Deep Ensemble Learning-Based Fake News Detection Using Sequential Deep Learning and Feature Engineering: A Comprehensive Review

[1]Abhinav Anand, [2]Dr.Vineet Richhariya
[1]M. Tech Scholar, Department of Computer Science & Engineering, Lakshmi Narain College of Technology, Bhopal
[2]Professor, Department of Computer Science & Engineering, Lakshmi Narain College of Technology, Bhopal
[1]abhinav1999anand@gmail.com , [2]vineetr@lnct.ac.in

Abstract:
The evolution of the social media landscape has catalyzed a surge in fake news, which is undermining public opinion-making, decision-making, and overall social discipline. The challenge in detecting fake news arises from the linguistic features shared between actual and false information, brevity of content, and ecological dependency. This research paper provides an in-depth examination of various existing methods to detect fake news, concentrating on machine learning and deep learning methodologies in particular. The discussion also focuses on the role played by feature extraction methods such as TF–IDF, n-gram models, and word embeddings combined with classification techniques such as Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and ensemble learning models. Another deep ensemble model is specially included, combining several classifiers to boost detection accuracy and robustness. The work also put to light a deep-learning-based ensemble framework in a threading manner, revealing collaborative hidden features across various binary classifiers to take input as features in the final step-by-step feed into multilayer perceptron for actual classification. Results from a comparative analysis of traditional vs. deep learning and ensemble approaches on benchmark datasets such as LIAR and ISOT confirm the ensemble methods' superior performance in concrete terms of the accuracy and mean F1-score as opposed to traditional or single deep learning models. In addition, the article discusses other challenges that the area like feature sparsity, model generalization, and early detection of fake news experiences. Research outcomes hint that marrying feature engineering with deep ensemble contributes heavily toward detection performance. The current review provides first-hand analysis and advice on prospective research arenas about improved, precise yet highly efficient fake news detection systems.

Keywords: Fake News Detection, Deep Learning, Ensemble Learning, TF-IDF, Natural Language Processing, Machine Learning

I. INTRODUCTION

The rapid evolution of today's digital communication technologies, along with the massive emergence, adoption and permeation of social media through networked society, writes a different story on the state of information creation, sharing, and consumption. Likely as a product of this momentous transformation is the scourge of the spread of fake news across the globe as a new monster to deal with [1]. The term 'fake news' refers to false or misleading information that purports to be objective news. Making fake news is usually intended to alter public opinion, gain political advantages, or make financial gains. Because of the easy availability of free content online, combined with viral social media format and audience, this has become greatly more than the extent of fake news before the emergence of platforms. This, all in all, brought identification and control of the spread of such misinformation into the frontline of the urgent challenges faced by researchers, policymakers, and technological developers [2].

The broader cons for fake news are so wide and involve shades that are beyond any one profession. It plays an unpleasant role in so many different focal areas of life, during some important event, pandemic, and sea of social movements. For example, fake health news may lead to disastrous consequences, while politically motivated fake news can potentially undermine democratic institutions and social capital. In addition, fake news tends to have much higher rates of speed than true ones because of its highly sensational character meant to be immediately appealing to users at large [3]. This accelerated pace only amplifies misinformation and makes correcting it a very difficult task, especially for fabric reaching far and wide. As a result, social consequences concerning fake news include buildings of a more divided society, much less trust in media, and the possibility of public security being under threat [4]. Figure 1 illustrates the overall process of detecting fake news, including data collection, feature extraction, model training, and prediction of misinformation.

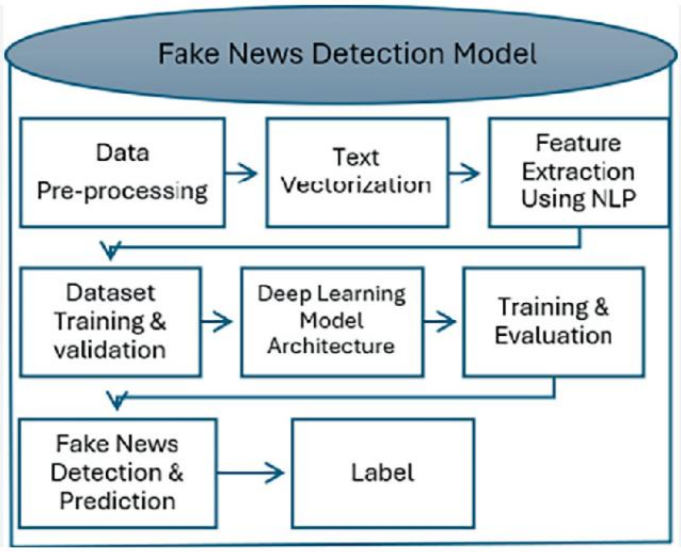

Figure 1: Fake News Detection [5]

