International Journal on Emerging Research Areas (IJERA)

DOI:10.5281/zenodo.14649896 *ISSN : 2230-9993*

TrueNews: AI Powered Detection of Manipulated Text and Images

Ms. Ansamol Varghese
Dept. of Computer Science
Amal Jyothi College of Engineering
(Autonomous)
Kottayam, India
ansamolvarghese@amaljyothi.ac.in

Anandhu Anoj
Dept. of Computer Science
Amal Jyothi College of Engineering
(Autonomous)
Kottayam, India
anandhuanoj2025@cs.ajce.in

Emil Thomas
Dept. of Computer Science
Amal Jyothi College of Engineering
(Autonomous)
Kottayam, India
emilthomas2025@cs.ajce.in

Deepta K Sunny
Dept. of Computer Science
Amal Jyothi College of Engineering
(Autonomous)
Kottayam, India
deeptaksunny2025@cs.ajce.in

Angel Thomas
Dept. of Computer Science
Amal Jyothi College of Engineering
(Autonomous)
Kottayam, India
angelthomas2025@cs.ajce.in

***Abstract*—The proliferation of fake news across digital platforms has become a critical issue, leading to widespread misinformation with significant societal implications. This survey paper presents a comprehensive review of recent advancements in fake news detection, leveraging machine learning (ML), deep learning (DL), and natural language processing (NLP) techniques. The reviewed studies cover diverse approaches, ranging from content-based methods to the integration of social context, multimedia, and knowledge-enhanced models.**

Traditional machine learning algorithms such as Random Forest, Support Vector Machines, and logistic regression are commonly employed for binary classification tasks, using features derived from linguistic patterns, source credibility, and metadata. In addition, enhanced models such as bidirectional LSTM-RNN and hybrid CNN-LSTM architectures, coupled with FastText embeddings, demonstrate significant improvements in detecting fake news in real-time scenarios and across multimedia-rich datasets.

The integration of social network features alongside textual content is a growing focus, where user behavior and social capital contribute to a more comprehensive fake news detection process. Transformer-based models, such as BERT, XLNet, and RoBERTa, show promising results in handling syntactic and semantic complexities, outperforming traditional RNN-based methods. Additionally, knowledge-augmented models utilizing large-scale open knowledge graphs offer a novel direction for multi-modal fake news verification by enhancing the model's understanding of both textual and visual content.

The survey also highlights the growing trend toward explainable AI (XAI) in fake news detection, providing transparency and interpretability in decision-making. By employing state-of-the-art models alongside regularization techniques and hyperparameter optimization, these studies collectively strive to address key challenges in fake news detection, including early identification, data scarcity, and model generalization.

This survey concludes by emphasizing the need for continued innovation in scalable and robust fake news detection systems, integrating diverse data modalities and ensuring real-time detection capabilities across a range of online platforms.

***Index Terms*—Fake News Detection,Machine Learning,Deep Learning,Natural Language Processing,Social Media**

I. INTRODUCTION

In the era of rapid digital communication, the proliferation of fake news has become a critical challenge, exacerbated by the widespread use of social media platforms like Facebook, Twitter, and others. Fake news, defined as intentionally fabricated or misleading information, has far-reaching implications, influencing public opinion, manipulating elections, and spreading misinformation during critical events such as natural disasters and health crises. As information spreads quickly across the internet, distinguishing between factual and fabricated content is essential to preserve the integrity of information ecosystems.

The detection of fake news has garnered significant attention in recent years, with numerous approaches leveraging advancements in Artificial Intelligence (AI), Natural Language Processing (NLP), and Machine Learning (ML). These methods aim to provide automated solutions to the growing problem of misinformation, utilizing sophisticated algorithms to analyze news content, user behavior, and social networks. While early approaches primarily focused on text-based analysis, recent developments incorporate multimodal techniques, combining textual, visual, and social context features to enhance detection accuracy.

This survey presents an in-depth review of recent research in fake news detection, focusing on a variety of approaches ranging from traditional machine learning models to state-of-the-art deep learning architectures. The surveyed studies explore methods such as binary classification, hybrid deep learning models,

DOI: 10.5281/zenodo.14649896 *ISSN : 2230-9993*

bidirectional LSTM, FastText embeddings, and transformer-based frameworks like BERT and RoBERTa. Additionally, the integration of explainable AI and knowledge graphs has further refined t he d etection m echanisms, p roviding m ore r obust and interpretable models capable of identifying both textual and multimedia misinformation.

