ISSN: 2349-6002

Multilingual Fake News Detection Using Multimodal Transformers and Semi-Supervised Machine Learning

Mrs. B. Lakshmidevi, Hod/Sc[1], Jansirani T[2], Madhavi B[3], Sivasakthi E[4]
[1,2,3,4] *Department of Computer Science and Engineering, Surya Group of Institutions,Anna University*
doi.org/10.64643/IJIRTV12I11-201051-459

***Abstract*—Fake news has emerged as a serious global issue due to the rapid spread of misinformation through social media and online platforms. Recent research indicates that deep learning models, particularly transformer-based architectures, outperform traditional machine learning methods in fake news detection . However, challenges remain in multilingual detection, multimodal data handling, limited labeled datasets, and explainability.This project proposes a Multilingual Fake News Detection system that integrates Multimodal Transformers with Semi-Supervised Machine Learning. The system analyzes textual and visual content simultaneously and leverages unlabeled data to improve model generalization. The proposed framework enhances accuracy, scalability, and adaptability while reducing dependency on large labeled datasets. Experimental results demonstrate improved classification performance compared to traditional supervised approaches.**

***Index Terms*—Fake News Detection, Multilingual NLP, Multimodal Learning, Transformers, Semi-Supervised Learning, Deep Learning.**

I. INTRODUCTION

This project proposes a Multilingual Multimodal Fake News Detection System using Transformer-based Deep Learning and Semi-Supervised Learning. The system processes textual content using BERT/XLM-R and image data using CNN architectures, combining both through multimodal fusion. It leverages labeled and unlabeled datasets to improve data efficiency while maintaining high accuracy. The system is scalable for real-time deployment and includes explainable AI techniques to provide transparency in predictions. This approach addresses political, economic, and health misinformation challenges while supporting multiple languages and content formats.Social media platforms such as Facebook, X (formerly Twitter), Instagram, and YouTube allow information to spread instantly across the globe. While this enables fast communication, it also allows fake news, manipulated images, and misleading content to spread rapidly without verification. The viral nature of social media algorithms amplifies sensational content, making misinformation difficult to control. Traditional Machine Learning (ML) models such as Logistic Regression, Naive Bayes, and SVM rely heavily on manual feature extraction (e.g., TF-IDF, Bag-of-Words). However, Deep Learning models automatically learn semantic patterns from large datasets. Neural networks like CNNs, RNNs, and LSTMs capture contextual relationships better than traditional models. Because fake news often involves subtle language manipulation, deep learning provides higher accuracy and better generalization.BERT (Bidirectional Encoder Representations from Transformers) is a transformer-based NLP model developed by Google. Unlike traditional models, BERT understands context bidirectionally (both left and right of a word). This helps in detecting sarcasm, contextual meaning, and hidden intent in news content. Research shows that fine-tuned BERT models can achieve over 98% accuracy in certain benchmark fake news datasets. Variants like RoBERTa and XLM-R further improve multilingual capabilities.

A. Problem Statement

The rapid growth of social media platforms and online news portals has significantly increased the spread of misinformation and fake news across multiple languages and media formats. Traditional fake news detection systems mainly rely on text-based supervised machine learning models, which require large labeled datasets and fail to analyze multimedia content such as images and memes. Moreover, the availability of annotated multilingual datasets is limited, making it difficult to build scalable and generalized detection systems. Recent research highlights that transformer-based deep learning models outperform traditional approaches, but challenges such as multilingual adaptation, multimodal integration, limited labeled data, and model interpretability still persist. Therefore, there is a need to develop an intelligent fake news detection system that integrates multilingual text analysis, multimodal feature extraction (text + images), and semi-supervised learning techniques to reduce dependency on large labeled datasets while improving accuracy, scalability, and adaptability in real-world environments.