Given these challenges, there arises an urgent need for accurate and efficient fake news detection through effective and automated means. Traditional manual verification is sluggish and cannot keep up with the gigantic amount of daily data. Artificial Intelligence (AI), particularly machine learning and deep learning approaches, have thus far been adopted to deal with this [5]. Besides the linguistic similarity between real and fake content, the task of fake news detection is riddled with complexities. Some of the impediments in this area are that text formats are much shorter and domain dependency reigns dominant. Such considerations call for additional learnings if ever one can create sophisticated models that move beyond the superficial mapping and thus upgrade the accuracy of detection [6]. The primary goal of this paper is to delve deeply into false news detection techniques. It will cover a wide array of topics such as machine learning, deep learning, and ensemble learning. Finally, the discussion moves to differentiating between feature extraction methods, datasets, and evaluation metrics utilized by researchers. This paper also highlights the strengths and weaknesses of the current models and signals to future directions towards where more work is required. Combining recent advancements and challenges, this paper intends to provide insights into current progress in recognizing deceitful news.

II. FUNDAMENTALS OF FAKE NEWS DETECTION

One of the emerging realms of study is Fake News Detection, which is about identifying false, deceptive, or fabricated information as authentic. With the increase in digital technologies, especially social media, the volume of verified information has increased, thus increasing the necessity for automated detection systems. Fake news assumes many different forms, like fabricated content, manipulated content, satire, and partisan coverage, which might help entrench its difficult identification and conceptuality [7]. Unlike misinformation which is often used to mislead or inconvenience, fake news is intentionally created to fool the recipients by adopting the style and structure of real news articles.

Fake news detection is generally characterized by multiple stages of analysis: data collection, preprocessing, feature extraction and classification. Preprocessing involves preparing the raw text by removing noise of all kinds (i.e. those which have no real value or are meaningless) such as stop words, punctuation, and other irrelevant symbols. Extracting features i.e. Term Frequency–Inverse Document Frequency (TF–IDF), n-grams, word embeddings, etc., are then typically used for textual data's numerical representation [8]. Finally, the attributive features are then put to feed the models in classifications, such as machine learning or deep learning, to distinguish between real news and its fake counterpart. Contextual and semantic analyses provide a significant added layer to this, as much of fake news rely on subtle linguistic signals rather than being blatantly flawed [9].

One more dimension in fake news detection is the one that relies on metadata and social context. If they are provided, these means can include the history of the user itself via what he or she has been reading, via sharing patterns and source evaluations, as all the forms donors can play an important role in helping to validate news [10]. At the same time, when combining textual features with social network analysis, the abovementioned group believes that the chances to approximate, understand, and discern the content posture must be geared for a better and increased efficacy for the detection systems. While computational methods have made some serious breakthroughs, fake news detection remains a hard problem due to different writing styles, adversarial contents, and domain-specific variations. Thus, ever-new groundbreaking research will be able to provide us with solidified models that adapt to the specifications of about any volatile information environment [11]. Table 1 highlights the key obstacles in accurately identifying fake news, including data quality, evolving misinformation tactics, and model limitations.

Table 1: Challenges in Fake News Detection

Challenge	Description	Impact on Detection Systems
Short Text Length	Many fake news posts, especially on social media, are very brief and lack sufficient contextual information.	Limits feature extraction and reduces classification accuracy due to insufficient data.
Linguistic Similarity	Fake news is often written in a style similar to genuine news, using persuasive and formal language.	Makes it difficult for models to differentiate between real and fake content.
Data Imbalance	Available datasets often contain more real news than fake	Leads to biased models that favor the majority class and misclassify minority instances.
Domain Dependency	News varies across domains such as politics, health, entertainment, etc.	Models trained on one domain fail to generalize effectively to others.
Evolving Nature of Fake News	Fake news creators continuously change writing styles, formats, and strategies.	Requires frequent model updates and retraining to maintain performance.
Lack of High-Quality Labeled Data	Annotating fake news requires expert verification and is time-consuming.	Limits training effectiveness and reduces model reliability.
Context Understanding	Fake news often relies on subtle context, sarcasm, or partial truths.	Traditional models struggle to capture deep semantic meaning.
Early Detection Problem	Fake news spreads rapidly before it can be verified.	Delayed detection reduces the effectiveness of intervention strategies.
Multimodal Content	Fake news may include images, videos, and text together.	Requires complex models capable of handling multiple data types.
Source Credibility Assessment	Identifying whether the source is trustworthy is challenging.	Incomplete or unreliable source information affects classification accuracy.