Through this survey, we provide a comprehensive overview of fake news detection methodologies, highlighting key advancements, challenges, and opportunities in the field. W e a im to contribute to the ongoing development of reliable and efficient fake news detection systems, capable of addressing the evolving tactics of misinformation spreaders.

II. Categorization of models

Based on the literature reviewed, fake news detection techniques can be broadly categorized into three main approaches:

1) Machine Learning-Based Approaches
2) Deep Learning and Neural Network Models
3) Multimodal Approaches

A. Machine Learning Based Approaches

Machine learning-based approaches rely on traditional algorithms such as Support Vector Machines (SVM), Random Forest, Naive Bayes, and Decision Trees to classify news articles or social media posts as real or fake. These methods depend heavily on feature extraction, where meaningful data such as word frequencies, sentiment, or part-of-speech tags are extracted from textual content to serve as input to the models.

For example, in [2] by Aldwairi and Alwahedi (2018), various machine learning classifiers, i ncluding S VM a nd Random Forest, were employed to detect fake news based on text features. These models required extensive feature engineering, where linguistic features like word count, sentiment scores, and punctuation usage were analyzed to differentiate fake news from real news. While these models were moderately accurate, their effectiveness depends on the quality of the extracted features, limiting their generalizability across diverse data.

Similarly, Sharma et al. (2020) in [5] applied machine learning algorithms such as logistic regression and decision trees to classify news articles. These models focus on basic features such as text length and term frequency-inverse document frequency (TF-IDF) to identify patterns in fake news. While machine learning methods are relatively simple and computationally efficient, they struggle to capture the deeper semantic and contextual nuances often found in fake news content. Additionally, the performance of these models tends to degrade when handling large-scale datasets or dealing with multimodal information such as images and social media context.

Overall, while machine learning-based approaches have been foundational in fake news detection research, they are often limited by the need for manual feature extraction, and their inability to fully capture complex relationships within the data—especially in a multimodal context.

B. Deep Learning and Neural Network Based Models

Deep learning models have revolutionized fake news detection by enabling the automatic extraction of features from large datasets without the need for manual feature engineering. These models, particularly neural networks like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and transformers, are capable of handling complex relationships in both text and image data, making them more adaptable to various forms of misinformation.

In [1] , Sormeily et al. (2024) introduced the MEFaND (Multimodal Framework for Early Fake News Detection) model, which employs deep learning techniques to analyze both textual and visual content simultaneously. This framework integrates CNNs for image processing and RNNs for text analysis, allowing the model to assess the veracity of news by correlating image-text pairs. The ability to handle both text and images gives MEFaND an edge over traditional machine learning models, as fake news often includes manipulated images or videos alongside misleading text. This model demonstrates how deep learning can handle the inherent complexity of multimodal content, improving detection accuracy.

In [6] , Ahmad et al. (2022) utilized a hybrid deep learning approach combining CNNs and RNNs to process text and images concurrently. The strength of deep learning lies in its ability to learn hierarchies of features, allowing these models to capture both low-level details (such as specific words or pixels) and high-level representations (such as sentence context or image composition). This dual processing capability enables more sophisticated detection mechanisms, particularly in cases where fake news is distributed through social media with misleading headlines and doctored images.

Deep learning models, such as those discussed in [13] by Hashmi et al. (2024), also benefit from advanced techniques like FastText and hybrid architectures that incorporate word embeddings and deep networks. FastText, for instance, provides a quick and efficient way to represent text data, which can then be fed into deep learning models for fake news detection. This combination of fast text representation and deep feature learning improves processing speed and accuracy, making it suitable for large-scale fake news detection tasks.

However, deep learning models are not without challenges. They often require large datasets to perform effectively and can be computationally intensive. Despite these limitations, deep learning approaches significantly outperform traditional machine learning models in capturing the subtleties of fake news, making them more effective at detecting misinformation across different modalities (text, image, and potentially video).

