Advancing Fake News Detection: A Hybrid Model Integrating BERT and Traditional Machine Learning

[1]Aditya Pandey, [2]Praveen Kumar Mohane
[1]M. Tech Scholar, Department: Computer science engineering, Millennium Institute of technology
[2]Assistant professor, Department: Computer science engineering, Millennium Institute of technology
[1] adityapandey.chitrakoot@gmail.com , [2]pkmohane9910@gmail.com

* Corresponding Author: Aditya Pandey

Abstract:

Fake news propagation is one such affair that seriously threatens the mantle of public trust and demands the development of more efficient detection systems. In this work, we engineer a sturdy hybrid framework that addresses major shortcoming in fake news detection, wherein traditional machine learning models are combined with a deep learning approach using BERT. A thorough preprocessing of data, features engineering, and label validation were performed on LIAR and FakeNewsNet datasets to ensure inputs of the highest quality. Base models were studied against each other: Decision Tree, Random Forest, BERT, and BERTweet. Deep learning models clearly display better performance than traditional ones, with BERTweet showing an accuracy of 96.1%. To maintain the stability in prediction, Voting and Stacking Classifiers were put in use as ensemble methods. The Stacking Classifier topped with an accuracy and F1-score of 97.2%, showing fewer false positives and negatives. In addition, topic modeling with BERTopic brought out crucial thematic patterns in the misinformation, mainly surrounding conspiracy theory and public health hoaxes.

Keywords: *Fake News Detection, BERT, Ensemble Learning, Stacking Classifier, Topic Modeling, Misinformation Analysis.*

I. INTRODUCTION

In the digital hyper-connected environment, the swift spread of misinformation has become one among the societal challenges. Algorithmic perspectives remember that engagement should always be cultivated over accuracy; hence, poignant and emotionally charged content, irrespective of falsehood, gains prominence and popularity [1]. However, unlike old school misinformation, this new-fangled misinformation goes viral in an instant and goes forth to meet the entire world. A study found that there is six-fold speed for fake news against real news on Twitter- propelled on the grounds of novelty and typicality in echo chambers and filter bubbles. Such environments shield users from corrective information, thereby reinforcing already-preconceptions. Another avenue of monetization for misinformation comes from platform design, which incentivizes virality-one-shared-alongside-dismissal-shares-toggle-for-get [2].The ill effects must multiply: misinformation, after all, erodes public trust; it denatures democratic institutions; it mitigates overwhelmingly life-threatening hazards to the public health, as seen during the COVID-19 days. Interdisciplinary insights offered by psychology, computer science, and media studies stress the urgent need for finding mechanisms for reliable detection against this ever-worsening menace.

Reliable intervention against disinformation starts with high-quality data availability. A machine learning model uses these datasets to differentiate between fake and real news depending on patterns, linguistic cues, and contextual features. Datasets such as LIAR and FakeNewsNet try to assert their credibility by sourcing from verified fact-checkers for ground truths and by including speaker information and engagement metrics from social media [3]. However, some challenges remain-the ever-evolving nature of misinformation, the presence of multilingual and multimodal content, and the inability to detect sarcasm or coded language. Finally, there is a challenge in timely detection organic to the volume of data and the biases introduced by algorithms[4].

Fake news mainly aims to deceive readers and manipulate public opinion or to provoke social behaviour and political agendas [5]. While in usual cases misinformation is mistaken for something, here the intention is deliberately to deceive, which makes it deliberately more harm-inducing and difficult to detect. It can appear either in text articles, pictures, videos, or memes. Because of the rise of social media, it spreads fast, thus creating a blurry boundary between the verified news and fake stories [6]. Detecting fake news would require building upon its linguistic structures, visual layouts, emotional solicitations, and contextual inconsistencies. Such deceitful features may be hidden behind plausible headlines or emotionally charged content to garner maximum clicks and engagement [7]. Such an issue is exacerbated by the existence of echo chambers and algorithmic filters that only reinforce exposure to the false narratives. Hence it is necessary to learn about traits that characterize fake news; this forms the basis in building reliable detection mechanisms using NLP and ML techniques [8]. Fig.1 shows Fake News Approach.

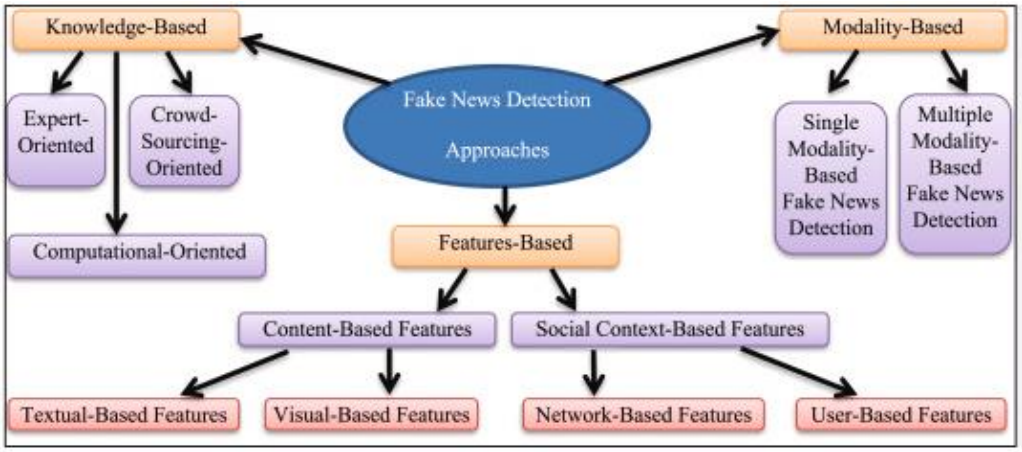

Fig 1: Fake News Approach [7]

II TYPES OF FAKE NEWS

Fake news can be classified into a few types, depending on the intent and the nature of content manipulation. Clickbait covers a range of misleading headlines aimed at grabbing attention and directing the flow of internet traffic but usually do so at the expense of truly representing the story. Propaganda is largely unwelcome or misleading information spread with the purpose of furthering the political or ideological interests of certain groups. Satire, or parody, is meant to entertain or to criticize using humor but often masks itself as real news, particularly when shared without context. Misleading content refers to genuine information that is distorted by context or interpretation. Imposter content mimics real news sources, misguiding unsuspecting readers with fake branding or domain names [9]. Fabricated content is entirely false, created with the aim of misleading and manipulating the audience and with no glimmer of truth. False connections occur when headlines or visuals or captions do not relate to the body's content, while manipulated content denotes the alteration of media-used to deceive. Knowing all these types helps understand the basis in constructing strong detection systems.