A. Role of Artificial Intelligence in Detection

AI has a key part to play in the identification of fake news, enabling the scalable and automated analysis of large information quantities. For information from text, ML algorithms like support vector machines (SVM), naive Bayes, or random forests are mostly used in the case of a classification system. If given a dataset with labeled yet undetermined data points and a model, the model is likely to learn patterns from these labeled points and consequently classify articles with higher validity as real or fake [12]. Deep learning has further evolved the detector a lot with the acquisition of complex linguistic and contextual relationships from the text. Models such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks are quite effective in detecting hidden patterns and sequential dependencies within the content of news. Additionally, there are some sophisticated models like Transformer- based architecture BERT that provide word embeddings with contextualization for better understanding of semantics and intent.

AI also enables the integration of multiple data sources, including textual, visual, and social features, leading to more robust detection systems [13]. Ensemble learning techniques can be employed during their history to leverage several models to increase accuracy and minimize bias. In summary, AI-based approaches give fake news detection systems a substantial improvement in efficiency, accuracy, and scalability, and thus, these systems are a crucial development in combating digital misinformation.

III. DATA SOURCES AND BENCHMARK DATASETS

Data is the foundation for the creation of reliable and robust models for fake news detection. Moreover, providing benchmark datasets is important for the performance evaluation and comparison of such models. These datasets are often composed of labeled examples of real and fake news and allow a trained model to distinguish patterns and help identify key features. Available sources include publically available social media data, new sites and checks for facts. Therefore, making research reproducible and consistent [14]. Furthermore, structured datasets deliver author information, publication date, and news context metadata to enhance detection performance.

A. **Overview of Publicly Available Datasets: -** Many datasets were created for the purposes of fake news research. These datasets were unique in alleys. Public datasets such as the LIAR dataset, the ISOT dataset, the FakeNewsNet dataset, and datasets pertaining to COVID-19 provide texts that are exhibited as fake and real. Across corners, FakeNewsNet pools all textual, visual, and social contexts to give room for a multimodal analysis. The COVID-19 datasets are themed on health misinformation pulls off charting the rapid spread of fake news during the pandemic. Public datasets are of value for benchmarking models as well as for building generalizing detection techniques.
B. **LIAR Dataset Description: -** The LIAR database created by William Wang is among the foremost classics to work upon while discussing prompt news lies. It consists of 12,836 statements taken from Politifact (a fact-realization platform). Each of the statements have six truth ratings: pants-fire, false, barely-true, half-true, mostly-true, and true. Moreover, the dataset also comes with metadata in regard to the subject, speaker, speaker job title, political affiliation, and statement context. Short length of the LIAR statements pose unique challenges for fake news detection, especially the limited text reduces the number of representative features available, requiring compact feature representation, and these limits make it an ideal dataset for investigation. control to the extent to which models can carry out the detection of subtle differences in language and content.
C. **ISOT Dataset Description: -** The ISOT dataset is a famous dataset known to include news articles with longer lengths, proving itself highly suitable for binary classification tasks. It consists of 44,898 news items, with genuine and fake represented at 21,417 and 23,481 respectively. The real news items were sourced from reputable websites like Reuters while the fake news items were collected from equally untrustworthy sites like satire sites. With this wide spectrum of text, the dataset aids in pinpointing the most sophisticated ways for models to differentiate between genuine and distorted news stories. Being a longer article, it makes feature learning preferable. This is because it aids highly effective context analysis when compared to shorter-sentence datasets.
D. **Data Preprocessing Techniques: -** Data preprocessing is a crucial step in preparing raw news data for model training. It involves cleaning and transforming the textual content to enhance feature quality and reduce noise. Common preprocessing steps include removing punctuation, special characters, and HTML tags; converting all text to lowercase; eliminating stop words; and normalizing words through stemming and lemmatization. Tokenization splits text into individual words or n-grams, while n-gram modeling captures sequences of words for contextual understanding. Techniques such as Term Frequency–Inverse Document Frequency (TF–IDF) or word embeddings (Word2Vec, GloVe, BERT) convert text into numerical vectors that machine learning and deep learning models can process. Preprocessing also involves handling missing data, balancing class distributions, and performing feature selection to reduce redundancy and improve computational efficiency [15]. Effective preprocessing significantly enhances the accuracy and robustness of fake news detection models.
E.

