A Robust Hybrid Ensemble Model with Deep Feature Engineering Techniques for Fake News Detection in Social Media

Lakshmi Holla
VTU University, India
laksganesha@gmail.com

K. S. Kavitha
Department of Computer Science and Engineering, Dayananda Sagar College of Engineering, Bangalore, India
dr.kavitha-cse@dayanandasagar.edu (corresponding author)

Received: 3 September 2025 | Revised: 3 October 2025, 14 October 2025, 21 October 2025, and 23 October 2025 | Accepted: 25 October 2025

 DOI: https://doi.org/10.48084/etasr.14505

ABSTRACT

This study introduces a Hybrid Ensemble Model that combines several powerful methods to enhance the accuracy and reliability of fake news detection. Term Frequency-Inverse Document Frequency (TF-IDF) was utilized to highlight important words, and a one-dimensional Convolutional Autoencoder (1-D CAE) was used to capture deeper patterns in text. Subsequently, Pearson correlation was applied to filter the irrelevant information, keeping only the most useful features. These refined features were then passed through a combination of three classifiers: Extreme Gradient Boosting (XGBoost), Deep Forest (DF), and an Adaptive Layer-Based BERT (ALB) model. Final predictions were made through a weighted voting mechanism. The evaluation metrics were estimated on the FA-KES and LIAR datasets and a comparison with previous models, including Transformer+T5, XLNet+GPT-3, and BERT+GPT-2, was conducted. The results revealed 98.75% accuracy, 97.7% precision, 100% recall, and an F1-score of 98.83% on the FA-KES dataset, as well as 98.63% accuracy, 98.32% precision, 98.90% recall, and an F1-score of 98.61% on the LIAR dataset. These results demonstrate that the proposed approach can effectively handle the challenges of detecting fake news in social media.

Keywords-fake news; term frequency-inverse document frequency; one-dimensional convolutional autoencoder; extreme gradient boosting; deep forest; adaptive layer-based BERT; Pearson correlation

I. INTRODUCTION

Social media is one of the main sources of information for the global population. Social Network Sites (SNSs) are platforms where users can share their content openly and participate actively in disseminating information [1]. Popular examples include Facebook, Instagram, WhatsApp, and Twitter [2, 3]. These platforms have replaced the traditional sources of news, since today's youth spend the majority of their time on social media [4]. However, throughout the 2016 US presidential election, a lot of fake news was dispersed and debated on social media and the candidates' support was affected [5-7].

Numerous techniques have been proposed to detect false information. Machine Learning (ML) was the primary approach used in these methods [8-10]. In [11], a geometric deep learning method was developed for Twitter false news identification. Content, user profiles and activities, social graphs, and news dissemination were among the diverse data types that could be combined thanks to the fundamental algorithms, which were an extension of standard convolutional neural networks to graphs. In [12], several parallel blocks of the single-layer deep CNN with varying kernel sizes and filters were combined with the BERT to create the proposed BERT-based deep learning method (FakeBERT). This model was built on top of a bidirectional transformer encoder-based pre-trained word embedding model, BERT.

Authors in [13] released Facebook tools that allowed people to flag bogus news, which inspired the usage of crowd signals. Specifically, they created a brand-new system called Detective that uses Bayesian inference to identify false information and learns from users' flagging accuracy over time. For example, it was helpful to expand their methodology to model and deduce the reliability of sources.

Additionally, authors in [14] introduced a novel problem of the detection of bogus news without supervision. They took news facts and user credibility as latent random variables and utilized social media interactions to gauge consumers' perceptions of news authenticity. To measure the trustworthiness of users and the legitimacy of the news at the same time, an effective Gibbs sampling technique was proposed.

Although previous models produced positive outcomes, they often require a sequence that is difficult to achieve when receiving a fake news article [15-18]. To overcome these issues, a novel method is proposed utilizing two types of datasets. The major focus of this study is to predict the accuracy of news items by training classification models on news characteristics using a labeled dataset of fake and actual news [19-21].

II. PROPOSED METHOD

Figure 1 illustrates the flow diagram of the proposed architecture.

