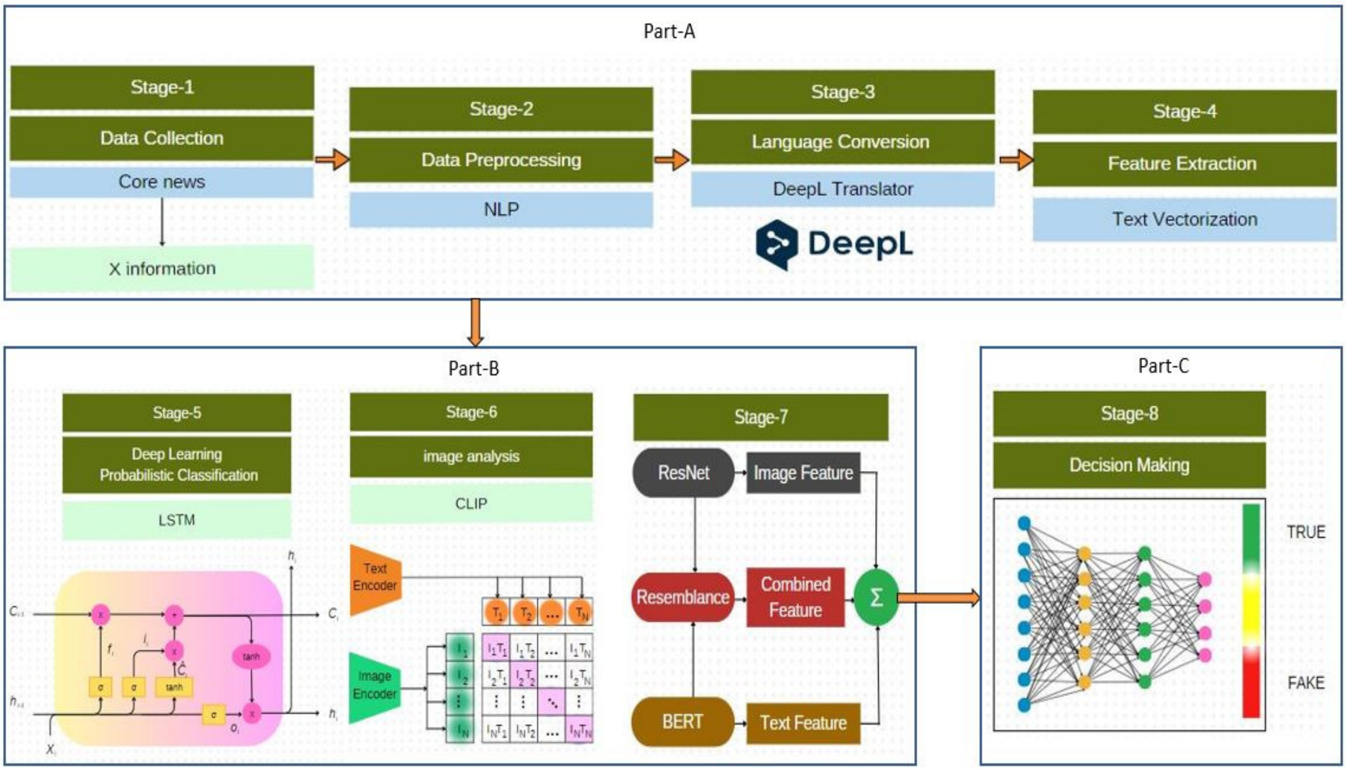

Fig. 2. Proposed model for fake news detection.