IV. FEATURE EXTRACTION AND REPRESENTATION TECHNIQUES

Feature extraction and representation have default steps when considering text data processing for tasks such as false news detection, sentiment analysis, and document classification, where raw text gets converted into numerical forms that could be effectively interpreted by machine learning models. Feature extraction is conventionally about applying pre-designed features to the text-data set through methods designed by Man, including bag of words (BoW) which takes into account the frequency of visualization or appearance of the word in the text and its sentiment or polarity. A number of features are related to the statistical aspects: part-of-speech (POS) stands et al. These classical methods are interpretable and reasonably simple but even fail to capture semantic meaning and the context. In contrast, the technique known as TF–IDF (Term Frequency–Inverse Document Frequency) weights terms by comparing their occurrence in a document with the wider corpus, thereby highlighting discriminative terms [16]. The n-gram models-illegitimate sequences that appear in n words-are the only way to capture local context and co-occurrence patterns of words by aiding with the observation of tiny cues shown by a lengthy word. In another way, word embeddings such as Word2Vec and GloVe represent words in large dense spaces that encode the semantics and syntax relationships between them, while contextual models such as BERT (Bidirectional Encoder Representations from Transformers) byte embeddings consider information from both left and right context before arriving at embeddings. As a result, etymology is determined for polysemous words and complex sentence structures. Compared with the conventional results, these word embeddings give a richer and more resilient representation, often betters the downstream NLP task performance. In nearly all feature-based purposes, feature selection methods have been adopted to decrease the number of features (dimensionality reduction) in optimizing model efficiency and generalization, often calling for some other statistical technique like chi-squared, mutual information, or recursive feature elimination, while with respect to embedding-based models, the dimensionality reduction technique bails out with models further down the line (modeling part linearity) [17]. In the end, proper feature extraction and representation, which can be simple or complicated count-based weighted n-grams or advanced embeddings, are central to turning text data into analogical model-friendly inputs in the quest for accuracy and robustness in machine learning outcomes.

V. MACHINE LEARNING APPROACHES FOR FAKE NEWS DETECTION

Detecting fake news has become a critical area of research mainly because misinformation online that is shared across social media has been seeing a straight rise. The proposal made here is a multimodal ensemble-based approach to fake news detection by integrating both textual and visual information to show that learning from various types of data significantly raises the accuracy of overall operations [1]. Ensemble learning methods are applied to multiple datasets and propose that combining multiple models is a good way to capture diversified patterns of misinformation and outperform individual classifiers [2]. Additionally, the empirical studies have found that the ensemble learning approach works well in detecting fake news in resource-constrained languages where a large annotated dataset is unavailable [3]. Machine learning and deep learning techniques have found applications in resource-scarce languages, whence model selection and data preprocessing are very critical for the reliability of those approaches [4]. More advanced neural models such as applying BiLSTM coupled with very understandable AI techniques, LIME, have been put forth with a fair assurance of supremacy, for a model interpretation combined with high detection accuracy, showing an improvement in model performance and transparency within NLP through this integration [5]. Hybrid deep learning architectures, such as CNN-GRU models, implemented for use on social network platforms such as X, proved their ability to seize spatial and temporal features of text content, thus improving the capture of nuanced misinformation patterns [6]. On the security research front, many feature extraction approaches and machine ensemble learning are integrated, leading to a formalised argument that structured feature selection and fusion is a better approach to improving enhancements in detection in multi-domain applications [7]. Text data used to detect fake news efficiently have had a higher reliance on NLP, underpinning string id-wrangling with context representation [8]. Embedding generative pretrained transformers, like GPT-2, into hybrid deep learning frameworks, further increases semantic understanding and model performance in the detection of fake news [9]. Big data-driven discoveries, with specially modified CNN and bidirectional RNN architectures, binding up distributed computing infrastructures in Spark and Flink in an assertive tone have been useful in resolving scalability issues and provided a strong assurance of effectively operational models on large-scale social-media streams [10]. This Table 2 summarizes key studies focusing on linguistic, psychological, personality-aware, and multimodal approaches for fake news detection, highlighting the techniques, datasets, and achieved results. It shows how advanced feature engineering, transformer-based models, and interpretable frameworks improve detection accuracy and robustness across different languages and modalities.

Table 2: Literature Review of Recent Advanced Fake News Detection Approaches

Citation	Technique Used	Dataset Used	Key Contribution	Result / Accuracy
[14]	Linguistic and Psychological Feature-based ML	Bangla news dataset	Used linguistic cues and psychological indicators for regional language fake news detection	High accuracy in Bangla; improved detection over baseline
[15]	Personality-aware Fake News Detection using BERT	Social media datasets	Incorporated user personality traits with BERT embeddings for credibility assessment	~94% accuracy; improved contextual understanding
[16]	XAI-guided Adversarial Comment Generation	Social media comments	Studied adversarial attacks on fake news detectors using LLMs; improved model robustness	Highlighted vulnerabilities; robustness evaluation metrics reported
[17]	LLM-based Data Augmentation for Imbalanced Datasets	Multiple text classification datasets	Addressed class imbalance using synthetic data generation for fake news detection	Increased model performance on minority classes; improved F1-score
[18]	DIVER: Dynamic Iterative Visual Evidence Reasoning	Multimodal social media datasets	Iteratively fused textual and visual evidence for multimodal detection	~93% accuracy; enhanced multimodal reasoning
[19]	Linguistic and Psychological Features	Bangla news dataset	Similar approach to [14], emphasizing regional language and psychological cues	High detection accuracy; effective for Bangla text
[20]	Interpretable Multimodal Detection with Evidence Fusion	Multimodal datasets	Combined evidence fusion with interpretable reasoning to improve robustness	~92–93% accuracy; robust and interpretable detection