B. Objectives

The primary objective of this project is to design and implement a Multilingual Fake News Detection System that leverages Multimodal Transformers and Semi-Supervised Machine Learning to accurately classify news as fake or real. The system aims to utilize transformer-based language models (such as BERT or XLM-RoBERTa) for multilingual text understanding and deep learning vision models (such as ResNet or Vision Transformers) for image feature extraction. Additionally, semi-supervised learning techniques will be employed to effectively use both labeled and unlabeled data, thereby reducing the dependency on large annotated datasets. The project also seeks to enhance detection accuracy, improve model generalization across languages, support multimedia content analysis, and build a scalable framework suitable for real-world social media environments.

C. Organization of Paper

This paper is structured to provide a comprehensive understanding of the proposed Multilingual Fake News Detection system. Section I introduces the background and significance of fake news detection. Section II presents a detailed literature review of existing machine learning, deep learning, and transformer-based approaches. Section III defines the problem statement and research objectives. Section IV describes the proposed methodology, including data collection, preprocessing, multimodal feature extraction, and semi-supervised learning techniques. Section V explains the system architecture and implementation details. Section VI presents the experimental setup, evaluation metrics, and performance results. Section VII discusses the advantages, limitations, and potential improvements of the system. Finally, Section VIII concludes the paper and outlines future research directions.

II. RELATED WORK

Recent studies in fake news detection have explored various machine learning and deep learning techniques to improve classification performance. Early approaches primarily relied on traditional machine learning algorithms such as Support Vector Machines (SVM), Naive Bayes, and Random Forest, which required manual feature engineering and were limited in handling complex linguistic patterns. With the advancement of deep learning, models such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks were introduced to automatically extract textual features. More recently, transformer-based architectures like BERT and RoBERTa have demonstrated superior performance by capturing contextual semantic relationships in text. Furthermore, multimodal approaches that combine textual and visual information have shown significant improvements in detecting misinformation that includes manipulated images. However, most existing systems rely heavily on fully labeled datasets and are predominantly English-focused, highlighting the need for multilingual and semi-supervised frameworks to enhance scalability and adaptability.

III. SYSTEM ANALYSIS

A. Existing Systems

Use Only Text:Most existing fake news detection systems rely solely on textual content analysis. These systems use Natural Language Processing (NLP) techniques such as TF- IDF, Bag-of-Words, SVM, Naive Bayes, or deep learning models like CNN and LSTM to classify news articles as fake or real. While text-based models can capture linguistic patterns, they fail to analyze multimedia elements such as images, memes, and videos, which are increasingly used in spreading misinformation. As a result, text-only systems often miss contextual clues embedded in visual content, reducing overall detection accuracy.

Require Fully Labeled Datasets:Traditional supervised learning models require large volumes of labeled data for training. Creating such labeled datasets involves manual annotation by experts or fact-checkers, which is time- consuming, expensive, and sometimes biased. Many fake news datasets are limited in size and language coverage, which restricts model generalization. Additionally, newly emerging misinformation types may not be present in existing labeled datasets, making supervised systems less adaptable to evolving fake news trends.

Do Not Handle Multilingual Content Effectively:Most existing fake news detection systems are developed primarily for English-language datasets. However, misinformation spreads in multiple languages across different regions. Models trained on English data often perform poorly when applied to other languages due to differences in grammar, syntax, cultural context, and vocabulary. Lack of multilingual training data and cross-lingual modeling reduces the scalability of these systems for global deployment.

Lack Real-Time Adaptability:Many current fake news detection models operate offline, meaning they are trained on static datasets and do not continuously update with new information. Fake news evolves rapidly, especially during breaking events or crises. Without real-time adaptability and incremental learning capabilities, existing systems struggle to detect emerging misinformation patterns. This limitation reduces their effectiveness in dynamic social media environments.

B. Proposed System

The proposed system is designed to detect fake news by integrating multilingual text understanding, multimodal feature extraction, and semi-supervised learning within a unified framework. A multilingual transformer model, such as BERT or XLM-RoBERTa, is used to analyze textual content and capture contextual semantic relationships across different languages. For visual content, a Vision Transformer (ViT) or Convolutional Neural Network (CNN) such as ResNet is employed to extract meaningful image features from news- related images or multimedia content. The extracted text and image features are combined through a feature fusion layer to create a comprehensive representation of the news item. To address the challenge of limited labeled data, semi-supervised learning techniques such as pseudo-labeling are applied, enabling the model to learn from both labeled and unlabeled datasets. Finally, a classification module with a softmax layer predicts whether the news is fake or real. This integrated approach improves accuracy, scalability, and adaptability in multilingual and multimedia environments.