The proposed framework addresses multilingual fake news detection across eleven languages including English, German, French, Finnish, Portuguese, Swahili, Vietnamese, Somali, Spanish, Polish, and Romanian. It acknowledges the complexities posed by linguistic diversity and cultural nuances, which traditionally challenge detection methods. Linguistic variability [21] is the diverse linguistic landscape across different languages, which introduces variations in vocabulary, syntax, and semantics, making it challenging to generalize fake news detection models across multiple languages. This can be represented mathematically as:

$$Variance(L) = \sum_{i=1}^{n}(x_i - \mu)^2 \quad (1)$$

where L represents the linguistic landscape or the set of languages under consideration, n is the number of different languages, x_i is a linguistic feature or characteristic (e.g. vocabulary, syntax, semantics) of the i^{th} language, and μ is the mean value of the linguistic feature across all languages. The variance formula encapsulates this variability, highlighting its impact on model generalization. Deep Learning Probabilistic Classification [22] and LSTM networks are used to process sequential data and capture temporal dependencies and context from the pre-processed textual features. For each time stamp t:

$$h_t c_t \leftarrow LSTM(X_t, h_{t-1}, c_{t-1}) \quad (2)$$

where h_t represents hidden state of textual feature at the time stamp t, c_t signifies the cell state at time stamp t. X_t represents the input feature vector derived from the current word at time stamp t. h_{t-1} signifies the hidden state for time stamp $t-1$ and c_{t-1} represents cell state for time stamp $t-1$.

The cell state is a kind of long-term memory that can carry information across many time stamps, allowing the, LSTM to maintain a context over long sequences. After processing the input X_t along with the previous hidden state h_{t-1} and the previous cell state, c_{t-1}, the LSTM outputs this new hidden state. It captures the current context and can be used for predictions or passed to the next time stamp, c_t is the updated cell state at the current time stamp t. The LSTM updates the cell state based on the input X_t, the previous hidden state h_{t-1}, and the previous cell state c_{t-1}. This updated cell state will be used in the next time stamp $t+1$. This process allows the LSTM to maintain and update a representation of the sequence's context as it processes each element in the sequence, making it well-suited for tasks involving sequential data like text, time series, and more.

For multimodal fake news detection, the proposed framework uses the dataset which includes text and image, and each sample is denoted as $x = (x_T, x_I)$. The label indicating the truth is y where $y = 0$ specifies that x is real news. Otherwise $y = 1$ specifies that x is fake news. Plenty of features representing the unimodal properties are retrieved from x_T, and x_I, they are then projected in a solitary value of $\hat{y}$, which is the predicted probability [23], and is expected to be fairly close to the ground truth, after being further fused.

$$\hat{y} = F_c\left(F_M\left(F_T(x_T), F_I(x_I)\right)\right) \quad (3)$$

where F_M is the feature blending model, F_c is the head of classification, and F_T and F_I are unimodal feature extractors for text and image respectively. The majority of the previous approaches utilize a variety of trained models to extract features from images and text in various semantic domains in order to simulate F_T and F_I, while other techniques are presented for F_M.

The proposed solution finds cross-modal similarity [24] by using the CLIP taking into account that the model satisfies the aforementioned characteristics since it is taught to offer the best linguistic description of a particular image and vice versa. The framework uses the pretrained BERT model to obtain the feature $f_{BERT} \in R^{n_{BERT}}$ of text x_T and $ResNet$ [25] to get profound depictions $f_{ResNet} \in R^{n_{ResNet}}$ of image x_I. CLIP uses f_{BERT} and f_{ResNet}, to encode text and image respectively. $f_{CLIP-T} \in R^{n_{CLIP}}$ and $f_{CLIP-I} \in R^{n_{CLIP}}$is applied to attain the features. To enhance the unimodal depiction, embedding sequencing is executed in the unimodal intra-modalities of text and image, as given below:

$$f_T = concat(f_{BERT}, f_{CLIP-T}) \quad (4)$$

$$f_I = concat(f_{ResNet}, f_{CLIP-I}) \quad (5)$$

where $f_T \in R^{n_{BERT}+n_{CLIP}}$and $f_I \in R^{n_{ResNet}+n_{CLIP}}$. It is challenging for networks to acquire their basic semantic association when they are directly fused as BERT and RestNet, extracting large cross-modal semantic gaps. Consequently, the dual features are limited to unimodal depiction, whereas the text-image pair's alignment features recovered through CLIP are first concatenated to create the multimodal representation, which is subsequently adjusted to eliminate redundancy. The combined feature is represented as $f_{Comb} \in R^{2\times n_{CLIP}}$, where

