NARASARAOPETA ENGINEERING COLLEGE (AUTONOMOUS)

Department of Computer Science and Engineering

Unifying Vision and Language for Robust Fake News Detection Using Novel Deep samples

PRESENTED BY

Shaik Siraz (22471A05O2)

Shaik Malka Jan Shafi (22471A05O9)

Nuti Nanda Kameswar (23475A0504)

Under the Guidance of,

Ch.Chandra SekharM.Tech, Ph.D

Department of Computer Science and Engineering,

Narasaraopeta Engineering College (Autonomous),

Narasaraopet- 522601.

OUTLINE

1. Abstract
2. Introduction
3. Literature Survey
4. Research Gaps
5. Problem Statement
6. Objectives
7. Block Diagram / Flow Diagram
8. Methodology
9. Implementation
10. Results and Analysis
11. Conclusion & Future Scope
12. References
13. Question and Answers
14. Acknowledgements

ABSTRACT

- Fake news identification has gained its relevance over the last few years as a result of large-scale propagation of fake information through social media.The paper presents a new method for detecting fake news that uses both text and image information together identification with multimodal learning that combines both text and image modalities. Utilizing the Fakeddit dataset, three new models were created and tested:(1)Retrained MLP Classifier with BERT + MobileNetV2 (91 accuracy), (2) CLIP + MLP (88.24accuracy), and (3) DistilBERT + EfficientNet + MLP (89 accuracy). The three models all achieve better performance than the baseline 88.83 from the original paper. This paper proves that combining different architectures beyond conventional literature can achieve better classification results in fake news.

INTRODUCTION

- Fake news is a major threat to public debate, democracy, and public health, especially due to rapid social media propagation.
- Increased availability of content creation tools and algorithmic amplification make distinguishing real from fake information difficult.
- Human fact-checking methods are accurate but slow and not scalable for mass monitoring.
- Early fake news detection approaches mainly used linguistic features and classical machine learning models.
- Deep learning models like BERT, LSTM, and CNNs capture complex context and semantics in text, improving detection accuracy.
- These models utilize strong pre-trained encoders (BERT, CLIP, DistilBERT, MobileNetV2, EfficientNet) and multi-layer perceptrons for classification.
- The research demonstrates better accuracy and efficiency compared to established baseline approaches.

NARASARAOPETA ENGINEERING COLLEGE (AUTONOMOUS)

LITERATURE SURVEY

No	Title	Author	Journal Name & Year	Methodology Adapted	Key Findings	Gaps
1	Fake News Detection using Deep Learning using a Systematic Literature Review	Mohammad Q. Alnabhan	https://ieeexplore.ieee.org/document/10614154	Text Cleaning and Normalization Tokenisation, Sequence Padding .	Systematic Survey of DL Algorithms, Dataset Analysis	Limited Use of Transfer Learning, Lack of Generalizable Models
2	Advancing Fake News Detection: Hybrid Deep Learning With FastText and Explainable AI	Subhan Ali	https://ieeexplore.ieee.org/document/10713872	Tokenization,lemmatization, duplicate removal, regex; BERT, LSTM, transformer-based	FastText hybrid models enhanced performance	Fake news mimics real news language, needing deeper context-aware models
3	Fake News Detection Landscape: Datasets, Data Modalities, AI Approaches, Their Challenges, and Future Perspectives	Seemab Hameed	https://ieeexplore.ieee.org/document/10937488	Text cleaning, image preprocessing; text-based and multimodal fusion models	Taxonomy of predictio-n architectures, multilingual and multimodal focus	Majority of studies focus on English, no standard benchmarks
4	Enhancing Fake News Detection by Multi-Feature Classification	Ahmed Hashim	https://ieeexplore.ieee.org/document/10343156	Textcleaning,normalization; global, spatial, temporal feature extractors, classifiers	Improved classification performance	Language limitations, no transfer learning
5	Fake News Classification Methodology With Enhanced BERT	AmnaZafa	https://ieeexplore.ieee.org/document/10742347/	Lowercasing all texts,Comparitve models	Enhanced BERT improves classification	Underutilization of enhanced BERT layers, Most of studies used plain BERT

LITERATURE SURVEY

- Researchers use a variety of methodologies for fake news detection, including deep learning models like BERT, CNN, LSTM, BiLSTM, transformer-based models, and hybrid approaches combining text and image features.
- Preprocessing techniques such as text cleaning, normalization, tokenization, lemmatization, stopword removal, image preprocessing, and feature selection (PCA, Chi-square) are standard for preparing data.
- Many studies focus on enhancing model interpretability by adding layers like fuzzy logic or explainable AI, and by developing multimodal architectures that combine textual and visual content.
- Key findings highlight improved classification performance, superior interpretability, effectiveness of feature-based models, and advantages of hybrid architectures over standalone models
- Deep learning, particularly transformer-based language models like BERT and its variants, dominates current methodologies due to their strong contextual understanding.