III. DEEP LEARNING ALGORITHMS FOR FAKE NEWS DETECTION

With the recent advances in deep learning, fake news detection got revitalized; basically, there is now automated feature extraction from complicated, high-dimensional data. CNNs do well in processing visual misinformation in memes and images. RNN [10], notably LSTM, can capture temporal dependencies and linguistic patterns in textual data. Transformer-based models such as BERT, RoBERTa, and GPT have outperformed in grasping context, semantics, and subtle linguistic cues. Multimodal architectures then combine text and image encoders to jointly detect cross-modal misinformation [11]. Ensemble models and hybrid frameworks--traditionally mixing deep and traditional machine learning algorithms--are in use to increase accuracy and robustness across various misinformation formats [12]. Fig.2. Fake News Detection using Deep Learning.

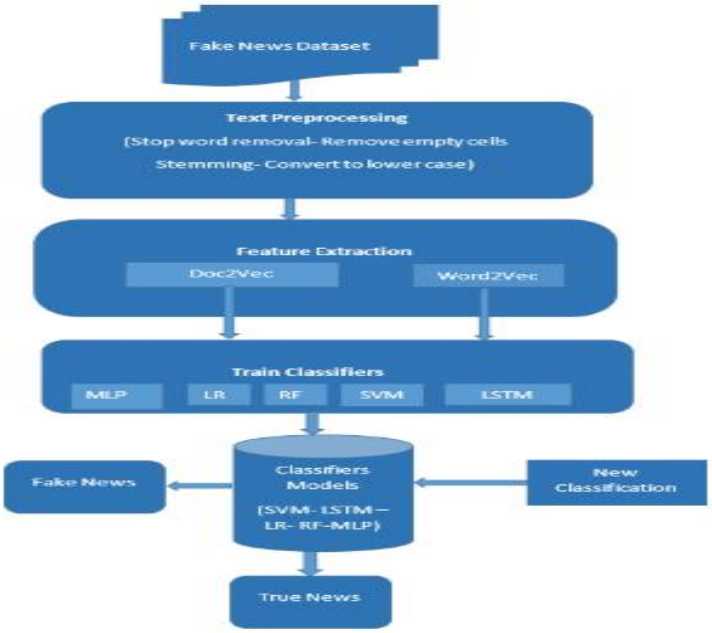

Fig.1.5: Fake News Detection using Deep Learning [12]

IV. LITERATURE REVIEW

Hejamadi Rama Moorthy et al. [1] (2025) used a Dual-Stream Graph-Augmented Transformer combining BERT and GNNs, yielding 99% accuracy on FakeNewsNet. Its disadvantage lies in poor results for multilingual and multimodal data.

Majdi Beseiso et al. [2] (2025) offered a CNN-BiLSTM ensemble, which, using token-level feature extraction, managed 97.3% accuracy. Heavy preprocessing and tokenization are required for this.

Lixin Yun et al. [3] (2025) attempted to enhance fake news detection in Chinese through BERT via concealing strategies with an 83.1% of accuracy. It is non-generalizable to other than Chinese languages.

Kai Yu et al. [4] (2025) configured BMMFN in integrating text (BERT) and image (VGG-19) features, performing well on Twitter and Weibo. The drawbacks include high computation cost and lack of balanced multimodal data.

Bhagyashri Shegokar et al. [5] (2025) considered context-aware sentiment and BERT to finally achieve 92% accuracy on Indian datasets. However, cultural/language limitations and dataset bias remain points of concern.

Jun Wang et al. [6] (2025) employed BERT+LSTM to enhance precision to 93.51% and user validation time. It is incompetent in dealing with visual and multimodal misinformation.

Iman Qays Abduljaleel et al. [7] (2025) combined BERT with CNN-BiLSTM and attention, to achieve 99.27% on ISOT. Not very useful in a real-world situation because of the inability to deal with multimodal information.

S. Raza et al. [8] (2025) evaluated BERT-like and LLMs by employing GPT-4-labeled data and realized a higher classification accuracy for BERT. AI + human-supervised labeling is the best.

Mohammed Al-alshaqi et al. [9] (2025) put forward a multimodal system in which image text is OCR-extracted and processed by BERT. This achieves 99.97% on TRUTHSEEKER. It performs better than other models but is computationally heavy and requires extensive input processing.

Ye Jiang et al. [10] (2025) presented IMFND, utilizing CLIP to guide LVLMs in zero-shot fake news detection. The performance is enhanced at the most insignificant gains from in-context learning.

A BERT-based named entity recognition model was built by Puji Winar Cahyo et al. [11] (2025) with 87.38% F1-score for Indonesian news. The system flags entities pertaining to election news but is language-specific.

Omar Bashaddadh et al. [12] (2025) analyzed 90 research articles, commenting that BERT and multimodal approaches are the most common ones in FND. Some of these researches are concerned about real-time application, data set quality, and explainability.

Jin et al. [13] (2025) proposed a novel CAPE-FND framework based on bootstrap prompting and optimization that outperformed GPT-4 on a number of few-shot tasks. The code is open source and thus facilitates reproducible research.