$$f_{Comb} = concat(f_{CLIP-T}, f_{CLIP-I}) \quad (6)$$

The introduction of the CLIP model enhances pretrained unimodal models like BERT and ResNet by focusing on extracting semantic information from large-scale image-text pairs. This allows ResNet to recognize noise patterns in photos and BERT to extract emotional elements from text. CLIP's training method disregards irrelevant noise and emotion, improving multimodal feature creation. The framework employs three projection heads P_T, P_I, and P_{Comb} with Multi-Layer Perceptrons to process features from different modalities, reducing coarse features and eliminating unnecessary data. This approach addresses the potential noisiness of multimodal features due to weak image-text correlations in some news articles, ensuring a more accurate analysis.

To address this, the cosine similarity between text and image features provided by CLIP is calculated, filtering out ambiguous multimodal features to enhance performance. The cosine similarity [26] is determined as follows:

$$sim = \frac{f_T \cdot (f_I)^T}{\|f_T\| \, \|f_I\|} \quad (7)$$

Now similarity is mapped into a range $[0 - 1]$ using Sigmoid function. The projected unimodal features for text, unimodal features for image, and multimodal features for text and image are calculated using (8), (9), and (10) [27], respectively.

$$M_T = P_T(f_T) \quad (8)$$

$$M_I = P_I(f_I) \quad (9)$$

$$M_{Comb} = Sigmoid\big(Std(sim)\big) \cdot P_{Comb}(f_{Comb}) \quad (10)$$

Then, the aggregated feature M_{Agg} [27] is obtained as per (11):

$$M_{Agg} = M_T + M_I + M_{Comb} \quad (11)$$

Finally, f_T, f_I, f_{Comb}, and M_{Agg} is imparted to a four-layer fully-connected network as F_T, F_I, F_M, and the classifier F_c, respectively, to predict the label $\hat{y}$. Now binary cross-entropy [28] also known as loss function is calculated as per (12):

$$\mathcal{L}_{CE} = y\log(\hat{y}) + (1-y)\log(1-\hat{y}) \quad (12)$$

where $\mathcal{L}_{CE}$ is the loss function, y is the true label which can be either 0 or 1, and $\hat{y}$ is the predicted probability that is true label.

A. Components of the Loss Function

First Term: $y\log(\hat{y})$: This term only contributes when $y = 1$. It measures the log-probability of the predicted probability when the actual label is 1. If the predicted probability $\hat{y}$ is close to 1, $\log(\hat{y})$ is close to 0, leading to a lower loss. If the predicted probability $\hat{y}$ is close to 0, $\log(\hat{y})$ becomes a large negative number, leading to a higher loss.

Second Term: $(1-y)\log(1-\hat{y})$: This term only contributes when $y = 0$. It measures the log-probability of the predicted probability when the actual label is 0. If the predicted probability $\hat{y}$ is close to 0, $\log(1-\hat{y})$ is close to 0, leading to a lower loss. If the predicted probability $\hat{y}$ is close to 1, $\log(1-\hat{y})$ becomes a large negative number, leading to a higher loss.

Binary cross-entropy is calculated because it effectively handles the probabilistic nature of predictions and ensures that the model is penalized appropriately for incorrect predictions, guiding the model's parameters towards better accuracy over iterations during training. The objective is to minimize the cross-entropy loss to predict the real and fake news correctly.

The proposed deep learning framework is implemented and tested on a Windows 7 operating system, using a multi-feature method to identify fake news in multilingual multimodal social networks. This system configuration provided the computational power for several sentiment analysis and multimodal data processing tasks. The experimental setup was reliable due to the 16 GB RAM capacity, which enabled efficient handling of large-scale datasets and the execution of advanced machine-learning algorithms.