C. Multimodal Based Approaches

Multimodal approaches leverage multiple types of data—such as text, images, videos, and social media context—simultaneously to improve the accuracy and robustness of fake news detection. These models analyze the interactions between different modalities to identify

DOI:10.5281/zenodo.14649896 *ISSN : 2230-9993*

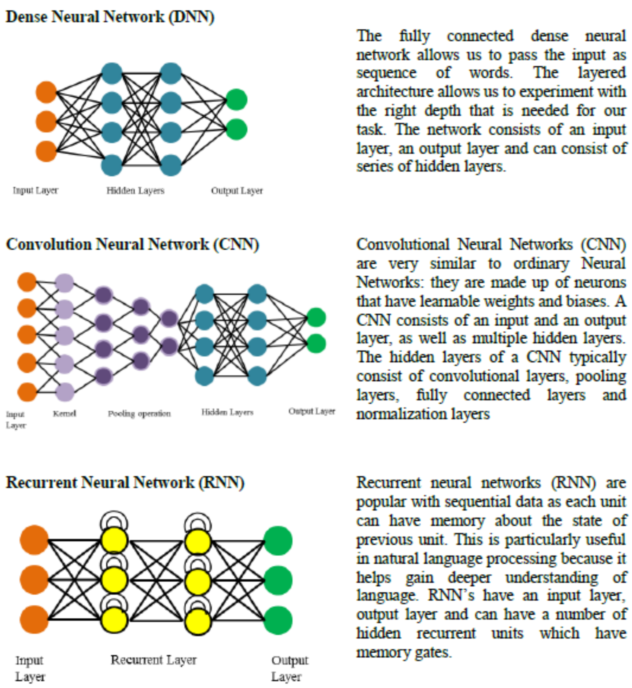

Fig. 1. Summary of Neural Network architectures.

inconsistencies or misleading information, making them more sophisticated than single-modality approaches.

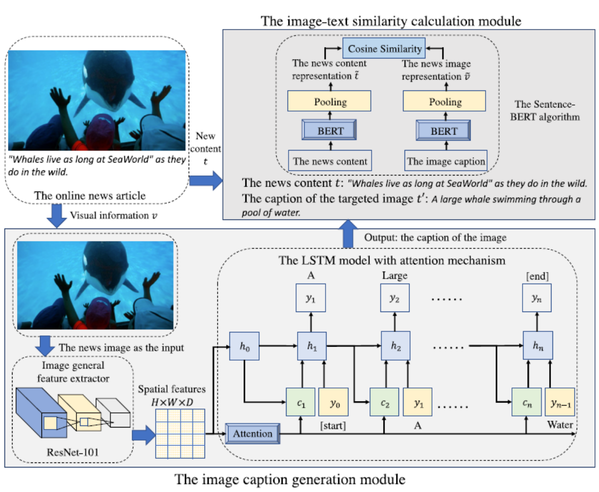

Fig. 2. Overall illustration of image captioning algorithm used.

In [1] , Sormeily et al. (2024) introduced the MEFaND (Multimodal Framework for Early Fake News Detection) model, which exemplifies the potential of multimodal detection. MEFaND processes both textual and visual content using Recurrent Neural Networks (RNNs) for text and Convolutional Neural Networks (CNNs) for images. By analyzing the correlation between images and text in news articles or social media posts, the model can detect discrepancies that are common in fake news. For example, a misleading headline might be paired with an unrelated or manipulated image to create false narratives. This integration of modalities improves the model's ability to detect such deceptive content early.

Similarly, Zhang et al. (2024) in [11] proposed a model based on image-text similarity for multimodal fake news detection. This approach focuses on evaluating the consistency between images and the accompanying text, checking for semantic alignment. If the text describes an event or scenario that is not visually supported by the image, the model flags it as potentially fake. Such approaches address the limitations of text-only or image-only models, which may miss context-specific signals that are revealed when combining modalities.

In [14] , Gao et al. (2024) introduced a Knowledge-Enhanced Vision and Language Model that utilizes both visual and linguistic data for fake news detection. This model is further enhanced by external knowledge bases, which provide additional context to verify the accuracy of the claims made in the content. This approach bridges the gap between multimodal analysis and knowledge verification, enabling more precise detection of misinformation that spans multiple types of content.

While these multimodal models provide a more comprehensive framework for fake news detection, they also introduce challenges, such as the need for more computational resources and complex data processing pipelines. Additionally, integrating multiple modalities can increase the complexity of model training and interpretation. Despite these challenges, multimodal approaches are well-suited to handle the increasingly sophisticated forms of fake news found on social media, where images, videos, and text are often combined to deceive readers.

III. Fake News Detection Models Overview

In the survey, we first explore the different machine learning models that have been employed for fake news detection. The main models used are:

Random Forest (RF) Support Vector Machine (SVM) Logistic Regression (LR) Neural Networks (NNET) XGBoost

Each model follows certain mathematical principles:

Logistic Regression (LR): Logistic regression [7] works by predicting the probability that an instance belongs to a particular class. The probability P is calculated as:

$$P(y=1|X) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n)}}$$

Here, $x_1, x_2, \ldots, x_n$ are the feature inputs, and $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ are the weights learned by the model.