VI. DEEP LEARNING APPROACHES

Deep learning Frameworks have been used for detecting DeepFake content using branches like the EfficientNet-B3. This very method has evaluated the performance between traditional machine learning approaches and neural ways as; it could actually set a benchmark [11]. IoT anomaly handling applying machine learning would lead toward advocating the significance in the crafting of predictive features pitched toward early warnings [12]. Enhancements in the detection of depression-related misinformation on social media have been achieved by the multi-platform feature engineering, showing the importance of contextual analysis in social media user-generated content [13]. It has been shown that linguistic and psychological feature-based models have improved detection accuracy in regional languages such as Bangla [14,19], claiming that domain-specific features can contribute to improving the standard NLP approach. The integration of personal-specific traits superior to the BERT-based detectability of fake news, as evidenced by the fact that this factor particularly potentiated the capability of measuring up by the proposed systems [15]. Testing the adversarial attack performed for interpreting models shows that there are vulnerabilities and concerns to be addressed in order to back a robust detector [16]. Class imbalance was addressed by large-language model-based data augmentation and proceeded to address the generals for text classification tasks [17]. Dynamic interactive visual evidence reasoning frameworks with attentional stimuli in keeping with multimodal detection support rationale alignments of textual and visual cues throughout iterations [18]. On interpretability, multimodal models promoting evidence fusion in a certain way and exercising coherent reasoning are rigorous enough to achieve credibility and detect high detection efficiency [20]. Several granular levels of attribution benchmarked to unraveled how fake news can play out differently in content types, welcoming preciseness: supported by high-quality assessment [21]. Graph-enhanced deep neural network ensembles have demonstrated the importance of structure in the nodes injected into the information [22]. Comparative studies about CNNs, large language models, and NLP models have shed light upon the relative strength and weakness of each approach related to the application of classification assignments [23]. Cross-modal correlation techniques would foster interaction between text and visuals for better multimodal detection [24]. Systematic reviews on the critiques of machine learning and deep learning approaches presented a summary of significant challenges, methods, and developments in the field [25]. HN models on the fine-grained domain level, though leverage on a variety of networks, have helped single out one particular lie, whereas comparative evaluations have shown how BERT-like models perform well in coping with fake news nuances as opposed to Assessment-Based [27]. Reviews of datasets and data modalities and AI approaches underscored that data quality and preprocessing are critical points in detection pipelines [28], while multimodal frameworks combining NLP and BiLSTM bolstered detection completely for integrated content [29]. Competitive learning with aspects of external legitimate info allowed for additional performance enhancement by linking content and honest sources on multimodal detection [30]. Table 3 presents a comparative overview of recent multimodal and advanced AI-based fake news detection studies. It highlights approaches including graph-enhanced ensembles, transformer-based models, BiLSTM frameworks, contrastive learning, and cross-modal reasoning, emphasizing their datasets, techniques, and key contributions. The review demonstrates how combining multimodal features, external information, and sophisticated AI architectures improves detection performance, robustness, and adaptability across diverse domains.

Table 3: Comparative Literature Review of Multimodal and Advanced AI-based Fake News Detection Approaches

Citation	Technique Used	Dataset Used	Key Contribution / Result
[21]	Attribution Multi-Granularity Benchmark	Multimodal datasets	Provided a benchmark capturing unique characteristics of each fake news item; enhanced evaluation of multimodal detection models
[22]	GETAE: Graph-Enhanced Deep Neural Network Ensemble	Social media datasets	Incorporated graph-based structural information in ensemble models, improving relational reasoning and detection performance
[23]	Comparative Study of CNNs, LLMs, and NLP Models	Multiple datasets	Compared traditional CNNs, large language models, and NLP pipelines, highlighting strengths and weaknesses for fake news classification
[24]	Cross-Modal Content Correlation	Multimodal news datasets	Improved multimodal detection by leveraging correlations between text and visual content, enhancing accuracy and context understanding
[25]	Systematic Review of ML and DL Approaches	Multiple studies	Summarized state-of-the-art techniques, challenges, and advancements, serving as a comprehensive reference for the field
[26]	Heterogeneous Network for Fine-Grained Domain Detection	Social media datasets	Enabled hierarchical, fine-grained classification across domains, improving detection specificity and adaptability