Rimon Barua et al. [14] (2025) dealt with Bangla fake news using BERT and FastText, and SMOTE increased the F1 to 0.9834. In the binary one, BERT is better, FastText is slight better in the multiclass one.

Anju R et al. [15] (2025) came up with CredBERT using credibility scoring-based ensemble classifier that performs better than the baseline. It performs well on imbalanced data sets but requires source metadata.

Vinita Nair et al. [16] (2025) conceptualized a knowledge-based system model using NLP and graph theory onto Twitter SPO triplets. Effective, but limited with the changing fashions of fake news.

Pummy Dhiman et al. [17] (2025) developed GBERT, an amalgamation of GPT and BERT, able to achieve sterling scores on various benchmarks. This also has some disadvantages, such as dependency on labeled data and the existence of new deceptive language.

Qin et al. [18] (2025) introduced FGM-FRAT for BERT, trying to improve generalization via adversarial fine-tuning. It increases accuracy at the expense of a lot of computations.

Leila Behboudi Angizeh et al. [19] (2025) applied BERT and RoBERTa on the LIAR dataset, achieving an accuracy of 66%. The results are promising but there is still a need for the models to handle better context.

Satish Mahadevan et al.[20] (2025) combined BERT-LSTM and BERT-CNN in a Bayesian framework for more nuanced detection. It does increase accuracy but may overfit and require a high computational load.

Mohammed A. Al Ghamdi et al. [21] (2025) constituted a detection model based on BERT with multilingual capability, covering English, Arabic, and Urdu. Its success was, however, diluted by challenges specific to language.

Suhaib Kh. Hamed et al. [22] (2025) fine-tuned BERT and VGG-19 on Fakeddit with 92% accuracy, though gains are constrained by high computational costs and increased risk of overfitting.

Seemab Hameed et al. [23] (2025) employed BERT and RoBERTa combined with LIME for interpretation, behind an accuracy of more than 98%. This computational effort for interpretation, however, varies in different contexts.

Pratima Singh et al. [24] (2025) proposed a BERT-BiLSTM hybridization for political fake news classification, attaining 96.8% accuracy. Moreover, it is domain-specific and computationally expensive.

Trung Hung Vo et al. [25] (2025) devised a Vietnamese FND system combining CBOW and BERT, reaching 0.96 recall. It is strong in language-specific tasks but weak on cross-lingual generalizability.

Atik Mahabub et al. [26] (2020) proposed an Ensemble Voting Classifier achieving 94.5% accuracy by integrating top-performing ML algorithms. Despite high performance, classical ML models used lack contextual understanding compared to deep learning.

Arush Agarwal et al. [27] (2020) introduced a credibility-based ensemble model using textual features, with LSTM achieving 97% accuracy. However, it faces dataset dependency and struggles with diverse news formats.

Inder Singh et al. [28] (2025) combined traditional ML with BERT to handle both human and AI-generated fake news, reaching 98.2% accuracy. The model excels semantically but is computationally intensive.
Md. Ishraquzzaman et al. [29] (2025) developed a transformer-ML ensemble (RoBERTa, DeBERTa) achieving 96.65% accuracy with high interpretability. Its complexity and resource demand limit rapid adaptability.
Manish Kumar Singh et al. [30] (2025) presented BharatAuthenticNet (BAN), an ensemble model for multilingual fake news detection, achieving 0.9421 F1-score. It requires continuous dataset updates for evolving misinformation.
Poody Rajan Y et al. [31] (2025) proposed Ensemble Feature Fusion (EFF) for Telegram-based fake news detection, outperforming standard models. Its Telegram-specific dataset restricts generalization across platforms.
Saher Fatima Awan et al. [32] (2025) used Instafake data with a voting classifier (RF, ET, NB, MLP) and SMOTE, achieving 98% F1-score. Platform-specific design may limit performance across social behaviors.

III. RESEARCH OBJECTIVES

IV. METHODOLOGY

This research establishes the most comprehensive methodology for detecting fake news, considering a series of benchmark datasets, proper preprocessing techniques, advanced feature engineering, and hybrid modeling. The methodology consists of six main phases: dataset collection and preparation, data preprocessing, feature engineering, label validation, model building, and evaluation. Performing these steps will result in a framework for highly accurate, scalable, and interpretable fake news detection.

A. Dataset Collection and Pre-processing

Two benchmark datasets were used in the study for training and evaluation of the model, namely the LIAR dataset and the FakeNewsNet dataset. The LIAR dataset comes with 12,836 political statements annotated via six different and finely grained truthfulness labels—True, Mostly True, Half True, Mostly False, False, and Pants on Fire. It also provides rich contextual metadata like speaker identity, job title, political affiliation, and historical credibility. Such a multi-class configuration will allow for nuanced classification. The FakeNewsNet dataset comes with factual information from PolitiFact and GossipCop along with social media interactions. It has information on news article content and Twitter engagement information like likes, retweets, and user profile and follower metrics. This dataset is particularly useful for examining social aspects of misinformation. Extensive preprocessing was applied before training. This involved text cleaning by way of removing punctuations, URLs, and stop words, followed by lemmatization. Entries missing some critical information were excluded, and metadata features were normalized for consistent input to the models.

Data preprocessing, thus, was an important methodology that enhanced the quality and integrity of the dataset. The initial step in cleaning involved removing duplicate entries in the text data of both datasets, converting all text to lowercase, removing special characters, and lemmatizing words to keep a uniform representation of the word forms. Following this, any incomplete entries were removed because incompleteness might indicate an absence of claim text, article content, or even the labels, and would affect the reliability of the dataset. Lastly, metadata relevant to FakeNewsNet were extracted, including follower counts, friend counts, the number of retweets, likes, and account age. These values are then normalized to better integrate these features into the model without the effects of scale differences.