RESEARCH GAPS

1. Subhan Ali: Fake news often mimics real news language, requiring deeper context-aware models.
2. Seemab Hameed: Majority of studies focus on English; absence of standard benchmark datasets.
3. Amna Zafar: Prior studies mostly use plain BERT; underutilization of enhanced BERT architectures.
4. Mohammad Q. Alnabhan: Limited use of transfer learning; lack of generalizable models across datasets.
5. Ahmed Hashim: Language limitations in datasets; limited application of transfer learning.
6. Insufficient availability of large, balanced, and standard benchmark datasets for robust evaluation.
7. **Overall**: Need for more real-world, scalable, and multilingual solutions to tackle the evolving nature of fake news.

PROBLEM STATEMENT

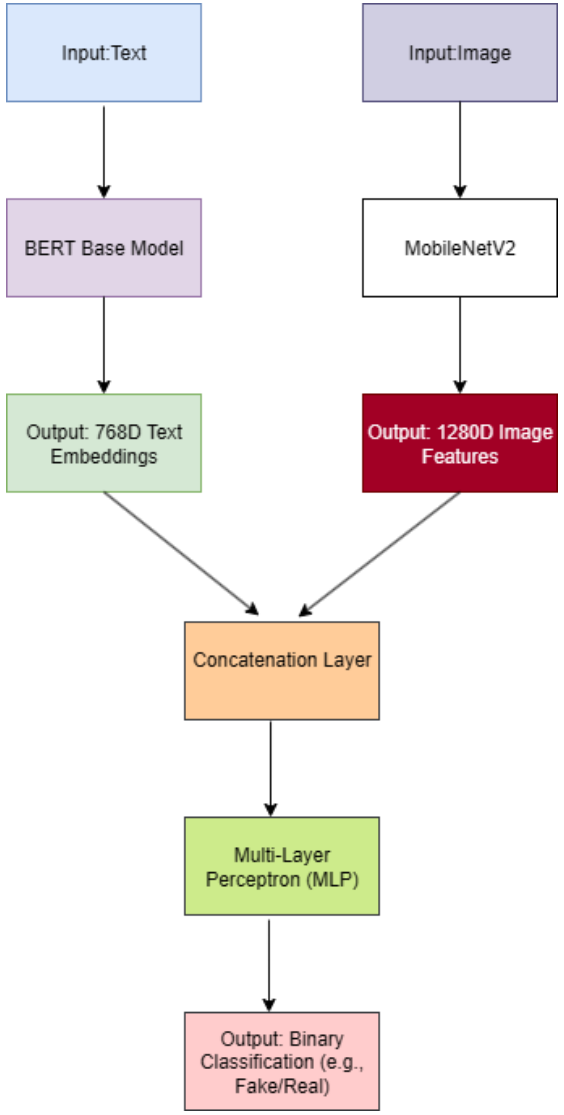
- Fake news spreads rapidly on social media, impacting public opinion, health, and democracy, creating an urgent need for effective automated detection systems.
- Existing fake news detection models often lack generalizability across diverse datasets, languages, and domains, limiting real-world applicability.
- Most current approaches rely heavily on English datasets and single modalities (text or image), while multimodal and multilingual methods need further development.
- Transfer learning techniques are underutilized despite their potential to improve model robustness and accuracy.
- There is a scarcity of large, balanced, annotated datasets representative of real-world news spread, negatively affecting model training and evaluation.
- Many models struggle to detect fake news early in its propagation, which is crucial to minimizing its harmful impact.

OBJECTIVES

Our Research work aims to:

- Develop new multimodal fake news detection models that effectively combine both text and image modalities for improved classification accuracy.
- Utilize strong pre-trained encoders like BERT, CLIP, DistilBERT for text and MobileNetV2, EfficientNet for images in hybrid architectures.
- Introduce three novel model architectures: BERT MobileNetV2 MLP, CLIP MLP, and DistilBERT EfficientNet MLP, to achieve efficient late fusion of multimodal features.
- Evaluate these models on the Fakeddit dataset, focusing on binary classification of news as real or fake.
- Achieve better performance in terms of precision and overall classification metrics compared to baseline methods.