IV. RESULTS AND DISCUSSION

The study's results are presented and explained in this section. The latter includes numerical data, statistical analysis, graphs, tables, and other relevant representations.

A. Dataset Description

Two publicly available datasets are used. One is "Fake News Detection" dataset from Kaggle, with 44,898 text entries sourced from Facebook, including article titles, full text, and veracity labels, and it may also contain metadata like author, publication date, and source. This dataset is intended for developing machine learning models for fake news detection using natural language processing to analyze word patterns, context, and semantics. The other is "Fakeddit" dataset, with

1,063,106 entries from Reddit, featuring text posts, images, and combinations thereof, annotated with veracity labels for multimodal detection model development. It supports advanced NLP techniques with diverse features such as post titles, body text, and associated images, facilitating robust misinformation detection across various content types on social media.

B. Sentiment Score

Sentiment scores are categorized into Real, and Fake. The overall combined emotion score for the Fake News Detection dataset and Fakeddit dataset is shown in Figure. 3.

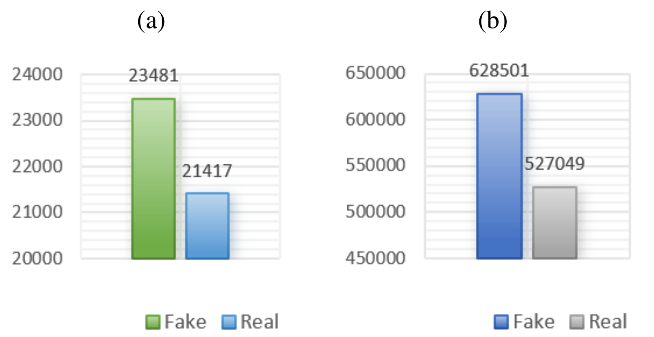

Fig. 3. Overall combined emotion score for (a) Fake News Detection dataset and (b) Fakeddit dataset.

C. Performance Evaluation

The performance of the proposed model has been measured using loss, accuracy, precision, recall, and F1-score [1]. Loss is a scale to measure how greatly or poorly the predictions of a model match the actual outcomes. Accuracy is defined as the ratio of the correctly predicted cases-both true positives and true negatives-to all instances, precision is the proportion of true positive predictions to the entire number of positive predictions, the ratio of true positive predictions to the entire number of actual positives is recall, and F1-score is the harmonic mean of precision and recall, supplying a single metric that balances both.

$$\text{Accuracy} = \frac{\text{True Positives+True Negatives}}{\text{Total Number of Instances}} \quad (13)$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives+False Positives}} \quad (14)$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives+False Negatives}} \quad (15)$$

$$\text{F1-score} = 2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \quad (16)$$

The performance of the fake news detection model is evaluated over various epochs, showing substantial accuracy improvement and loss reduction during training. Table II presents the loss and accuracy metrics for Fake News Detection, and Fakeddit datasets over epochs 1 to 35.

The summarized performance of the three models trained on different datasets for fake news detection reveals notable insights into their learning dynamics over 35 epochs. The Fake News Detection (FND) model starts at epoch 1 with a loss of 54.59% and an accuracy of approximately 78.34%. As epochs increase the loss drastically reduces and accuracy surges. Ultimately achieving a loss of 6.24% and an accuracy of 99.22% at epoch 35, indicates highly effective feature learning and minimal errors The model shows the fastest improvement in the Fake News Detection dataset, achieving high accuracy and low loss by 35 epochs, indicating the dataset's suitability or the model's effectiveness for this task.