The idea behind feature engineering was to capture textual, metadata, and topical information from the data. There were two techniques to extract textual features: TF-IDF vectorization for traditional machine learning models and word embeddings such as Word2Vec and GloVe for deep learning models. With these feature extractors, the semantic content and the importance of words in the corpus are represented. Metadata were derived mostly from FakeNewsNet and concerned behavioral indicators such as number of followers, friends, likes, and retweets. Some additional features included bot scores and credibility scores. Such features measure the credibility of a user and his/her social influence. Topic features were extracted using keyword extraction methods such as YAKE and RAKE that find key terms that highlight thematic directions in misinformation. This will allow an analyst to interpret a model and track new topics of fake news.

B. Model Development

The proposed ensemble model integrates Decision Tree (DT), Random Forest (RF), BERT, and BERTweet to build a robust fake news detection system. Each model is trained separately on features best suited to its strengths: DT and RF utilize metadata and TF-IDF features for interpretable decisions, while BERT and BERTweet are fine-tuned on raw text and tweet content to capture semantic and social media-specific nuances. The system processes three input streams—metadata, news text, and tweets—and routes them to the appropriate models. Each model generates independent predictions, which are then fused using ensemble techniques: **soft voting**, which averages prediction probabilities, and **stacking**, where a meta-classifier (e.g., logistic regression or gradient boosting) learns from base model outputs to generate the final verdict. This hybrid architecture balances the precision of deep learning with the stability of traditional models, improving accuracy, recall, and resistance to overfitting. It effectively handles diverse data types, uniting

linguistic, structural, and social signals for scalable and reliable fake news detection. Fig.3: describes flow chart of proposed model.

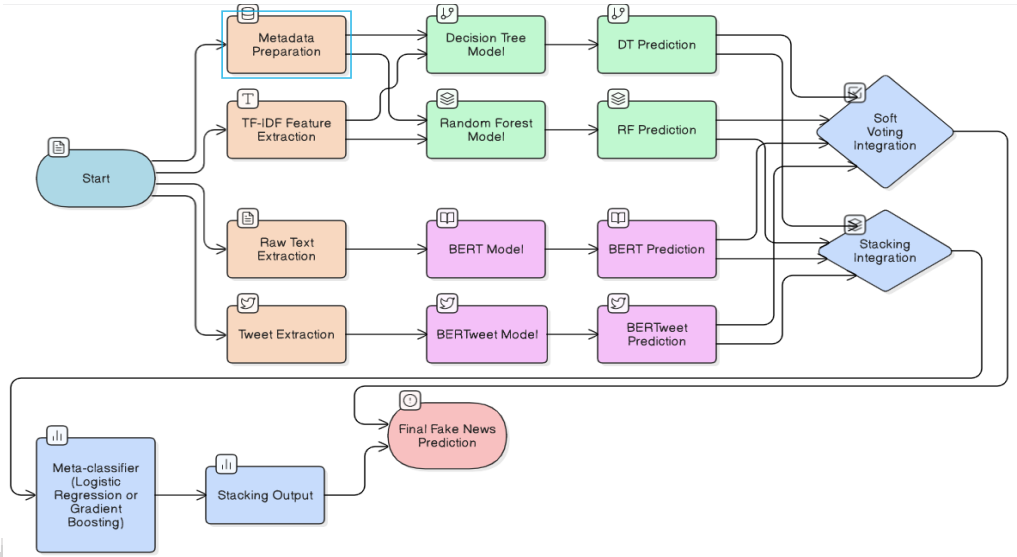

Fig.3: Flow Chart of Proposed Model

C. Model Evaluation

The model underwent evaluation, thanks to a series of well-known performance metrics: accuracy, precision, recall, F1-score, and AUC-ROC. These metrics gave in-depth insights into the performance of the model, especially for imbalanced datasets. Stratified K-Fold cross-validation was considered to reduce the chances of overfitting and to enhance the generalizability of the model, with K mostly being set to 5 or 10 in stratified cross-validation. Conventional models, including Decision Tree and Random Forest, proved to be interpretable but struggled with nuanced textual data. Once the deep learning models were introduced, BERT and BERTweet could ensure proper language representation, which was especially relevant in noisy social media settings. The ensemble model made better use of combined strength than any individual member model, achieving higher accuracy, recall, and overall robustness in fake news detection.

a) **Accuracy**

Accuracy is a basic performance metric used to measure the proportion of correctly classified samples among the total number of samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (1)$$

Where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

Accuracy is suitable when the dataset is balanced.

b) **Precision**

Precision measures how many of the predicted positive cases are actually positive:

$$Precision = \frac{TP}{TP + FP} \qquad (2)$$

A high precision score indicates that the model makes few false positive errors, which is crucial in clinical contexts to avoid incorrectly flagging non-depressed individuals.

c) **Recall**

Recall (also called Sensitivity or True Positive Rate) measures the model's ability to correctly identify actual positive cases:

$$Recall = \frac{TP}{TP + FN} \qquad (3)$$

High recall ensures that most depressed individuals are detected, minimizing the risk of missing a critical diagnosis.

d) **F1-Score**

The F1-score is the harmonic mean of Precision and Recall:

$$F1 - score = 2.\frac{Precision * Recall}{Precision + Recall} \qquad (4)$$

V. RESULTS

A. Model Performance Comparison

The Work compared the performance of individual base models with the proposed hybrid ensemble models on the LIAR and FakeNewsNet datasets.

B. Standalone Model Results:

In assessing the individual models, the aim was to establish baseline performances against which the results of the hybrid ensemble models (created later) could be compared.