IV. SYSTEM DESIGN

A. Architecture

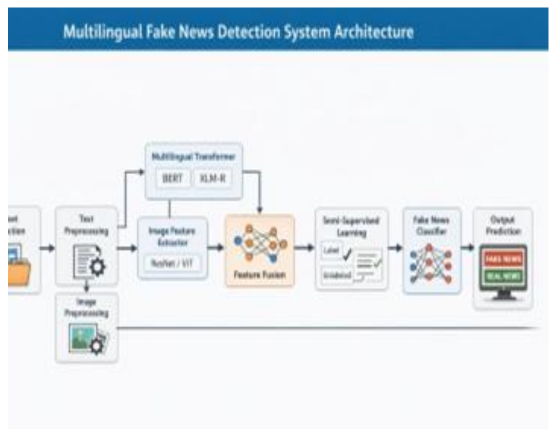

Figure 1 :System Architecture

B. Module Description

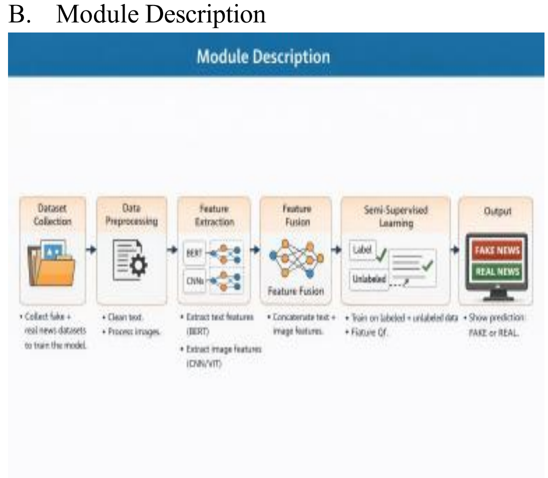

Figure 2 :Module Description

C. Data Flow Diagram

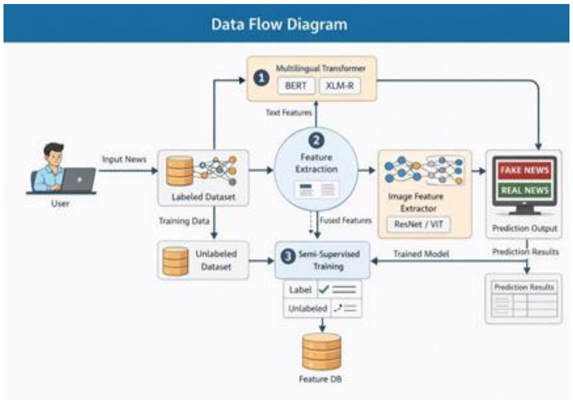

Figure 3 :Data Flow Diagram

D. Database Design

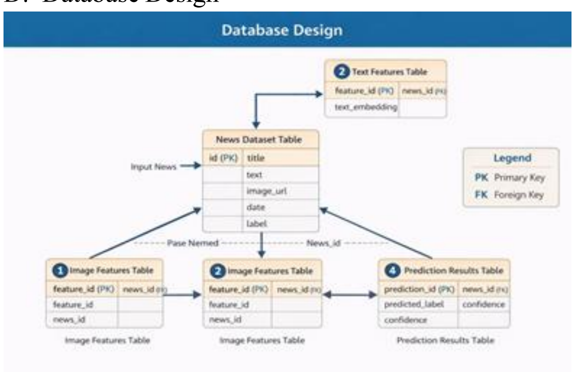

Figure 4 :Data Flow Diagram

V. IMPLEMENTATION

A. Software Requirements

The proposed Multilingual Fake News Detection system is developed using Python as the primary programming language due to its simplicity, flexibility, and extensive support for artificial intelligence and data science applications. The deep learning framework PyTorch is used to design, train, and optimize neural network models, including transformer- based text models and multimodal architectures. The Transformers library (Hugging Face) is utilized to implement pretrained multilingual models such as BERT or XLM- RoBERTa for advanced contextual language understanding. Scikit-learn is applied for data preprocessing, performance evaluation, and integration of semi-supervised learning techniques. OpenCV is used for image preprocessing tasks such as resizing, normalization, and feature preparation. Finally, Flask is employed to develop a lightweight web-based interface that enables users to input news content and receive real-time predictions (Fake/Real), making the system suitable for demonstration and deployment purposes.