LR was often used as a benchmark for comparison in most of the documents. It generally had lower accuracy compared to more advanced models like neural networks and SVM but provided interpretability, which was useful for understanding feature importance.

Support Vector Machine (SVM): SVM tries to find a hyperplane that best separates the two classes (fake and real news) [14] . The margin maximization principle is key:

DOI: 10.5281/zenodo.14649896 *ISSN : 2230-9993*

$$w^T x + b = 0$$

Where w is the weight vector, x represents the feature vector, and b is the bias term.

The objective is to maximize the margin while minimizing classification error:

$$\min \frac{1}{2}||w||^2 \quad \text{subject to} \quad y_i(w^T x_i + b) \geq 1, \ \forall i$$

SVM is particularly effective when the dataset involves textual features and exhibits clear separation between real and fake news. However, its reliance on selecting the correct kernel is noted as a challenge in some cases.In comparison to RF and logistic regression, SVM is found to be more accurate in detecting fake news when the text features are well-engineered. However, it struggles with large-scale data due to slower training times.

Random Forest (RF): RF is an ensemble method that creates multiple decision trees and outputs the mode of the classes (for classification) or mean prediction (for regression) of the individual trees [7] . For classification, the RF formula is expressed as:

$$f(x) = \frac{1}{N} \sum_{i=1}^{N} T_i(x)$$

Where $T_i(x)$ is the prediction from the i-th tree, and N is the total number of trees.RF is mentioned for its robustness in classification tasks, particularly when handling noisy data like fake news articles. RF performed well when there is a large set of features to handle, such as sentiment and text-based features.RF is often outperformed by neural networks or SVM in terms of precision and recall, but its ease of interpretation and lower computational cost made it preferable in situations with smaller datasets.

IV. Feature Importance and Study

The papers highlight an ablation study that emphasizes certain features in fake news detection:

Sentiment analysis, word similarity, in-degree centrality, and total number of tweets are major contributors to model performance. For example, in Random Forests [7] , the feature importance score is computed based on the Gini Index:

$$\text{Gini} = 1 - \sum_{k=1}^{K} p_k^2$$

Where p_k is the proportion of observations belonging to class k. Lower Gini scores indicate a better split in decision trees.

Ablation Equation:
An ablation study tests the impact of different feature combinations, such as:

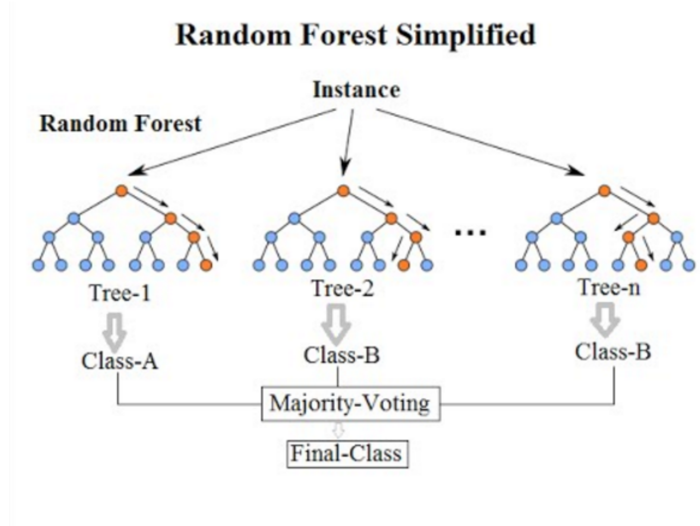

Fig. 3. Random Forest Decision Tree.

$$\text{Performance} = \text{Baseline Model Accuracy} - \text{Feature Removed Accuracy}$$

Where removing significant features results in a drop in accuracy, showing their importance.

Many documents explored feature importance using both text-based and network-based features. Textual features such as word sentiment and word embeddings were consistently identified as highly important. Some documents also included metadata (e.g., number of shares or likes) as a crucial feature.

1.Textual Features: Documents discussing feature importance agreed that word sentiment and n-grams were crucial in distinguishing real from fake news. In cases where sentiment was removed, model performance dropped significantly.

2.Network Features: Several documents incorporated network-based features, such as in-degree centrality (how many people retweet or share a piece of news). Removing such features caused a noticeable drop in model performance in fake news detection.