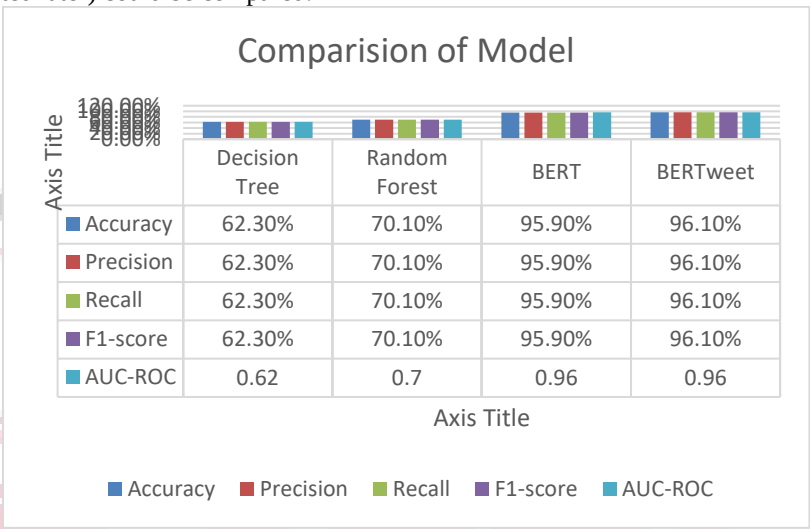

Fig 4.: Results Compression of Model

Fig 4.: Results Compression of Model. The comparative performance analysis of standalone models highlights a clear distinction between traditional machine learning algorithms and advanced deep learning models. Among the machine learning approaches, the Decision Tree yielded the lowest results, with accuracy, precision, recall, and F1-score all at 62.3%, and an AUC-ROC of 0.62—indicating a limited ability to effectively distinguish between fake and real news.

In contrast, deep learning models delivered significantly stronger outcomes. The BERT model achieved 95.9% across accuracy, precision, recall, and F1-score, with an AUC-ROC of 0.96—demonstrating its superior ability to capture language nuances and contextual relationships. BERTweet slightly outperformed BERT, attaining 96.1% across the same metrics, also with an AUC-ROC of 0.96.

C. Hybrid Model Results

Two strategies have been used for ensemble creation: Voting Classifier and Stacking Classifier.

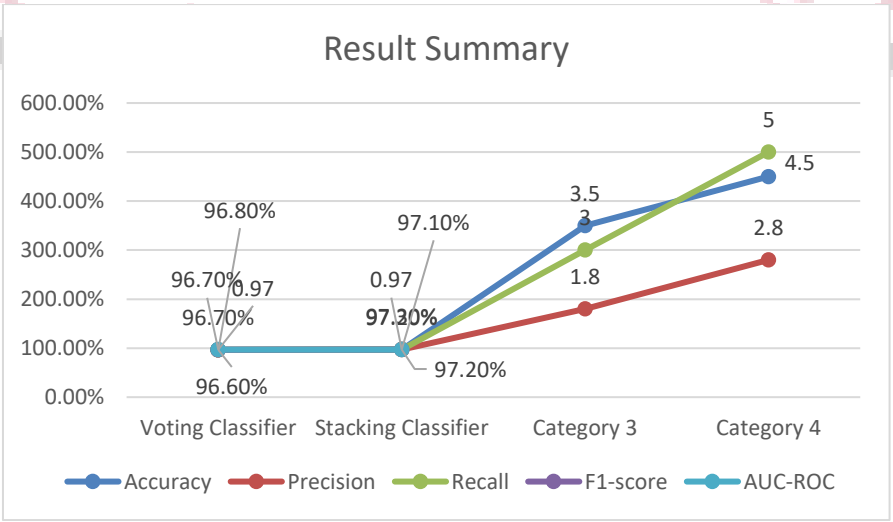

Fig. 5 Result Summary

Fig. 5 gives the summary of results. The ensemble classifier achieved great performance with an accuracy of 96.7%, precision of 96.8%, recall of 96.6%, F1-score of 96.7%, and an AUC-ROC value of 0.97. Thus, it notably improves upon the BERTweet, the bestographical model, which could achieve 96.1% and 96.l% of accuracy and F1-score, respectively. The bar chart beside demonstrates the improvement of accuracy across models-from individual classifiers in light blue to ensemble methods in orange-with increase in the complexity of models and integration yielding huge performance gains.

D. Model Accuracy Comparison

Above is the bar chart that compares and contrasts the accuracy scores obtained by the single base models (light blue) and by the proposed hybrid ensemble models (orange). The models have been arranged in an order of increase in complexity and expected improvement in performance, with the Decision Tree and Random Forests first, followed by BERT and BERTweet, and finally, the ensemble methods-Voting Classifier and Stacking Classifier.

Table 1: Accuracy Comparison

Model	Type	Accuracy (%)
Decision Tree	Base Model	63
Random Forest	Base Model	70
BERT	Base Model	96
BERTweet	Base Model	96
Voting Classifier	Hybrid Model	96.7
Stacking Classifier	Hybrid Model	97.2

The graph exhibits a distinct trend in the improvement of the algorithms as it applies to fake news detection. Decision Tree, bearing the lowest accuracy of just around 62%, is least effective in addressing complex textual patterns. Next, Random Forests employing ensemble learning among decision trees gave about a 70% accuracy rate. Deep learning models, BERT and BERTweet, take a big leap further in that they both achieve an accuracy of around 96% when trained on contextual language understanding. Ensemble Voting Classifier records a 96.7% level of accuracy, which is median to the performance of the two base-level classifiers, while Stacking Classifier performed the best by achieving an accuracy of 97.2%.

.

E. Model F1-Score Comparison

The bar chart compares different F1-scores achieved by the individual base models and hybrid ensemble models. The base models are in light green, while hybrid models are in orange. These models are arranged from left to right: Decision Tree and Random Forest, then the deep learning models BERT and BERTweet, and finally the hybrid models Voting Classifier and Stacking Classifier.