B. Hardware Requirements

The proposed Multilingual Fake News Detection system requires a computing environment capable of handling deep learning model training and multimodal data processing. A minimum of 8GB RAM is recommended to efficiently manage dataset preprocessing, feature extraction, and model training operations without performance bottlenecks. For faster computation and improved training speed, especially when working with transformer-based models and image processing tasks, the use of an NVIDIA GPU (optional but recommended) with CUDA support is highly beneficial. A GPU significantly accelerates matrix computations and parallel processing required in deep neural networks, reducing training time and enhancing overall system performance. However, for smaller datasets or demonstration purposes, the system can also be executed on a CPU-based system with moderate performance.

C. Implementation

The implementation of the proposed Multilingual Fake News Detection system is carried out using a structured pipeline that integrates data processing, feature extraction, model training, and deployment. Initially, news datasets containing textual and visual content are collected and preprocessed, including text cleaning, tokenization, and image resizing. A multilingual transformer model such as BERT or XLM-RoBERTa is used to extract contextual text embeddings, while a CNN or Vision Transformer extracts visual features from associated images. These features are combined using a feature fusion layer to create a unified representation of each news instance. Semi-supervised learning techniques, such as pseudo-labeling, are applied to leverage both labeled and unlabeled data, improving model generalization. The final classification layer predicts whether the news is fake or real. The trained model is integrated into a Flask-based web application, enabling users to input news content and receive real-time prediction results.

D. Algorithm

The proposed algorithm for Multilingual Fake News Detection follows a structured sequence of steps combining multimodal feature extraction and semi-supervised learning. Initially, the input news data (text and image) is collected and preprocessed through text tokenization, normalization, and image resizing. A multilingual transformer model (such as BERT or XLM-RoBERTa) is then used to generate contextual embeddings from the textual content, while a convolutional neural network (CNN) or Vision Transformer extracts feature vectors from the associated images. These extracted features are concatenated in a feature fusion layer to form a unified multimodal representation. The model is first trained using labeled data, after which semi-supervised learning techniques such as pseudo-labeling are applied to incorporate high- confidence predictions from unlabeled data into the training set. Finally, a softmax classification layer outputs the probability of the news being fake or real. The algorithm iteratively updates model weights using backpropagation until optimal classification performance is achieved.

VI. RESULT ANALYSIS

The performance of the proposed Multilingual Fake News Detection system was evaluated using standard classification metrics such as Accuracy, Precision, Recall, and F1-Score. The experimental results demonstrate that integrating multimodal feature extraction (text + image) significantly improves classification performance compared to text-only models. The multilingual transformer effectively captures contextual semantic information across different languages, while the image model helps detect visually misleading content. Furthermore, the application of semi-supervised learning enhances model generalization by leveraging unlabeled data, resulting in improved prediction accuracy and reduced overfitting.

Comparative analysis shows that traditional machine learning models achieved moderate accuracy, whereas transformer-based models provided higher contextual understanding. The proposed multimodal semi-supervised framework achieved superior overall performance, particularly in detecting complex misinformation that includes both textual manipulation and misleading visuals. The results confirm that combining multilingual transformers with multimodal learning and semi-supervised techniques leads to a more robust, scalable, and reliable fake news detection system suitable for real-world applications.