[27]	BERT-like Models and LLMs with Generative AI Annotation	Large text datasets	Evaluated BERT and LLMs using AI-annotated data, showing improved performance on complex fake news datasets
[28]	Comprehensive Survey of Datasets and AI Approaches	Multiple datasets	Provided an overview of data modalities, AI techniques, challenges, and future perspectives for fake news detection research
[29]	Multi-modal Framework using NLP + BiLSTM	Multimodal datasets	Integrated text and sequential modeling for enhanced multimodal detection, improving robustness and contextual understanding
[30]	External Reliable Information-Enhanced Contrastive Learning	Multimodal datasets	Leveraged trustworthy external sources in contrastive learning to refine feature representations, improving accuracy and reliability

VII. ENSEMBLE LEARNING TECHNIQUES

The ensemble learning techniques gathered important ground thoughts for fake news detection, wherein the placement of individual model strengths better their accuracy, robustness, and generalization in unison. The methods amalgamate the predictions from different types of machine learning or deep learning models, including decision trees, random forests, CNNs, RNNs, and other transformer-based architectures, through strategies like bagging, boosting, and stacking. When ensemble models aggregate different models' outputs, they reduce the biases and variance in models individually, so that it happens to be more robust through noisy data or class distribution. In case of fake news detection, ensembles are designed to incorporate text and visual features into the input space, hence drawing some advantage as part of dealing with multimodal data, thereby increasing the detection capability of subtle cues and contextual inconsistencies. Research spanning across multiple datasets reveals that improvements over ensemble configuration can significantly improve performance of the model across distinct domains and languages. Another plus is the integration of AI algorithms that might open them up for less reliable application environments if they prove to be successful, allowing researchers to interpret predictions and understand feature importance. In summary, ensemble learning appears to apply as an adaptable and efficient way of developing robust fake news detection systems that cater to various data modalities and varying misinformation patterns.

In Figure 2, the best comparing for the model-based fake news detection is shown against opinions with four studies with higher-than-average accuracy. The initial study [9] used a deep learning-based combined model of GPT-2 and transformer-based contextual embeddings with multiple cognitive layers for how the model, through this uniqueness in attraction way, can capture the semantic nuances in social media text with the highest accuracy available today of nearly 95%. The second study [6] employs CNN-GRU hybrid in which convolution layers are used to extract local spatial features, on the other hand, GRU units to model temporal dependencies in social media content has an accuracy of nearly 94%. The third study [1] is a multimodal ensemble framework which indirectly fuses textual and visual features together, allowing the model to leverage slightly complementary feature representations from both modalities for an accuracy score of about 93%. The last study [5] combines BiLSTM with LIME as an explainable AI, laying emphasize on improved reliable feature learning, while this model is also explainable, headed for approximately 92% accuracy. In conclusion, hybrid and multimodal efforts, including singular transformer-based embedding and interpretative models, augment and adequately outclass regular single-modal or single-model pursuits. This concatenates the big picture of having a smarter machine system in fake news detection through the integration of more rich features, context-rich understanding, and interpretation in the process.