BLOCK DIAGRAM OR FLOW DIAGRAM



METHODOLOGY

Data Description:

- Develop new multimodal fake news detection models that effectively combine both text and image modalities for improved classification accuracy.
- Utilize strong pre-trained encoders like BERT, CLIP, DistilBERT for text and MobileNetV2, EfficientNet for images in hybrid architectures.
- Introduce three novel model architectures: BERT MobileNetV2 MLP, CLIP MLP, and DistilBERT EfficientNet MLP, to achieve efficient late fusion of multimodal features.
- Evaluate these models on the Fakeddit dataset, focusing on binary classification of news as real or fake.
- Achieve better performance in terms of precision and overall classification metrics compared to baseline methods.

Dataset Split	# Samples	Real	Fake
Train	40,000	20,000	20,000
Validation	5,000	2,500	2,500
Test	5,000	2,500	2,500

2. Data Preprocessing

The pre-process consisted of the following steps:

1. Text Cleaning: Lowercasing, special character removal, and tokenization.

2. Image Handling: Images were resized to 224x224 and normalized with ImageNet statistics.

3. Filtering: The rows with missing image files or blank titles were filtered out.

4. Label Encoding: Two-class labels were encoded as binary (0 = real, 1 = fake).

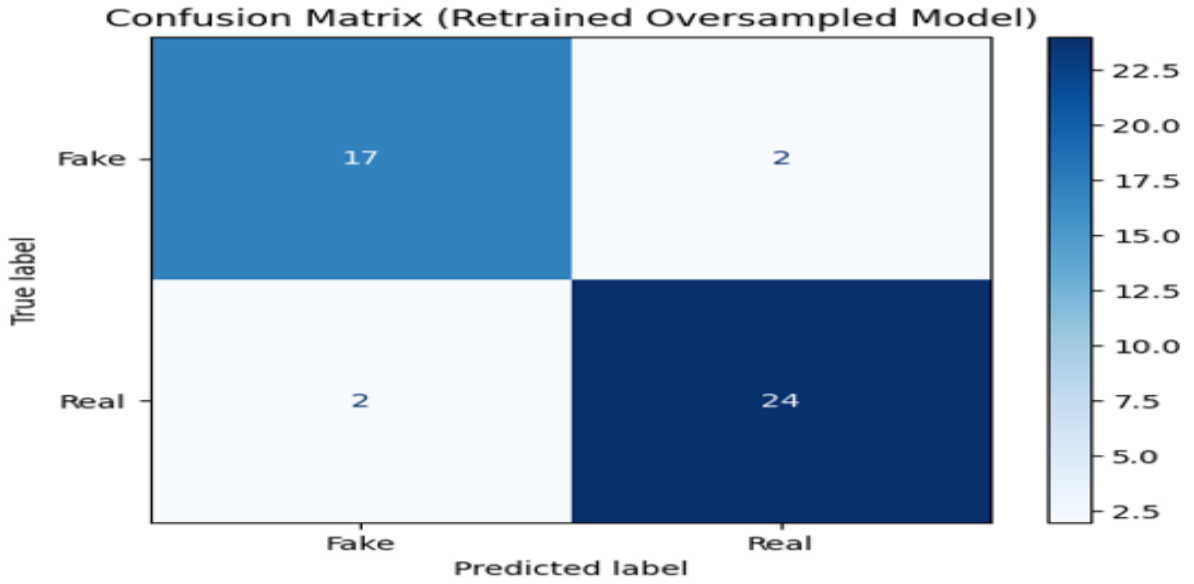
5. The data set was divided into three parts: training (80), validation (10) and test (10). Each sample was converted to tensor form appropriate for model input.