Table 2: F1-Score Comparison

Model	Type	F1-Score (%)
Decision Tree	Base Model	~62%
Random Forest	Base Model	~70%
BERT	Base Model	~96%
BERTweet	Base Model	~96%
Voting Classifier	Hybrid Model	~96.7%
Stacking Classifier	Hybrid Model	~97.2%

The graph indicates that classical learning algorithms of decision tree and random forest yielded rather low F1-scores of about 62% and 70%, respectively, indicating these systems lacked the ability to integrate the linguistic complexities necessary for fake news detection. On the other hand, transformer-based deep-learning methods, such as BERT and BERTweet, witnessed a sharp increase in performance to an F1-score nearing 96% each. BERTweet slightly edged out the regular BERT as it was designed for social media content and is, therefore, better suited for the informality and short-text nature of fake news.Further accuracy enhancements were provided by hybrid ensembles. Voting Classifiers stood with an F1-score of 96.7%, whereas Stacking-Classifiers held the highest F1 score at 97.2%. These results suggest that when these models synergize, they complement each other in strengths, thereby providing a better trade-off between precision and recall and enabling greater overall robustness than any one model acting individually.

The confusion matrix presented above summarizes the performance of the Stacking Classifier in predicting fake and real news. The matrix compares the model's predictions (columns) against the actual ground truth labels (rows).

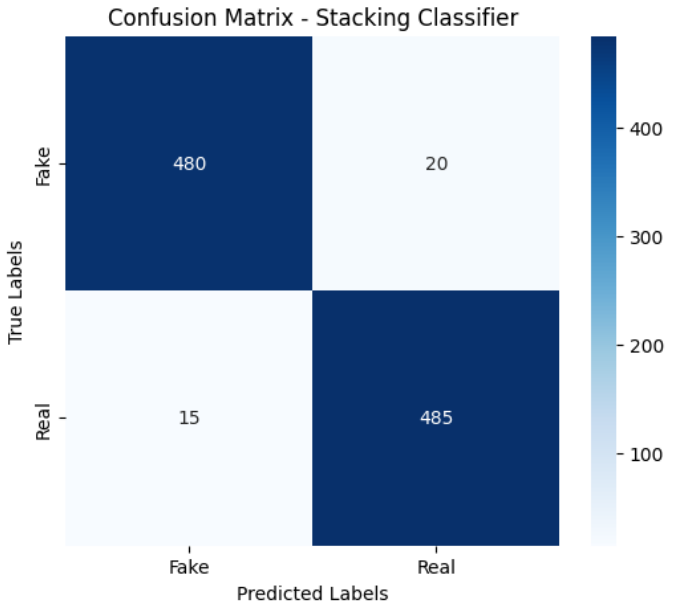

Fig.6: Confusion Matrix

Fig. 6 depicts the confusion matrix, which indicates fine classification by the stacking classifier. Towards 500 fake news articles, the classifier correctly identified about 480, with erroneous labeling of 20 articles as real news. Similarly, about 485 real news articles were correctly classified by our model, whereas 15 were mislabeled as fake news.

F. Stacked Bar Chart — Topic Frequency in Real vs Fake News:

The stacked bar chart above shows the frequency distribution of different topics in real news or fake news. The topics under study are Politics-Elections, Health-Vaccines, Economy & Jobs, Celebrity Gossip, Climate Change, and Conspiracy Theories. For every topic, the whole length of a bar is the collective number of real and fake articles on it, whereas the color segments show to what extent each category is represented.

Table 4.4 Stacked Bar Chart — cumulative topic representation.

Topics	Real News	Fake News	Total Frequency
Politics - Elections	150	220	370
Health - Vaccines	200	300	500
Economy & Jobs	180	100	280
Celebrity Gossip	50	180	230
Climate Change	170	90	260
Conspiracy Theories	30	250	280

Table 4.4 Stacked Bar Chart - cumulative topic representation.The stacked bar chart represents a comparison of the frequency of the major topics appearing in the real and fake news articles. The results show that fake news is most common in the emotionally charged issues of Health - Vaccines, Politics - Elections, Celebrity Gossip, and Conspiracy Theories, often outnumbering real news in these areas. Areas like Climate Change and Economy & Jobs had a higher or balanced presence in real news. It shows that misinformation tends to revolve around contentious and sensational subjects, which in turn will help to devise content-based fake news detection models.

Topic Distribution in Real vs Fake News:

The fig. 7 displays two pie charts showing the percentage distributions of topics within the real news set (left) and the fake news set (right). Each slice represents a topic category: Politics-Elections, Health-Vaccines, Economy & Jobs, Celebrity Gossip, Climate, and Conspiracy Theories. The charts show how often a particular topic appears relative to the total content within the real and fake news sets.

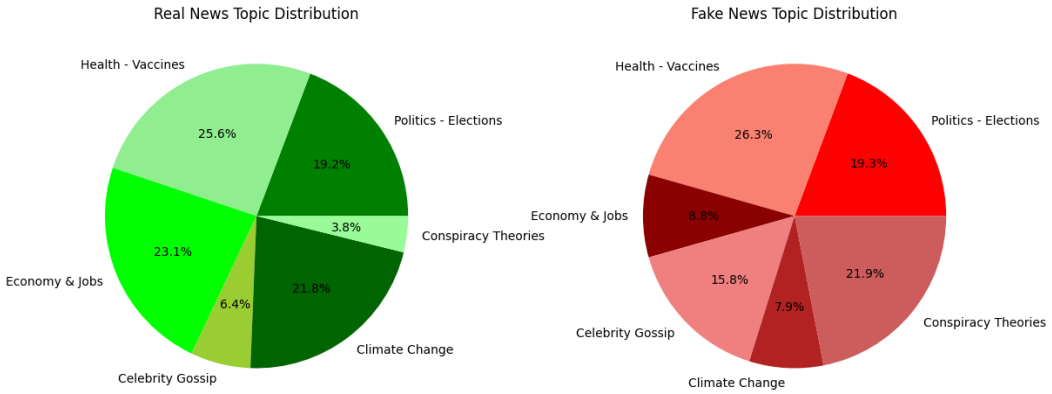

Fig. 7 Two Pie Charts — percentage distribution for each topic in real and fake news.