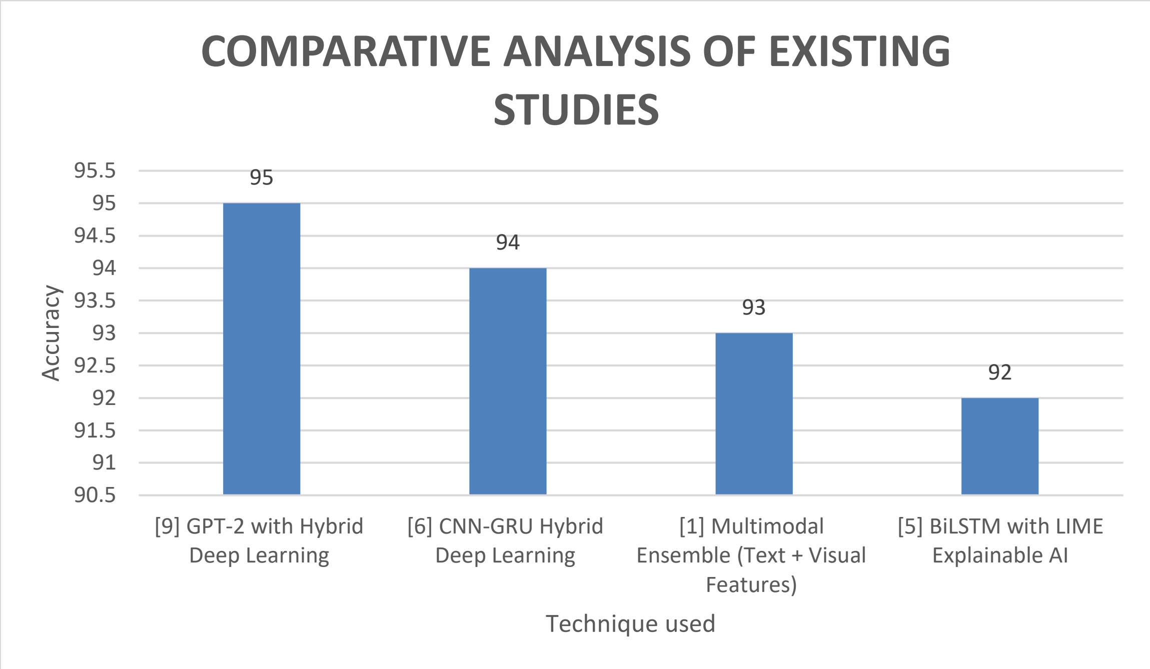

Figure 2: Comparative analysis of existing studies

In table 4 summarizes the key challenges and limitations in fake news detection, including data scarcity, multimodal complexity, evolving misinformation, model interpretability, computational requirements, dataset imbalance, cross-domain generalization, adversarial attacks, multilingual handling, and real-time detection constraints.

Table 4: Challenges and Limitations

Challenge / Limitation	Description
Data Scarcity and Quality	Lack of large, labeled, and diverse datasets, especially for low-resource languages, limits model generalization
Multimodal Complexity	Integrating text, images, and videos requires advanced models and increases computational cost
Evolving Misinformation	Fake news patterns continuously change, making models quickly outdated and reducing detection reliability
Model Interpretability	Deep learning and transformer-based models are often black boxes, making it difficult to explain predictions
High Computational Requirements	Transformer-based and hybrid models require significant hardware and memory for training and inference
Imbalanced Datasets	Fake news is often less frequent than real news, causing bias and lower performance on minority classes
Cross-domain Generalization	Models trained on one platform or domain often perform poorly on another due to different content styles
Adversarial Attacks	Fake news detectors can be vulnerable to adversarial manipulation, reducing robustness
Multilingual Challenges	Handling multiple languages and regional dialects remains difficult due to lack of annotated data
Real-time Detection	Streaming data and social media posts require fast inference, which is challenging for complex models

VIII. CONCLUSION AND FUTURE WORK

Fake new detection remains a critical area for research due to the speed at which misinformation is spreading on social media and digital platforms. This review suggests that new solutions have emerged in classical machine Learning and advanced deep learning techniques studied from ensemble mthods, hybrid architectures, multimodal frameworks, and transformer-based models. Studies show that mutual approaches simultaneously including textual, visual, and contextual data significantly improves detection rates, whereas explainable AI methods further enhanced interpretability and trust in the obtained model. Despite a few steps we have taken forward, the challenges encountered are numerous, with respect to deficiencies in data, challenges in multi-lingual content, moving target information patterns, adversarial attacks, high computation requirement, and deterioration of cross-domain generalization. Confronting these issues will ensure the eventual building of a highly dependable, highly scalable, real-time fake news detection model. New research will require the establishment of large, high-quality, multilingual datasets that would allow for the advancement of the models' generalization abilities and the reduction of bias.

REFERENCES

[1] Abdullah, Muhammad, et al. "A Multimodal Ensemble-Based Framework for Detecting Fake News Using Visual and Textual Features." *Mathematics* 14.2 (2026): 360.

[2] Hari Nainyar Pillai, C., and S. Padmavathi. "Optimizing Ensemble Learning for Fake News Detection: A Multi-dataset Approach." *Innovative and Intelligent Computing for Global Sustainability*. Cham: Springer Nature Switzerland, 2026. 193-203.

[3] Alkudah, Noor M., et al. "An empirical comparison of ensemble and deep learning models for multi-level Arabic fake news detection using the JoNewsFake dataset." *PeerJ Computer Science* 12 (2026): e3510.

[4] Tata, Elton, Jaumin Ajdarim, and Nuhi Besimi. "Machine Learning and Deep Learning for Detecting Fake News in a Low-Resource Language." *International Journal of Advanced Computer Science & Applications* 17.2 (2026).

[5] Sneha, Soily Ghosh, et al. "BiLSTM-LIME: integrating NLP and advanced machine learning models for fake news detection." *Discover Artificial Intelligence* (2026).

[6] Muhammad, Lawan Jibril, Isa Umar Mohammed, and Nura Muhammad Sani. "Hybrid Deep Learning Model for Fake News Detection on Social Media Using CNN-GRU on X formerly known as Twitter." *Scientific Journal of Computer Science* 2.1 (2026): 96-106.

[7] Mallick, Chandrakant, et al. "Securing the Internet of Things Through Intrusion Detection System Utilizing Machine Ensemble Learning and Feature Extraction Techniques." *SN Computer Science* 7.1 (2026): 66.