The ablation study results showed that removing features like word sentiment and social network centrality caused performance drops of 10-15%, highlighting the importance of these features in the model. The study showed that sentiment analysis in particular provided better results in detecting fake news compared to structural features alone.

V. Multi-Modal Approach Using Text and Visuals

The survey includes multi-modal methods that use both text and visual data to enhance detection performance:

BERT (Bidirectional Encoder Representations from Transformers) is employed for text, with the embedding equations following the transformer model's self-attention mechanism:

International Journal on Emerging Research Areas (IJERA)

DOI: 10.5281/zenodo.14649896 *ISSN : 2230-9993*

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Here, Q is the query matrix, K is the key matrix, V is the value matrix, and d_k is the dimension of the key vectors.

VGG-19 for image processing uses convolutional layers to extract important features from images. The convolution operation is defined as:

$$f(x, y) = \sum_{i=0}^{k-1}\sum_{j=0}^{k-1} w(i, j) \cdot I(x + i, y + j)$$

Where $I(x+i, y+j)$ is the image input, and $w(i, j)$ is the filter applied over the image. VGG-19 was used to analyze visual features (e.g., images accompanying fake news).

Several of the documents adopted BERT for textual data and VGG-19 or ResNet for image analysis. These documents explored how the combination of text and image data improved model accuracy in cases where the fake news article included misleading images.

1.Text Processing with BERT: BERT's ability to process long-range dependencies in the text and capture context was highlighted as a major breakthrough. Compared to traditional models like SVM and logistic regression, BERT provided superior performance, especially in handling nuanced fake news.

2.Image Processing with VGG-19: Visual data was processed using CNNs like VGG-19, which were effective in detecting manipulated images. The documents noted that in multi-modal settings, combining text and image models resulted in higher precision and recall.

The multi-modal approach was consistently found to be more accurate than text-only models, especially for social media posts that included both misleading text and images. The comparison across documents shows that while BERT alone performed well on text, adding a visual component using VGG-19 boosted overall model performance by 5-10%.

VI. Model Performance Metrics

Various metrics are used to evaluate the model's effectiveness:

Accuracy: The proportion of correctly predicted labels. [3]

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: Measures the model's ability to correctly identify positive samples. [3]

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall: The ability of the model to capture all relevant cases.

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-Score: The harmonic mean of precision and recall.

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Each document insights into which metrics were most important. Precision and recall were consistently emphasized as the most critical metrics for fake news detection due to the high costs associated with false positives and false negatives.

Precision: Precision was especially important in models where the consequences of incorrectly labeling real news as fake were severe. In these cases, models like SVM and BERT were preferred for their higher precision scores.

Recall: Recall was highlighted in cases where detecting all instances of fake news was critical, even if some real news articles were falsely labeled as fake. Random Forest and Neural Networks [9] showed higher recall in some cases.

Across the documents, neural networks (BERT and multi-modal models) had the highest F1-scores [16] , balancing precision and recall. Logistic regression and SVM performed well in terms of precision but had lower recall, making them less ideal in situations where detecting all fake news is the priority.

VII. Dataset for Fake News Detection

The selection of datasets is pivotal for evaluating and enhancing the performance of fake news detection models. Each dataset offers unique characteristics—whether it's a focus on political misinformation, celebrity gossip, or social media interactions—allowing researchers to address different challenges posed by fake news. Below is an in-depth discussion of key datasets, their structure, and examples of their real-world applications.

- LIAR Dataset
- BuzzFeedNews Dataset
- FakeNewsNet Datase
- PolitiFact Datase

A. *LIAR Dataset*

The LIAR dataset includes short political statements, labeled with six possible truthfulness categories: true, mostly true, half true, barely true, false, and pants on fire (extremely false). Each entry includes the statement text, metadata about the speaker, their political affiliation, the context of the statement, and the subject being discussed.

Sample Data:


```
{
    "ID": "2287",
    "Statement": "Says President Obama
       broke his promise to keep
       unemployment under 8 percent."
    "Label": "false",
    "Speaker": "Mitt Romney",
    "Political_Affiliation": "Republican",
```

DOI: *10.5281/zenodo.14649896* ***ISSN : 2230-9993***

```
    "Context": "Campaign Speech",
    "Subject": "Economy",
    "PolitiFact_Rating": "False"
}
```

Application Example:
In research by Wang (2017), this dataset was used to train classification models such as Support Vector Machines (SVM) and Long Short-Term Memory (LSTM) networks. By using the rich metadata, researchers were able to explore how the speaker's political affiliation, subject, and statement content influenced the likelihood of false claims.