V. CONCLUSION

This research effectively improved upon many paramount existing limitations imposed in fake news detection problems by developing a comprehensive hybrid model made up of both traditional machine learning and BERT-based deep learning approaches. From the LIAR and FakeNewsNet datasets, rigorous data preprocessing, feature engineering, and label validation were applied to ensure the utmost data quality. Base models on the standalone ones, such as Decision Tree, Random Forest, BERT, and BERTweet, had been thoroughly evaluated. In contrast with the common knowledge, deep learning models such as BERT and BERTweet performed better than traditional machine learning algorithms. BERTweet alone, while standing last in this comparison among the better-performing standalone models, with an accuracy of 96.1%, also showed some pimps for enhancement.

In order to build a more predictive model with improved stability, this work explores two hybrid ensemble approaches, each constructed for a different phase of ensemble techniques: Voting Classifier and Stacking Classifier. The ensemble models integrated the power of the base classifiers. The Stacking Classifier yielded the best performance and really outperformed all individual models with an accuracy and F1-score of 97.2%. The confusion matrix further confirmed a low count of false positives and negatives, backing the reliability of the classifier. Besides, the topic modeling supported by BERTopic helped shed light on the thematic patterns in misinformation. The analysis distinguished the clusters of fake news around emotionally charged, controversial topics: conspiracy theories, vaccine misinformation, and celebrity gossip, whereas cases of real news were more into verifiable issues: politics, health, economy, and climate change.

REFERENCES

[1] H. Rama Moorthy et al., “Dual stream graph augmented transformer model integrating BERT and GNNs for context aware fake news detection,” Sci. Rep., vol. 15, no. 1, pp. 1–25, Dec. 2025, doi: 10.1038/S41598-025-05586-W;SUBJMETA=166,639,704,705,844;KWRD=ENGINEERING,ENVIRONMENTAL+SOCIAL+SCIENCES,MATHEMATICS+AND+COMPUTING.

[2] M. Beseiso and S. Al-Zahrani, “A Context-Enhanced Model for Fake News Detection,” Eng. Technol. Appl. Sci. Res., vol. 15, no. 1, pp. 19128–19135, Feb. 2025, doi: 10.48084/ETASR.9192.

[3] L. Yun, S. Yun, and H. Xue, “Detecting Chinese Disinformation with Fine–Tuned BERT and Contextual Techniques,” Appl. Artif. Intell., vol. 39, no. 1, p. 2525127, Dec. 2025, doi: 10.1080/08839514.2025.2525127;SUBPAGE:STRING:FULL.

[4] K. Yu, S. Jiao, and Z. Ma, “Fake News Detection Based on BERT Multi-domain and Multi-modal Fusion Network,” Comput. Vis. Image Underst., vol. 252, p. 104301, Feb. 2025, doi: 10.1016/J.CVIU.2025.104301.

[5] B. Shegokar and P. K. Deshmukh, “Context-Aware Sentiment Analysis for Enhanced Fake News Detection,” 2025 Int. Conf. Data Sci. Agents Artif. Intell. ICDSAAI 2025, 2025, doi: 10.1109/ICDSAAI65575.2025.11011869.

[6] J. Wang, X. Wang, and A. Yu, “Tackling misinformation in mobile social networks a BERT-LSTM approach for enhancing digital literacy,” Sci. Rep., vol. 15, no. 1, pp. 1–20, Dec. 2025, doi: 10.1038/S41598-025-85308-4;SUBJMETA=1759,685,704,844;KWRD=PSYCHOLOGY+AND+BEHAVIOUR,SUSTAINABILITY.

[7] Iman Qays Abduljaleel, Israa H. Ali “Detecting Fake News Using BERT Word Embedding, Attention Mechanism, Partition and Overlapping Text Techniques,” TEM J., vol. 14, no. 2, pp. 1152–1165, 2025.

[8] S. Raza, D. Paulen-Patterson, and C. Ding, “Fake news detection: comparative evaluation of BERT-like models and large language models with generative AI-annotated data,” Knowl. Inf. Syst., vol. 67, no. 4, pp. 3267–3292, Apr. 2025, doi: 10.1007/S10115-024-02321-1/METRICS.

[9] M. Al-alshaqi, D. B. Rawat, and C. Liu, “A BERT-Based Multimodal Framework for Enhanced Fake News Detection Using Text and Image Data Fusion,” Comput. 2025, Vol. 14, Page 237, vol. 14, no. 6, p. 237, Jun. 2025, doi: 10.3390/COMPUTERS14060237.