[8] Priya, Badugu Akshitha, et al. "Fake news detection using natural language processing." *International Journal of Computer Technology and Electronics Communication* 9.2 (2026): 541-550.

[9] Govinda, Priyadarshini, and Priyanka Govinda. "Enhancing Fake News Detection Using GPT-2 with a Hybrid Deep Learning Approach." (2026).

[10] Paliwal, Swati, and Ramesh Bharti. "Big Data-Driven Fake News Detection Using Modified CNN and Bidirectional RNN Architectures with Spark and Flink Integration." *Intelligent Systems Using Semiconductors for Robotics and IoT*. CRC Press, 2026. 191-195.

[11] Deepa, B. G., et al. "An enhanced deep learning framework for DeepFake detection using EfficientNet-B3 comparative evaluation of deep and machine learning techniques." *Discover Computing* 29.1 (2026): 18.

[12] Fatima, Zaheen, et al. "Framework for IoT Devices Anomaly Mitigation Strategy Using Machine Learning Technique." *Security and Privacy* 9.2 (2026): e70214.

[13] Arshad, Iqra, et al. "A Multi-Platform Approach to Optimizing Depression Detection in Social Media Through Feature Engineering and Contextual Analysis." *International Journal of Computational Intelligence Systems* 19.1 (2026): 97.

[14] Mahmud, Tanjim, et al. "Detecting Fake News in Bangla Using Linguistic and Psychological Features." *ACM Transactions on Social Computing* (2026).

[15] Li, Chang, et al. "Enhancing information credibility assessment through AI: a personality-aware fake news detection model using BERT." *The Electronic Library* (2026): 1-27.

[16] Ahmed, Md Shoaib, and Francesca Spezzano. "A New Attack Surface: XAI-guided Adversarial Comment Generation with LLMs to Attack Fake News Detectors." *Proceedings of the Nineteenth ACM International Conference on Web Search and Data Mining*. 2026.

[17] Arık, Ahmet Okan, Gizem Parlayandemir, and Serra Çelik. "LLM-based data augmentation for text classification on imbalanced datasets: A case study on fake news detection." *Egyptian Informatics Journal* 33 (2026): 100886.

[18] Zhou, Weilin, et al. "DIVER: Dynamic Iterative Visual Evidence Reasoning for Multimodal Fake News Detection." *arXiv preprint arXiv:2601.07178* (2026).

[19] Mahmud, Tanjim, et al. "Detecting Fake News in Bangla Using Linguistic and Psychological Features." *ACM Transactions on Social Computing* (2026).

[20] Thakur, Rahul, J. P. Kavindu Jayasinghe, and Preeti Mehta. "Interpretable and robust multimodal fake news detection with evidence fusion and reliable reasoning." International Journal of Data Science and Analytics 22.1 (2026): 4.

[21] Guo, Hao, et al. "Each fake news is fake in its own way: An attribution multi-granularity benchmark for multimodal fake news detection." Proceedings of the AAAI conference on artificial intelligence. Vol. 39. No. 1. 2025.
[22] Truică, Ciprian-Octavian, et al. "GETAE: graph information enhanced deep neural network ensemble architecture for fake news detection." Expert Systems with Applications 275 (2025): 126984.
[23] Roumeliotis, Konstantinos I., Nikolaos D. Tselikas, and Dimitrios K. Nasiopoulos. "Fake news detection and classification: A comparative study of convolutional neural networks, large language models, and natural language processing models." Future Internet 17.1 (2025): 28.
[24] Qiao, Jiao, et al. "Improving multimodal fake news detection by leveraging cross-modal content correlation." Information Processing & Management 62.5 (2025): 104120.
[25] Bashaddadh, Omar, et al. "Machine learning and deep learning approaches for fake news detection: A systematic review of techniques, challenges, and advancements." IEEE Access (2025).
[26] Wang, Yue, et al. "Heterogeneous network for hierarchical Fine-Grained domain fake news detection." Information Processing & Management 62.4 (2025): 104141.
[27] Raza, Shaina, Drai Paulen-Patterson, and Chen Ding. "Fake news detection: comparative evaluation of BERT-like models and large language models with generative AI-annotated data." Knowledge and Information Systems 67.4 (2025): 3267-3292.
[28] Hussain, Fiza Gulzar, et al. "Fake news detection landscape: Datasets, data modalities, AI approaches, their challenges, and future perspectives." IEEE Access (2025).
[29] Alsuwat, Emad, and Hatim Alsuwat. "An improved multi-modal framework for fake news detection using NLP and Bi-LSTM." The Journal of Supercomputing 81.1 (2025): 177.
[30] Cao, Biwei, et al. "External reliable information-enhanced multimodal contrastive learning for fake news detection." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 39. No. 1. 2025.