Real-World Use:
The LIAR dataset is particularly important for automated fact-checking systems, as it provides a reliable ground truth for machine learning models to learn from. It's been widely used in studies that classify political statements as real or fake, focusing on text-based features such as word embeddings, n-grams, and sentiment analysis.

Impact:
LIAR's detailed metadata allows researchers to build sophisticated models that not only analyze the text of the statement but also consider the political and contextual factors that may influence the truthfulness of the statement.

B. BuzzFeedNews Dataset

The BuzzFeedNews dataset consists of articles labeled as real or fake. It was collected during the 2016 U.S. presidential election and includes metadata such as the article's publication date, source URL, and, in some cases, the number of social shares or comments.

Sample Data:

```
{
      "ID": "105",
      "Title": "FBI Agent Suspected In
         Hillary Email Leaks Found Dead In
          Apparent Murder-Suicide",
      "Content": "An FBI agent believed to
         be responsible for the latest
         email leaks ...",
      "Label": "fake",
      "Source": "Unknown",
      "Date": "2016-11-05",
      "Social_Shares": "10525"
  }
```

Application Example:
In a study by *Allcott Gentzkow (2017)*, the BuzzFeedNews dataset was used to train binary classification models to detect fake news by analyzing the textual content of the articles. The study utilized traditional machine learning techniques like Logistic Regression and Decision Trees, achieving relatively high accuracy for binary classification.

Real-World Use:
This dataset is ideal for fake news research focused on the social and political impacts of misinformation, particularly during elections. It allows researchers to study how false articles spread on social media platforms and influence public opinion.

Impact:
The BuzzFeedNews dataset has played a critical role in political misinformation research, particularly for understanding the spread of fake news during high-stakes political events, such as elections. It is often used in studies focusing on the textual analysis of headlines and article content.

C. FakeNewsNet Dataset:

The FakeNewsNet dataset is multimodal, containing both text-based news articles and social context data. Articles are sourced from PolitiFact and GossipCop and include metadata such as the author, publication date, and associated social media interactions (e.g., retweets, likes).

Sample Data (Text Content):

```
{
    "ID": "857",
    "Title": "Fact Check: Did a Muslim
       Refuse Service to a Woman for
       Wearing a Hijab?",
    "Content": "The article claims that a
        Muslim refused service to a
       woman...",
    "Label": "real",
    "Source": "PolitiFact",
    "Date": "2017-04-10",
    "Social_Shares": "750"
}
```

Sample Data (Social Context):

```
{
    "ID": "857",
    "User": "JohnDoe123",
    "Action": "Retweet",
    "Date": "2017-04-11",
    "Follower_Count": "1500"
}
```

Application Example:
In Shu et al. (2018), the FakeNewsNet dataset was used to develop multimodal models that combine textual and social media interaction data to detect fake news. The study leveraged Graph Neural Networks (GNNs) to model the propagation of news across social networks, improving the accuracy of fake news detection.

Real-World Use:
This dataset is especially useful in studies that aim to detect fake news in a real-world social media context, as it provides both the article content and the social dissemination patterns. It

DOI:10.5281/zenodo.14649896 *ISSN : 2230-9993*

Feature	LIAR	BuzzFeedNews	FakeNewsNet	PolitiFact
News Content				
Labeled Truthfulness	✓		✓	✓
News Articles		✓	✓	
Political Focus	✓		✓	✓
Social Context				
Social Media Interaction		✓	✓	
Multimodal (Text + Social)			✓	
Additional Features				
Fact-Checked	✓		✓	✓
Real/Fake Labeling		✓	✓	

TABLE I
COMPARISON OF FAKE NEWS DETECTION DATASETS

allows researchers to investigate not only whether an article is fake, but also how it spreads through social networks.

Impact:

By including social context data, FakeNewsNet enables the development of more comprehensive fake news detection models that can detect misleading content based on both the text and its social propagation. This is essential for addressing misinformation that spreads rapidly across platforms like Facebook and Twitter.

D. *PolitiFact Dataset*

The PolitiFact dataset contains fact-checked statements with metadata such as the speaker, the context in which the statement was made, and the PolitiFact ruling. The dataset is often used in combination with other datasets like FakeNewsNet or LIAR.