3.Model Architecture:

Three hybrid models were introduced and trained:
(i) **BERT + MobileNetV2 + MLP**: This model employs a pre-trained BERT base model to embed text into 768dimensionalvectors.TheimagesarefedthroughMobileNetV2 to get 1280-dimensional features. These vectors are concatenated and fed through an MLP for classification.
(ii) **CLIP + MLP: CLIP (Contrastive Language-Image Pretraining)** is employed to derive 512-dimensional unified embeddings from both image and text. These embeddings are directly fed into an MLP with two hidden layers and a final sigmoid output layer.
(iii) **DistilBERT + EfficientNet + MLP: DistilBERT** transforms the text into 768-dimensional features, whereas EfficientNet-B0 maps images into 1280 features. The concatenated vector is fed into an MLP with dropout and ReLU activations All models employ late fusion approaches and have a uniform binary classification output. Multimodal Feature Fusion: Let $T \in Rdt$ be the text embedding and $I \in Rdi$ be the image embedding. The resulting fused representation is:
$[F = MLP([T, \| ,I])$ (1) where $[\cdot \| \cdot]$ is vector concatenation, and MLP is the multilayer perceptron for final classification.

Evaluation Metrics and Data Presentation :

1. The study assesses the precision, precision, recall, and F1 score of the three models presented.
2. This facilitates a straightforward numerical comparison of the proposed methodologies.
3. Confusion Matrices: Confusion matrices were developed to demonstrate the specific strengths and weaknesses of each model.
4. These matrices illustrate the percentage of correct and incorrect predictions for both the 'actual' and 'fake' categories. Bar charts: The precision, recall and F1 scores for each category (fake and real) were represented using bar charts for the different models



MATERIALS AND METHODS:

A.Text-Based Approaches (Single-Modality): Textual fake news detection has been traditionally addressed

- Through deeplearningmodels such as CNNs,RNNs,and,more recently, transformers utilized a BiLSTMCRF model with semantic attention to capture contextual dependency in the classification of rumor
- It presented a domain-adaptive variant of BERT for fake news detection in the political sphere.
- Transformer models such as RoBERTa and XLNet have also demonstrated better performance on such datasets as LIAR and BuzzFeed.
- Even with these advances, single-modality techniques have difficulty with content that includes deceptive multimedia elements .There has therefore, been a focus on multimodal techniques.

B. Image-Based Approaches (Single-Modality) :

- Image-based approaches employ convolutional neural networks to detect visual patterns in manipulated images introduced a VGG19-based pipeline in 2023 for detecting doctored political images.
- Employed a CNN-RNN hybrid framework for detecting spatial and temporal semantics of misinformation .
- Though these approaches successfully examine visual content, they fail to capture the contextual information of related textual data.

C. Multimodal Fake News Detection:

- New multimodal models combine text and image modalities for more semantic representation base paper applied ensemble fusion of BERT + ResNet and XLNet + DenseNet to obtain 88.83 accuracy.
- New multimodal solutions have appeared to enhance early fusion and cross-modal alignment investigated CLIP embeddings to match text and image meanings for detecting fake news.
- It surpassed the performance of the traditional encoders Following work can be done in incorporating attention mechanisms across modalities, testing More recently, co-attention and graph neural networks (GNNs) have been introduced by recent frameworks.

D. Limitations in Existing Work:

- In spite of these developments, some problems persist. Many approaches are either computationally intensive at a large scale (e.g., ViT) or don't generalize well to noisy user-generated data.

- Certain models don't have strong fusion of features and stick to shallow concatenation. Furthermore, cross-modal inconsistencies aren't well captured in late fusion approaches.

- Our models to be proposed try to tackle these problems using light encoders (such as MobileNetV2 and EfficientNet).

- Underutilization of transfer learning methods, despite their potential to improve performance and adaptability.


Fake News Detection System
Enter the news text and upload an image to detect authenticity.
News Text:
lokha movie
Upload Image:
Choose File lokha.webp
LOKAH
Analyze News
Upload Image:
Choose File lokha.webp
LOKAH
Analyze News
Prediction Result
This news appears to be REAL.

IMPLEMENTATION

Software Specifications:

- Browser: Chrome
- Operating System: Windows11
- Python (COLAB)
- Flask

Hardware Specifications:

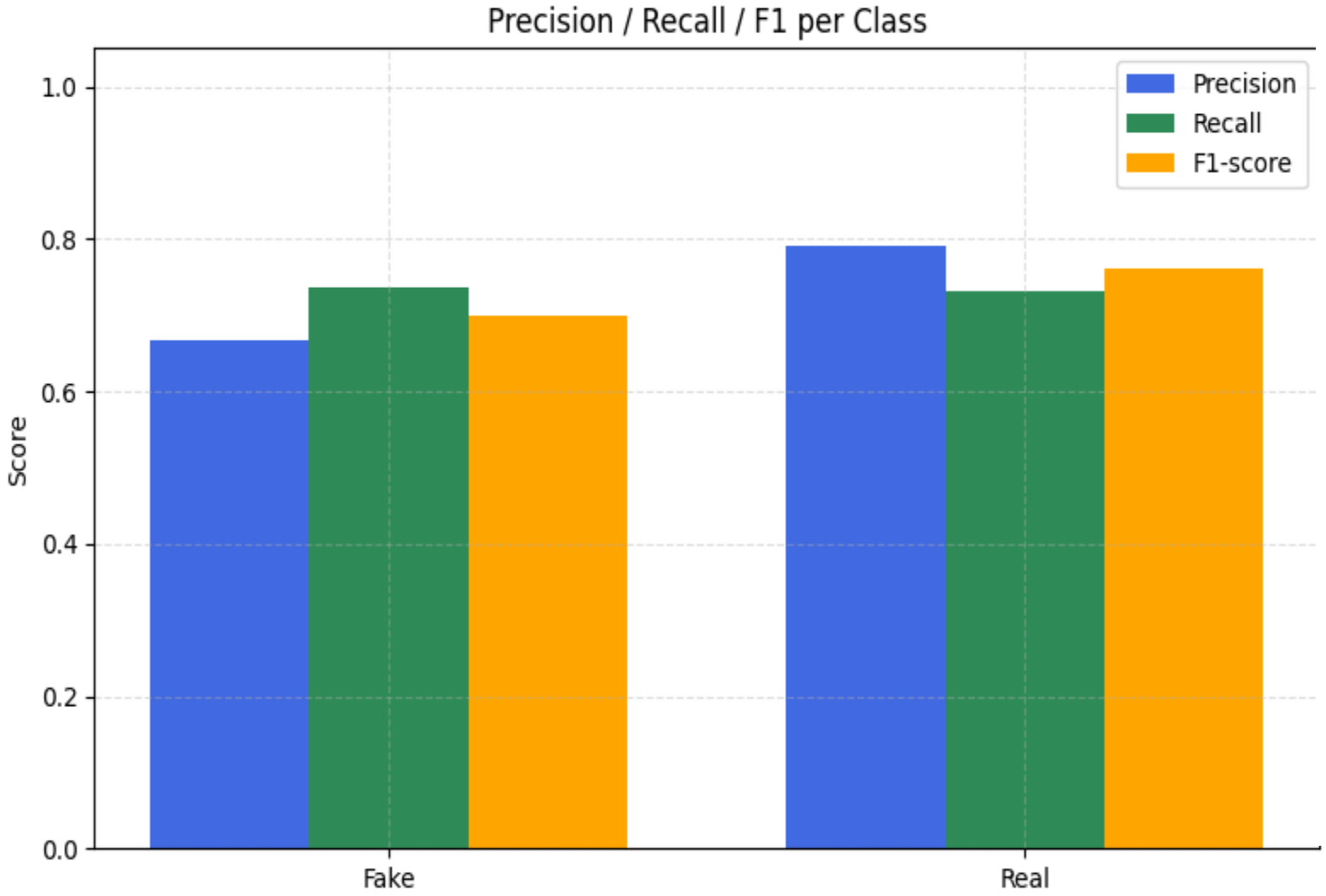
- Processor: Intel® Dual Core
- 2.0GHz Hard Disk: 1TB
- RAM: 8GB