TABLE II THE LOSS AND ACCURACY VALUES AT DIFFERENT EPOCHS

Dataset	Epoch	Loss	Accuracy
Fake News Detection (Text)	1	0.545883238	0.783434092
	20	0.096545405	0.973273932
	35	0.062357760	0.992151856
Fakeddit (Text)	1	0.842953742	0.597167969
	20	0.476794219	0.741308594
	35	0.214338526	0.91796875
Fakeddit (Text, image)	1	0.831954653	0.588267879
	20	0.294515759	0.868783931
	35	0.178612368	0.931298734

In contrast, the Fakeddit(Text) model begins with a higher loss of 84.30% and a lower accuracy of 59.72%, reflecting initial challenges in processing textual data. It improves with an increase in epochs, and further with a loss of 21.43% and accuracy of 91.80% by epoch 35, it shows steady but slower progress. The Fakeddit(Text and Image) model starts with a loss of 83.20% and an accuracy of 58.83% at epoch 1. As epochs progress, the loss decreases and accuracy improves, reaching 17.86% loss and 93.12% accuracy by epoch 35, demonstrating enhanced performance with multimodal data and highlighting dataset complexities.

The model starts with high loss but shows significant improvement by epoch 35, benefiting from multimodal data. It performs well on the Fakeddit multimodal dataset, indicating effective learning and faster improvement with the Fake News Detection dataset. The Fakeddit dataset is challenging, especially with text-only data, but incorporating images significantly improves performance, highlighting the benefits of multimodal learning approaches. Table III shows the test data classification report.

TABLE III CLASSIFICATION REPORT OF THE PROPOSED MODAL

Dataset	Dataset type	Precision	Recall	F1-score	Accuracy
Fake News Detection	Multilingual text	0.99	0.99	0.99	0.99
Fakeddit	Monolingual text	0.90	0.91	0.91	0.91
Fakeddit	Image and monolingual text	0.93	0.95	0.93	0.93

D. Comparison with State-of-the-Art

The proposed method, utilizing LSTM and CLIP on text in eleven languages and images, demonstrates superior performance compared to state-of-the-art techniques in fake news detection. For text-only approaches, the proposed method achieves an accuracy of 99.22% when state of the art works achieve 81.38% [9] and 94% [10, 15]. In text-image modalities, the proposed method reaches 93.12% accuracy, surpassing the models referred in [19] that score 91.94% accuracy, and the model described in [20] that scores 88.7% accuracy.

Table IV demonstrates the proposed method's effectiveness in leveraging multilingual-multimodal data, setting a new benchmark in fake news detection accuracy. This approach's versatility in handling diverse data types surpasses methods that typically focus on text or single-language datasets.

TABLE IV COMPARISON WITH STATE-OF-THE-ART

Authors (published year)	Dataset	Model	Accuracy
[9] (2023)	Monolingual Text	RM, SVM, DT, K-NN, LDA	81.38%
[10] (2023)	Monolingual Text	IFND	94%
[11] (2022)	Monolingual Text, images	SENAD, CNN	76%
[15] (2023)	Monolingual Text	XGBoost, CRT, RF, NN, SVM, LR	94%
[16] (2023)	Monolingual Text	EC, DT, RF, LR	92.45%
[17] (2023)	Bilingual Text	CNN, LSTM, BERT	86.86%
[18] (2023)	Bilingual Text	CNN, RNN-LSTM	92.48%
[19] (2022)	Monolingual Text, image	TL, LSTM, CNN	91.94%
[20] (2024)	Bilingual Text, images	CLIP, GCN	88.7%
Proposed model	Text in 11 languages, images	LSTM, CLIP	Text: 99.22% Text-image: 93.12%

V. CONCLUSION AND FUTURE SCOPE

The proposed method for fake news detection, utilizing Long Short-Term Memory (LSTM) and Contrastive Language-Image Pretraining (CLIP) on text in eleven languages and images, demonstrates remarkable performance, achieving an accuracy of 99.22% for text-only data and 93.12% for text-image data. The model's ability to handle multilingual-multimodal data showcases its versatility, making it a powerful tool for combating misinformation. Unlike previous studies that focused on single-language text or specific data types, this method offers a more comprehensive solution for fake news detection. Future research should enhance the model's robustness and efficiency by testing it in different domains, improving scalability, and enabling real-time detection to swiftly and accurately identify misinformation.