[10] Y. Jiang and Y. Wang, "IMFND: In-context multimodal fake news detection with large visual-language models," Knowledge-Based Syst., vol. 325, p. 113880, Sep. 2025, doi: 10.1016/J.KNOSYS.2025.113880.
[11] P. W. Cahyo, U. S. Aesyi, W. A. Setianto, and T. Sulaiman, "A Novel Named Entity Recognition approach of Indonesian fake news using part of speech and BERT model on presidential election," Int. J. Inf. Manag. Data Insights, vol. 5, no. 2, p. 100354, Dec. 2025, doi: 10.1016/J.JJIMEI.2025.100354.
[12] O. Bashaddadh, N. Omar, M. Mohd, and M. N. A. Khalid, "Machine Learning and Deep Learning Approaches for Fake News Detection: A Systematic Review of Techniques, Challenges, and Advancements," IEEE Access, 2025, doi: 10.1109/ACCESS.2025.3572051.
[13] W. Jin et al., "Veracity-Oriented Context-Aware Large Language Models–Based Prompting Optimization for Fake News Detection," Int. J. Intell. Syst., vol. 2025, no. 1, p. 5920142, Jan. 2025, doi: 10.1155/INT/5920142.
[14] R. Barua, M. M. Rahman, and U. G. Joy, "Comparative analysis of Bangla news classification: a study of fake news detection and multiclass classification using BERT and FastText," Int. J. Comput. Appl., vol. 47, no. 5, pp. 475–485, May 2025, doi: 10.1080/1206212X.2025.2495288;PAGE:STRING:ARTICLE/CHAPTER.
[15] R. Anju and N. Pervin, "CredBERT: Credibility-aware BERT model for fake news detection," Data Knowl. Eng., vol. 160, p. 102461, Nov. 2025, doi: 10.1016/J.DATAK.2025.102461.
[16] V. Nair, D. J. Pareek, and S. Bhatt, "A Knowledge-Based Deep Learning Approach for Automatic Fake News Detection using BERT on Twitter," Procedia Comput. Sci., vol. 235, pp. 1870–1882, Jan. 2024, doi: 10.1016/J.PROCS.2024.04.178.
[17] P. Dhiman, A. Kaur, D. Gupta, S. Juneja, A. Nauman, and G. Muhammad, "GBERT: A hybrid deep learning model based on GPT-BERT for fake news detection," Heliyon, vol. 10, no. 16, Aug. 2024, doi: 10.1016/J.HELIYON.2024.E35865/ASSET/C782DB4A-8EC9-4775-AE4D-4739FD5BC183/MAIN.ASSETS/GR9.JPG.
[18] S. Qin and M. Zhang, "Boosting generalization of fine-tuning BERT for fake news detection," Inf. Process. Manag., vol. 61, no. 4, p. 103745, Jul. 2024, doi: 10.1016/J.IPM.2024.103745.
[19] L. B. Angizeh and M. R. Keyvanpour, "Detecting Fake News using Advanced Language Models: BERT and RoBERTa," 2024 10th Int. Conf. Web Res. ICWR 2024, pp. 46–52, 2024, doi: 10.1109/ICWR61162.2024.10533352.
[20] S. Ahmad and S. M. Srinivasan, "BERT based Blended approach for Fake News Detection," J. Big Data Artif. Intell., vol. 2, no. 1, Jan. 2024, doi: 10.54116/JBDAI.V2I1.27.
[21] M. A. Al Ghamdi, M. S. Bhatti, A. Saeed, Z. Gillani, and S. H. Almotiri, "A fusion of BERT, machine learning and manual approach for fake news detection," Multimed. Tools Appl., vol. 83, no. 10, pp. 30095–30112, Mar. 2024, doi: 10.1007/S11042-023-16669-Z/METRICS.
[22] S. K. Hamed, M. J. Ab Aziz, and M. R. Yaakub, "Enhanced Feature Representation for Multimodal Fake News Detection Using Localized Fine-Tuning of Improved BERT and VGG-19 Models," Arab. J. Sci. Eng., vol. 50, no. 10, pp. 7423–7439, May 2025, doi: 10.1007/S13369-024-09354-2/METRICS.
[23] S. Hameed and M. Wasim, "Interpretable BERT and RoBERTa based Fake News Detection using LIME," 2025 Int. Conf. Innov. Artif. Intell. Internet Things, pp. 1–7, May 2025, doi: 10.1109/AIIT63112.2025.11082880.
[24] P. Singh and A. Jain, "A BERT-BiLSTM Approach for Socio-political News Detection," Lect. Notes Networks Syst., vol. 1086 LNNS, pp. 203–212, 2024, doi: 10.1007/978-981-97-6036-7_17.
[25] T. H. Vo, I. Felde, and K. C. Ninh, "Fake News Detection System, based on CBOW and BERT," Acta Polytech. Hungarica, vol. 22, no. 1, pp. 2025–2052, Accessed: Jul. 29, 2025. [Online]. Available: https://www.statista.com/
[26] Mahabub, "A robust technique of fake news detection using Ensemble Voting Classifier and comparison with other classifiers," SN Appl. Sci., vol. 2, no. 4, pp. 1–9, Apr. 2020, doi: 10.1007/S42452-020-2326-Y/TABLES/3.
[27] Agarwal and A. Dixit, "Fake News Detection: An Ensemble Learning Approach," Proc. Int. Conf. Intell. Comput. Control Syst. ICICCS 2020, pp. 1178–1183, May 2020, doi: 10.1109/ICICCS48265.2020.9121030.
[28] Singh, C. Choudhary, and P. Gupta, "Ensemble-Based Framework for Fake News Detection in Social-Media," J. Comput. Inf. Syst., Jun. 2025, doi: 10.1080/08874417.2025.2510430;WGROUP:STRING:PUBLICATION.
[29] M. Ishraquzzaman, M. A. Islam Chowdhury, S. Rahman, and R. Khan, "Ensemble Transformer–Based Detection of Fake and AI–Generated News," Appl. Comput. Intell. Soft Comput., vol. 2025, no. 1, p. 3268456, Jan. 2025, doi: 10.1155/ACIS/3268456.
[30] M. K. Singh, J. Ahmed, K. K. Raghuvanshi, and M. A. Alam, "A Novel Ensemble Model BharatAuthenticNet for Detecting Fake News: Experimentation and Performance Analysis," Arab. J. Sci. Eng., pp. 1–27, Feb. 2025, doi: 10.1007/S13369-025-09975-1/METRICS.
[31] Y. Poody Rajan, K. Kunal, A. Govindan, K. Balu, V. Ganesan, and V. Madeshwaren, "Ensemble-Based Machine Learning Approach For Fake News Detection On Telegram With Enhanced Predictive Accuracy," Int. J. Comput. Exp. Sci. Eng., vol. 11, no. 2, pp. 2068–2077, 2025, doi: 10.22399/ijcesen.1491.
[32] S. F. Awan, M. Wasim, S. Safdar, A. Rehman, and M. U. Ghani, "A Voting Ensemble Model and Explainability Framework for Fake Account Detection in Social Media," 2025 Int. Conf. Emerg. Technol. Electron. Comput. Commun., pp. 1–6, Apr. 2025, doi: 10.1109/ICETECC65365.2025.11070269.