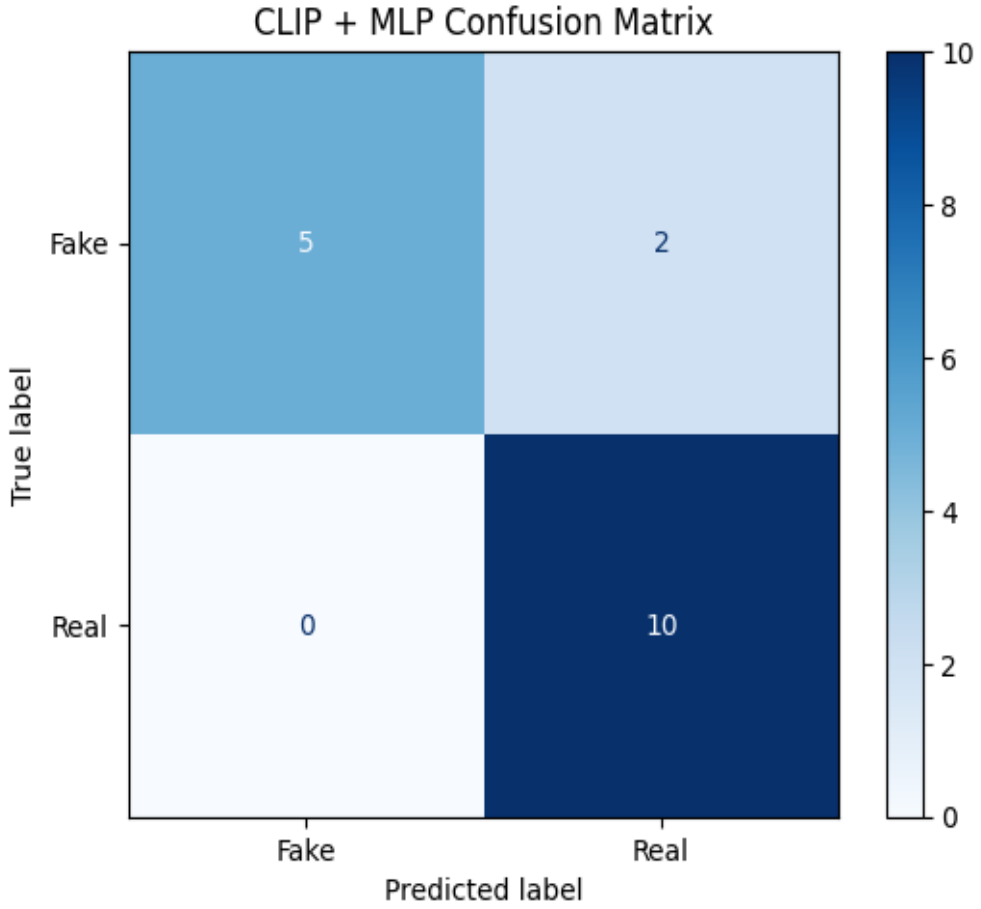
IMPLEMENTATION

```python
from tqdm import tqdm
import numpy as np

device = torch.device("cuda" if torch.cuda.is_available() else "cpu")

mobilenet = models.mobilenet_v2(pretrained=True).features
mobilenet.to(device).eval()

img_transform = transforms.Compose([
    transforms.Resize((224, 224)),
    transforms.ToTensor(),
    transforms.Normalize(mean=[0.485, 0.456, 0.406],
                         std=[0.229, 0.224, 0.225])
])

def extract_image_features(img_path):
    try:
        img = Image.open(img_path).convert("RGB")
        img = img_transform(img).unsqueeze(0).to(device)
        with torch.no_grad():
            feat = mobilenet(img)
            feat = torch.nn.functional.adaptive_avg_pool2d(feat, (1, 1))
        return feat.view(-1).cpu().numpy()
    except Exception as e:
        print(f"Image error: {e}")
        return None
```

```python
import torch
from torchvision import models, transforms
from PIL import Image
from tqdm import tqdm
import numpy as np
from transformers import DistilBertTokenizer, DistilBertModel # Using DistilBERT
from sklearn.model_selection import train_test_split
from sklearn.neural_network import MLPClassifier
from sklearn.metrics import accuracy_score, classification_report
import os
import pandas as pd # Ensure pandas is imported
import joblib # Import joblib for saving the model
from google.colab import drive # Import drive for mounting

# Ensure df_valid_for_resnet (or similar filtered DataFrame) is available
# Using df_valid_for_resnet assuming the previous check was successful and this DataFrame contains valid samples
if 'df_valid_for_resnet' in locals() and not df_valid_for_resnet.empty:
    print("✅ Using 'df_valid_for_resnet' DataFrame for training with", len(df_valid_for_resnet), "rows.")
    df_to_process = df_valid_for_resnet.copy() # Use a copy
elif 'df_filtered_valid_images' in locals() and not df_filtered_valid_images.empty:
     print("✅ 'df_valid_for_resnet' not found or empty. Using 'df_filtered_valid_images' DataFrame with", len(df_filtered_valid_images), "rows.")
     df_to_process = df_filtered_valid_images.copy()
elif 'df' in locals() and not df.empty:
    print("⚠️ Filtered DataFrame not found. Using original 'df' DataFrame with", len(df), "rows. Image errors may occur.")
    df_to_process = df.copy()
else:
    print("Error: No valid DataFrame loaded. Please load and preprocess your data first.")
    df_to_process = pd.DataFrame() # Create an empty DataFrame


if not df_to_process.empty:
    # --- Image Feature Extraction (EfficientNet) ---
    print("\nExtracting image features using EfficientNet...")
```