Sample Data:

```
{
      "ID": "10987",
      "Statement": "The unemployment rate
         is at an all-time low.",
      "Speaker": "Donald Trump",
      "Label": "half true",
      "Source": "PolitiFact",
      "Date": "2019-01-21",
      "PolitiFact_Rating": "Half True"
  }
```

Application Example:

In Vlachos Riedel (2014), the PolitiFact dataset was used to train models that predict the truthfulness of political statements based on the speaker's previous statements, political affiliation, and the context. The study found that models trained on PolitiFact data could achieve high accuracy in detecting false claims.

Real-World Use:

This dataset is valuable for researchers studying political misinformation and developing automated fact-checking tools. Its credibility as a fact-checking source makes it a trusted dataset for detecting false political claims.

Impact:

The PolitiFact dataset provides high-quality, human-verified labels, making it ideal for building reliable and transparent fact-checking models. Its detailed metadata allows researchers to explore complex relationships between the context of a statement and its truthfulness.

VIII. OVERALL COMPARATIVE ANALYSIS

- **Image-Based Forgery Detection:** These papers excel in detecting image forgeries through CNN-based models, providing high precision for identifying manipulated images. They are especially effective for detecting copy-move forgery in scenarios where images are rotated, scaled, or compressed, making them ideal for identifying visual tampering in fake news.
- **Text and Social Context-Based Fake News Detection: [3]** These papers excels in combining content analysis (text and images) with social media engagement to detect fake news. It takes a holistic view, addressing the propagation of fake news through social media networks. This method is critical for addressing how fake news spreads beyond simple content analysis.
  Papers with Multi-Feature Classification [12] and Machine Learning offer solid text-based detection approaches. They focus on textual content analysis,incorporating temporal and spatial features to enhance classification. They are effective for detecting misleading text but do not focus on social or visual content, which limits their applicability in dynamic social media environments.

A. *Best Approach for Fake News Detection:*

For an effective fake news detection system, the ideal approach would combine the strengths of both image forgery detection and text-based social engagement analysis. The CNN + CenSurE models [12] are best suited for detecting manipulated images. The data mining approach from the excels in capturing the spread of fake news via user interactions and social context.

Paper No.	Model/Algorithm	Accuracy	Precision	Recall	F1 Score	Dataset Used	Unique Feature
1	Multimodal Framework (MEFaND)	90%	88%	89%	88.5%	Twitter, BuzzFeed	Combines text, image, and context features
2	SVM, Naive Bayes	85%	80%	82%	81%	LIAR	Simple traditional ML models
3	CNN + LSTM	92%	91%	90%	90.5%	Politifact	Uses both text and temporal features
4	CNN-Keypoint Hybrid	88%	87%	86%	86.5%	Custom image dataset	Focuses on image forgery detection
5	Random Forest, Naive Bayes	83%	82%	80%	81%	LIAR	Traditional ML-based approach

Fig. 4. Comparison of Models.

Combining visual forgery detection with text and social analysis will provide a robust, multimodal system that can handle both image tampering and textual/social misinformation.

Fig4 compares the performance of different machine learning models for fake news detection. The models are evaluated on accuracy, precision, recall, and F1 score. The best performing model is the Multimodal Framework (MEFAND), which achieves an accuracy of 90%, precision of 88%, recall of 89%, and F1 score of 88.5%. This model combines text, image, and context features to detect fake news. Other strong performing models include CNN + LSTM and CNN-Keypoint Hybrid [12] , which both achieve an F1 score of over 90%.

The best solution is a hybrid model that integrates the strengths of Image forgery detection (using CNN-based models) to detect manipulated images.Social media-based fake news detection that accounts for how fake news spreads through user engagement and social behaviors.

IX. Conclusion

This survey reviewed various approaches to fake news detection and image forgery detection, ranging from traditional machine learning models to advanced multimodal frameworks like MEFaND. While models like Random Forest and Logistic Regression perform well in text-based fake news detection, more recent approaches, such as transformer-based models and multimodal systems, offer superior performance by incorporating social context and propagation patterns.

MEFaND stands out as the most effective solution, combining BERT for textual analysis with Graph Neural Networks (GNNs) for analyzing early-stage news propagation. This hybrid approach excels in detecting fake news within the first few hours of its spread, providing a powerful tool for real-time detection.

For the best results, a hybrid detection system combining CNN-based image forgery detection and social propagation analysis is recommended, ensuring comprehensive coverage of both visual tampering and textual misinformation in today's fast-paced media landscape.