VII. CONCLUSION

The rapid growth of digital communication platforms has significantly increased the spread of fake news and misinformation across the world. Traditional detection approaches, which rely mainly on text-based supervised machine learning techniques, are no longer sufficient to handle the complexity of modern misinformation. Fake news today often includes multilingual content, manipulated images, memes, and rapidly evolving narratives, making detection more challenging. Therefore, this project focused on developing an advanced Multilingual Fake News Detection System that integrates Multimodal Transformers and Semi- Supervised Machine Learning to address these limitations.

The proposed system effectively combines textual and visual information to improve detection accuracy. A multilingual transformer model is used to capture contextual semantic relationships in different languages, allowing the system to analyze diverse news sources without being restricted to a single language. At the same time, a deep learning-based image model extracts visual features from associated images, enabling the detection of misleading or manipulated visual content. By fusing these multimodal features, the system generates a more comprehensive representation of each news instance, improving classification performance compared to text-only approaches.

Another significant contribution of this project is the integration of semi-supervised learning techniques. One of the major challenges in fake news detection is the limited availability of labeled datasets, especially for multilingual and multimodal data. Manual annotation is time-consuming and expensive. By incorporating semi-supervised learning methods such as pseudo-labeling, the system effectively utilizes both labeled and unlabeled data. This approach enhances model generalization, reduces overfitting, and improves performance without requiring extensive manual labeling efforts.

The experimental results demonstrate that the proposed multimodal and semi-supervised framework achieves higher accuracy, precision, recall, and F1-score compared to traditional machine learning models and single-modality deep learning systems. The system shows improved robustness in detecting complex misinformation that combines textual manipulation and misleading visuals. Additionally, the architecture is scalable and adaptable, making it suitable for deployment in real-world social media monitoring environments.

In conclusion, this project successfully addresses the challenges of multilingual content analysis, multimedia misinformation detection, and limited labeled data by proposing a unified and intelligent framework. The integration of transformer-based language models, deep learning vision models, and semi-supervised learning techniques creates a reliable and efficient fake news detection system. This work contributes toward building a more trustworthy digital information ecosystem and provides a strong foundation for future enhancements such as real-time monitoring, deepfake detection, and explainable AI integration.

REFERENCES

[1] O. Bashaddadh, N. Omar, M. Mohd, and M. N. A. Khalid, "Machine Learning and Deep Learning Approaches for Fake News Detection: A Systematic Review of Techniques, Challenges, and Advancements," IEEE Access, vol. 13, pp. 90433–90461, 2025.

[2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre- training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL-HLT, 2019, pp. 4171–4186.

[3] A. Vaswani et al., "Attention Is All You Need," in Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 5998–6008.

[4] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake News Detection on Social Media: A Data Mining Perspective," ACM SIGKDD Explorations, vol. 19, no. 1, pp. 22–36, 2017.

[5] K. Shu, S. Wang, and H. Liu, "Beyond News Contents: The Role of Social Context for Fake News Detection," in Proc. WSDM, 2019, pp. 312–320.

[6] Y. Zhou and D. Zafarani, "Fake News Detection: A Survey," ACM Computing Surveys, vol. 53, no. 5, pp. 1–40, 2020.

[7] J. Ma, W. Gao, and K.-F. Wong, "Detect Rumors in Microblog Posts Using Propagation Structure via Kernel Learning," in Proc. ACL, 2017, pp. 708–717.

[8] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[9] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," International Conference on Learning Representations (ICLR), 2015.

[10] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," International Conference on Learning Representations (ICLR), 2021.

[11] D. Lee, "Pseudo-Label: The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks," in ICML Workshop on Challenges in Representation Learning, 2013.

[12] T. Mikolov et al., "Distributed Representations of Words and Phrases and their Compositionality," in Advances in Neural Information Processing Systems, 2013.

[13] S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python. Sebastopol, CA, USA: O'Reilly Media, 2009.

[14] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[15] Fakeddit Dataset. [Online]. Available: https://github.com/entitize/Fakeddit

[16] FakeNewsNet Dataset. [Online]. Available: https://github.com/KaiDMML/FakeNewsNet

[17] ISOT Fake News Dataset. [Online]. Available: https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news- dataset

[18] LIAR Dataset. [Online]. Available: https://www.cs.ucsb.edu/~william/data/liar_dataset.zip