REFERENCES

[1] S. Vermeer, D. Trilling, S. Kruikemeier, and C. de Vreese, "Online News User Journeys: The Role of Social Media, News Websites, and Topics," *Digital Journalism*, vol. 8, no. 9, pp. 1114–1141, Oct. 2020, https://doi.org/10.1080/21670811.2020.1767509.

[2] P. Jiao, A. Veiga, and A. Walther, "Social media, news media and the stock market," Journal of Economic Behavior & Organization, vol. 176, pp. 63–90, Aug. 2020, https://doi.org/10.1016/j.jebo.2020.03.002.

[3] S. Sahoo Ranjan and B. B. Gupta, "Multiple features based approach for automatic fake news detection on social networks using deep learning," Applied Soft Computing, vol. 100, Mar. 2021, Art. no. 106983, https://doi.org/10.1016/j.asoc.2020.106983.

[4] J. Reis C. S., A. Correia, F. Murai, A. Veloso, and F. Benevenuto, "Supervised Learning for Fake News Detection," IEEE Intelligent Systems, vol. 34, no. 2, pp. 76–81, Mar. 2019, https://doi.org/10.1109/MIS.2019.2899143.

[5] M. Asghar Zubair, A. Habib, A. Habib, A. Khan, R. Ali, and A. Khattak, "Exploring deep neural networks for rumor detection," Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 4, pp. 4315–4333, Apr. 2021, https://doi.org/10.1007/s12652-019-01527-4.

[6] R. Kaliyar Kumar, A. Goswami, and P. Narang, "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach," Multimedia Tools and Applications, vol. 80, no. 8, pp. 11765–11788, Mar. 2021, https://doi.org/10.1007/s11042-020-10183-2.

[7] K. Miller, "Human Writer or AI? Scholars Build a Detection Tool," Feb. 13, 2023. https://hai.stanford.edu/news/human-writer-or-ai-scholars-build-detection-tool.

[8] V. Singh K., I. Ghosh, and D. Sonagara, "Detecting fake news stories via multimodal analysis," Journal of the Association for Information Science and Technology, vol. 72, no. 1, pp. 3–17, May 2021, https://doi.org/10.1002/asi.24359.

[9] H. Che, B. Pan, M.-F. Leung, Y. Cao, and Z. Yan, "Tensor Factorization With Sparse and Graph Regularization for Fake News Detection on Social Networks," IEEE Transactions on Computational Social Systems, pp. 1–11, 2023, https://doi.org/10.1109/TCSS.2023.3296479.

[10] D. Sharma Kumar and S. Garg, "IFND: a benchmark dataset for fake news detection," Complex & Intelligent Systems, vol. 9, no. 3, pp. 2843–2863, Jun. 2023, https://doi.org/10.1007/s40747-021-00552-1.

[11] S. Uppada Kumar, K. Manasa, B. Vidhathri, R. Harini, and B. Sivaselvan, "Novel approaches to fake news and fake account detection in OSNs: user social engagement and visual content centric model," Social Network Analysis and Mining, vol. 12, no. 1, May 2022, Art. no. 52, https://doi.org/10.1007/s13278-022-00878-9.

[12] Y. Kang, Z. Cai, C.-W. Tan, Q. Huang, and H. Liu, "Natural language processing (NLP) in management research: A literature review," Journal of Management Analytics, vol. 7, no. 12, pp. 1–34, May 2020, https://doi.org/10.1080/23270012.2020.1756939.

[13] J. Pennebaker, L. Boyd Ryan, N. Jordan Kayla, and K. Blackburn, "The Development and Psychometric Properties of LIWC2015," Sep. 2015, https://doi.org/10.15781/T29G6Z.