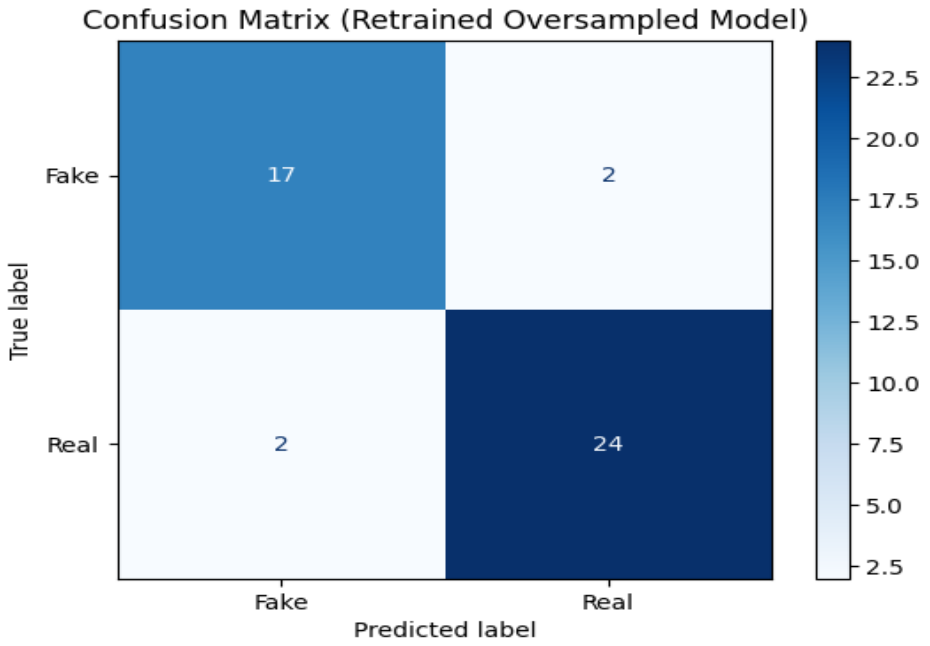
RESULTS & ANALYSIS





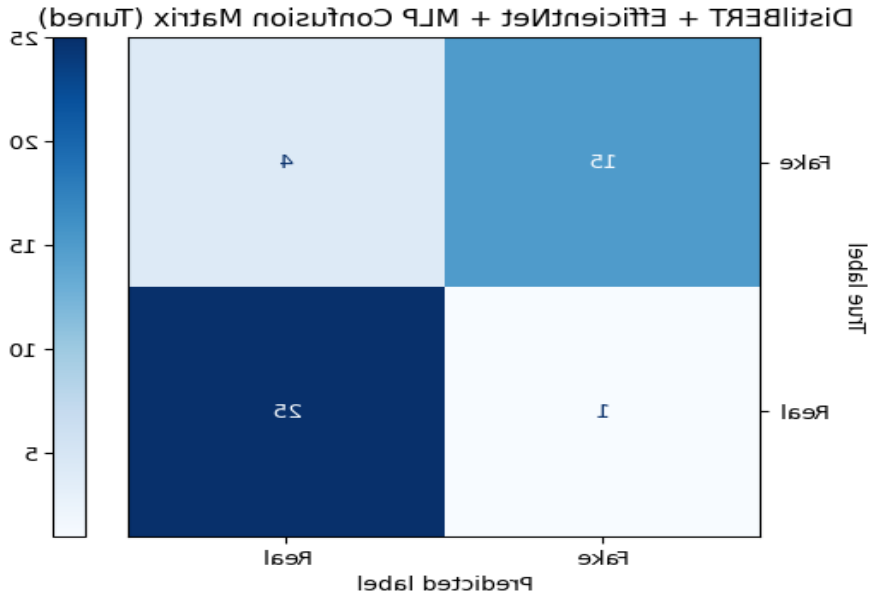
RESULTS & ANALYSIS

Approach	Accuracy	Precision	Recall	F1-Score
BERT + MobileNetV2 + MLP	91.03	0.92	0.91	0.91
CLIP + MLP	88.23	0.89	0.87	0.88
DistilBERT + EfficientNet + MLP	82.00	0.83	0.81	0.82