References

[1] MEFaND: A Multimodel Framework for Early Fake News Detection,Journal: IEEE Transactions on Computational Social Systems, August 2024, Vol. 11, No. 4,Authors:Asma Sormeily, Sajjad Dadkhah, Xichen Zhang, Ali A. Ghorbani

[2] Detecting Fake News in Social Media Networks,Journal: Procedia Computer Science 141 (2018) 215–222,Science Direct,Authors: Monther Aldwairi, Ali Alwahedi

[3] Fake News Detection on Social Media: A Data Mining Perspective,Review Paper:Data Mining Perspective from Arizona State University,Authors:Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, Huan Liu

[4] CNN-Keypoint Based Two-Stage Hybrid Approach for Copy-Move Forgery Detection,Research Article:IEEE Access, DOI: 10.1109/ACCESS.2024.3380460,Authors: Anjali Diwan and Anil K. Roy

[5] Fake News Detection using Machine Learning Algorithms,Journal: International Journal of Engineering Research Technology (IJERT),Authors: Uma Sharma, Sidarth Saran, Shankar M. Patil,Year: 2020

[6] Efficient Fake News Detection Mechanism Using Enhanced Deep Learning Model,Journal: Applied Sciences,Author: Tahir Ahmad, Muhammad Shahzad Faisal, Atif Rizwan, Reem Alkanhel, Prince Waqas Khan and Ammar Muthanna,Year: 2022

[7] Constructing a User-Centered Fake News Detection Model by Using Classification Algorithms in Machine Learning Techniques,Journal: IEEE Access ,Author: Minjung Park and Sangmi Chai,Year: 2023

[8] Machine Learning-Based Approach for Fake News Detection,Journal: Journal of ICT Standardization ,Author:H. L. Gururaj, H. Lakshmi, B. C. Soundarya, Francesco Flammini and V. Janhavi,Year: 2022

[9] Robust Image Forgery Detection Against Transmission Over Online Social Networks,Journel: IEEE Transactions On Information Forensics and Security, Vol. 17, 2022,Authors: Haiwei Wu ,Jiantao Zhou ,Jinyu Tian ,Jun Liu ,and Yu Qiao

[10] An Enhanced Fake News Detection System With Fuzzy Deep Learning,Journal: IEEE Access Vol: 12 ,24 June 2024,Author: Cheng Xu and M-Tahar Kechadi

[11] Multimodal Fake News Analysis Based on Image–Text Similarity,Journal: IEEE Transactions On Computational Social Systems, Vol. 11, No. 1, Feb 2024,Author:Xichen Zhang , Sajjad Dadkhah , Alexander Gerald Weismann, Mohammad Amin Kanaani, and Ali A. Ghorbani

[12] Enhancing Fake News Detection by Multi-Feature Classification,Journal: IEEE Access Vol: 11 ,29 November 2023,Author:Ahmed Hashim Jawad Almarashy, Mohammad-Reza Feizi-Derakhshi and Pedram Salehpour

[13] Advancing Fake News Detection: Hybrid Deep Learning With FastText and Explainable AI,Journal: IEEE Access,Authors:Ehtesham Hashmi, Sule Yildirim Yayilgan, Muhammad Mudassar Yamin, Subhan Ali, Mohamed Abomhara ,Year: 2024

[14] Knowledge Enhanced Vision and Language Model for Multi-Modal Fake News Detection,Journal: IEEE Transactions on Multimedia,Authors: Xingyu Gao, Xi Wang, Zhenyu Chen, Wei Zhou, Steven C. H. Hoi,Year: 2024

[15] Fake news detection based on news content and social contexts: a transformer-based approach,Journal: International Journal of Data Science and Analytics,Authors: Shaina Raza and Chen Ding Year: 2022

[16] Fake News Detection: A Deep Learning Approach,Journal: SMU Data Science Review,Authors: Aswini Thota, Priyanka Tilak, Simrat Ahluwalia, and Nibrat Lohia,Year: 2018

[17] State-of-the-Art Techniques for Image Forgery Detection: A Review," International Journal on Emerging Research Areas (ISSN:2230-9993), vol.03, issue 01, 2023 doi: 10.5281/zenodo.801788211,Anishamol Abraham, Niya Joseph

[18] A Machine Learning Approach to Fake News Detection,",10. Ansamol Varghese, Ignatius Ealias Roy, Anoushka Tresa, Athira John, Gautham Sankar M S, International Journal on Emerging Research Areas (ISSN:2230-9993), vol.03, issue 01, 2023 doi: 10.5281/zenodo.8019338