[14] N. A. Alabdulkarim, and M. A. Haq, "Exploring Sentiment Analysis on Social Media Texts" Engineering, Technology & Applied Science Research, Vol. 14, No. 3, pp. 14442-14450, June, 2024. DOI: https://doi.org/10.48084/etasr.7238

[15] M. Park and S. Chai, "Constructing a User-Centered Fake News Detection Model by Using Classification Algorithms in Machine Learning Techniques," IEEE Access, vol. 11, pp. 71517–71527, 2023, https://doi.org/10.1109/ACCESS.2023.3294613.

[16] A. Altheneyan and A. Alhadlaq, "Big Data ML-Based Fake News Detection Using Distributed Learning," IEEE Access, vol. 11, pp. 29447–29463, 2023, https://doi.org/10.1109/ACCESS.2023.3260763.

[17] Z. Guo, Q. Zhang, F. Ding, X. Zhu, and K. Yu, "A Novel Fake News Detection Model for Context of Mixed Languages Through Multiscale Transformer," IEEE Transactions on Computational Social Systems, pp. 1–11, 2023, https://doi.org/10.1109/TCSS.2023.3298480.

[18] G. Güler and S. Gündüz, "Deep Learning Based Fake News Detection on Social Media," International Journal of Information Security Science, vol. 12, no. 2, pp. 1–21, Jun. 2023, https://doi.org/10.55859/ijiss.1231423.

[19] S. Uppada Kumar, P. Patel, and B. Sivaselvan, "An image and text-based multimodal model for detecting fake news in OSN's," Journal of Intelligent Information Systems, vol. 61, no. 2, pp. 367–393, Oct. 2023, https://doi.org/10.1007/s10844-022-00764-y.

[20] Y. Zhou, A. Pang, and G. Yu, "Clip-GCN: an adaptive detection model for multimodal emergent fake news domains," Complex & Intelligent Systems, vol. 10, no. 4, pp. 5153–5170, Aug. 2024, https://doi.org/10.1007/s40747-024-01413-3.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2016, pp. 770–778, https://doi.org/10.1109/CVPR.2016.90.

[22] F. Laghrissi Ezzahra, S. Douzi, K. Douzi, and B. Hssina, "Intrusion detection systems using long short-term memory (LSTM)," Journal of Big Data, vol. 8, no. 1, May 2021, Art. no. 65, https://doi.org/10.1186/s40537-021-00448-4.

[23] Y. Bizzoni, S. Degaetano-Ortlieb, P. Fankhauser, and E. Teich, "Linguistic Variation and Change in 250 Years of English Scientific Writing: A Data-Driven Approach," Frontiers in Artificial Intelligence, vol. 3, Sep. 2020, https://doi.org/10.3389/frai.2020.00073.

[24] M. Li et al., "CLIP-Event: Connecting Text and Images with Event Structures." arXiv, Apr. 29, 2022, https://doi.org/10.48550/arXiv.2201.05078.

[25] W. Wolfram, "Variation and Language: Overview," Encyclopedia of Language & Linguistics, Dec. 2006, https://doi.org/10.1016/B0-08-044854-2/04256-5.

[26] T. Silva, Filho, H. Song, M. Perello, Nieto, R. Santos, Rodriguez, M. Kull, and P. Flach, "Classifier calibration: a survey on how to assess and improve predicted class probabilities," Machine Learning, vol. 112, no. 9, pp. 3211–3260, Sep. 2023, https://doi.org/10.1007/s10994-023-06336-7.

[27] A. Ghorbanali and M. Sohrabi Karim, "A comprehensive survey on deep learning-based approaches for multimodal sentiment analysis," Artificial Intelligence Review, vol. 56, no. 1, pp. 1479–1512, Oct. 2023, https://doi.org/10.1007/s10462-023-10555-8.

[28] P. Hurtik, S. Tomasiello, J. Hula, and D. Hynar, "Binary cross-entropy with dynamical clipping," Neural Computing and Applications, vol. 34, no. 14, pp. 12029–12041, Jul. 2022, https://doi.org/10.1007/s00521-022-07091-x.