Confusion Matrix

Model Accuracies



Confusion Matrix for DistilBERT Mode

CONCLUSION and FUTURE SCOPE

Conclusion:

- The proposed work introduces a multimodal fake news detection system consisting of strong textual and visual feature extractors with light-weight classifiers for enhancing classification accuracy while preserving computational tractability.
- We introduced and experimented with three hybrid models: BERT + MobileNetV2 + MLP, CLIP + MLP, and DistilBERT + EfficientNet + MLP. The models were tested on the cleaned Fakeddit dataset multimodal samples.
- The block diagram serves as a visual blueprint of the overall methodology. It illustrates the late-fusion architecture employed by our models.

Future Work:

- **Explore incorporation of attention mechanisms across modalities to enhance cross-modal feature interaction and improve detection accuracy.**
- **Investigate generative data augmentation techniques to address limited training samples and improve model generalization.**

REFERENCES

- [1] F. Almeida and R. Silva, ”Bi-modal hybrid CNN-BERT model for fake news classification,” Comput. Hum. Behav. Rep., vol. 9, p. 100152, 2023.
- [2] D. Chen and Z. Li, ”Graph neural networks for fake news detection: A review,” IEEE Trans. Comput. Soc. Syst., vol. 11, no. 1, pp. 80–91, 2024.
- [3] Y. Choi et al., ”Image-text coherence networks for fake news detection,” Pattern Recognition, vol. 139, p. 109405, 2023. [4] L. Dai et al., ”Light multimodal transformers for realtime fake news classification,” IEEE Internet Comput., vol. 28, no. 2, pp. 32–41, 2024.
- [5] H. Gao et al., ”Contrastive learning for robust fake news detection across modalities,” Proc. ACM Multimedia, 2025.
- [6] S. Guo and L. Tang, ”Multi-modal semantic alignment networks for misinformation detection,” KnowledgeBased Systems, vol. 280, p. 110929, 2024.
- [7] A. Gupta et al., ”Transformer-enhanced multimodal fake news classifier,” Information Sciences, vol. 654, pp. 78–95, 2024.
- [8] K. Lakshminadh, D.C.V. Guptha, J. Sai, K. Rajesh, S. Moturi, Y. Neelima, and D.V. Reddy, “Advanced Pest Identification: An Efficient Deep Learning Approach Using VGG Networks,” in Proc. 2025 IEEE Int. Conf. Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), 2025, doi: 10.1109/IATMSI64286.2025.10984619.
- [9] M. Li et al., ”Generative augmentation for multimodal fake news detection,” ACM Trans. Multimedia Comput. Commun. Appl., 2025.

REFERENCES

- [10] Y. Liu and X. Zhang, ”Cross-modal transformer for multimodal fake news detection,” IEEE Access, vol. 11, pp. 21567–21576, 2023.
- [11] Y. Peng and J. Wang, ”Multi-modal fake news detection using vision-language transformers,” IEEE Trans. Multimedia, 2025.
- [12] S. Rafi, M.S. Reddy, M. Sireesha, A.L. Niharika, S. Neelima, and K. Nikhitha, “Detecting Sarcasm Across Headlines and Text,” in Proc. 2025 IEEE IATMSI, 2025, doi: 10.1109/IATMSI64286.2025.10984543.
- [13] S.S.N. Rao, C. Sunitha, S. Najma, N. Nagalakshmi, T.G.R. Babu, and S. Moturi, “Advanced Water Quality Prediction: Leveraging Genetic Optimization and Machine Learning,” in Proc. 2025 IEEE IATMSI, 2025, doi: 10.1109/IATMSI64286.2025.10984615.
- [14] S.N.T. Rao, T.C. Dulla, V.K. Kolla, G.S. Kurakula, M. Suneetha, S. Moturi, and D.V. Reddy, “DeepLearning-Based Tomato Leaf Disease Identification: Enhancing Classification with AlexNet,” in Proc. 2025 IEEE IATMSI, 2025, doi: 10.1109/IATMSI64286.2025.10984969.
- [15] S. Raza et al., ”Survey on multimodal misinformation detection,” IEEE Access, vol. 11, pp. 123456–123470, 2023.


Any Questions ?


THANK
YOU