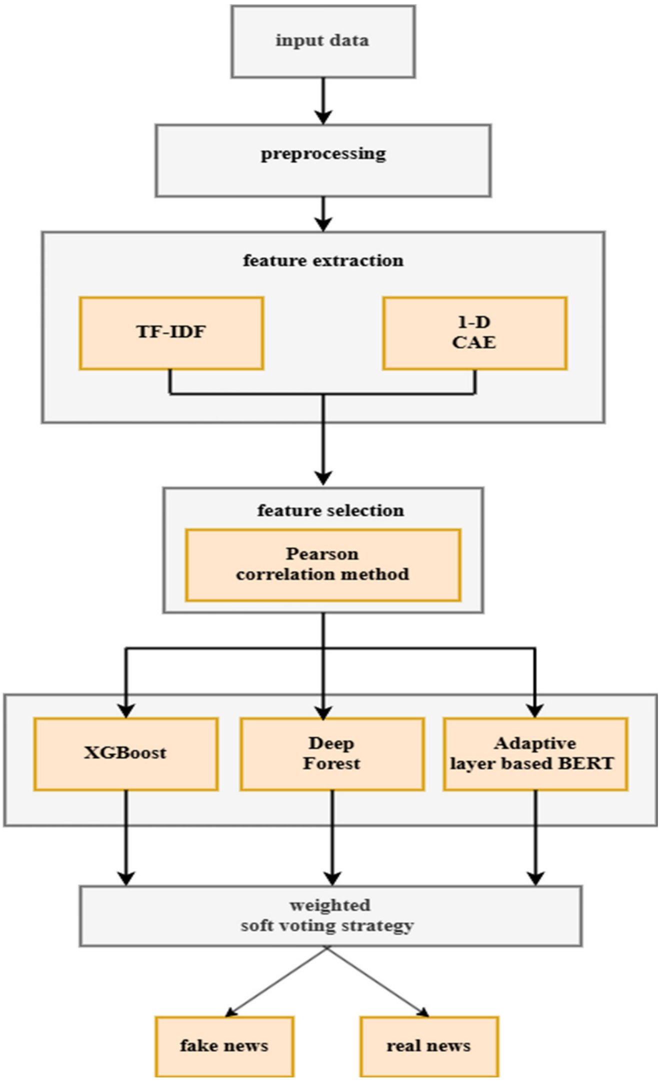


Fig. 1. Flow diagram of the proposed method.

A. Dataset Description

This study utilized two publicly available datasets: FA-KES and LIAR.

1) FA-KES Dataset

The FA-KES dataset contains false information on the Syrian crisis, with the labels 0 (fake) and 1 (credible). The Syrian Violations Documentation Center (VDC) provides ground truth information that was used to calculate an article's credibility. Additionally, the FigureEight crowdsourcing platform (formerly Crowd Flower) was employed to extract information (date, location, and number of victims) for every article.

2) LIAR Dataset

The LIAR dataset is openly accessible for detecting false news. It consists of 12.7K human-labeled short statements from POLITIFACT.COM's API, and each argument was evaluated for veracity by a POLITIFACT.COM editor.

B. Data Preprocessing

Data preprocessing is an essential step for transforming the data into a consistent and analyzable format, tailored for fake news detection. In this study, data preprocessing involved several steps for eliminating extraneous data that could introduce noise, including:

- Removing URLs: URLs and HTML elements in the source code of the web page, and the inclusion of websites with information in English.
- Removing Stopwords: Common words like "the", "is", "and" that often do not carry semantic meaning for NLP models.
- Tokenization: The process of breaking text into meaningful units (tokens), such as words or subwords.
- Stemming: The same procedure was employed, deleting "and" if the remaining portion of the word contained more than or equal to three letters.
- Lemmatization: Input words or tokens were reduced to their lemmas.

Stemming and lemmatization were not applied sequentially (i.e., one after the other) but were rather used as alternative options depending on the experiment or model requirements. This distinction avoided redundancy and potential over-normalization of the text.

C. Feature Extraction Using TF-IDF and 1-D CAE

The feature extraction phase utilized TF-IDF to assign weights to terms based on their frequency, while a 1-D CAE captured complex, high-level features from the text.

1) Term Frequency-Inverse Document Frequency

A numerical statistical method (TF-IDF) is employed in Natural Language Processing (NLP) applications, including text mining and data extraction. This helped the model focus on words specific to false or real news items, excluding low-relevance, frequently used keywords, like phrases that are used a lot, like breaking news.

The TF-IDF score for a term t in document d is computed as:

$$TF-IDF(t,d) = TF(t,d) \times \log \frac{N}{DF(t)} \quad (1)$$

where $TF(t, d)$ measures the term's frequency in a specific post, DF is the number of documents containing t, and N is the total number of documents. The term $log\ (\frac{N}{DF(t)})$ highlights words that are uncommon in the corpus. For instance, emotionally charged phrases, like "hoax" or "scam", tend to have higher TF-IDF scores in false news stories, making them crucial classification characteristics in a dataset of mixed fake and real news. The chronicle is arranged with every word in descending order.

2) One-Dimensional Convolutional Autoencoder

In the detection of fake news on social media, a well-designed 1D-CAE is used to learn meaningful feature representations from text data without relying on label information. Figure 2 depicts the Structure of the 1D-CAE.

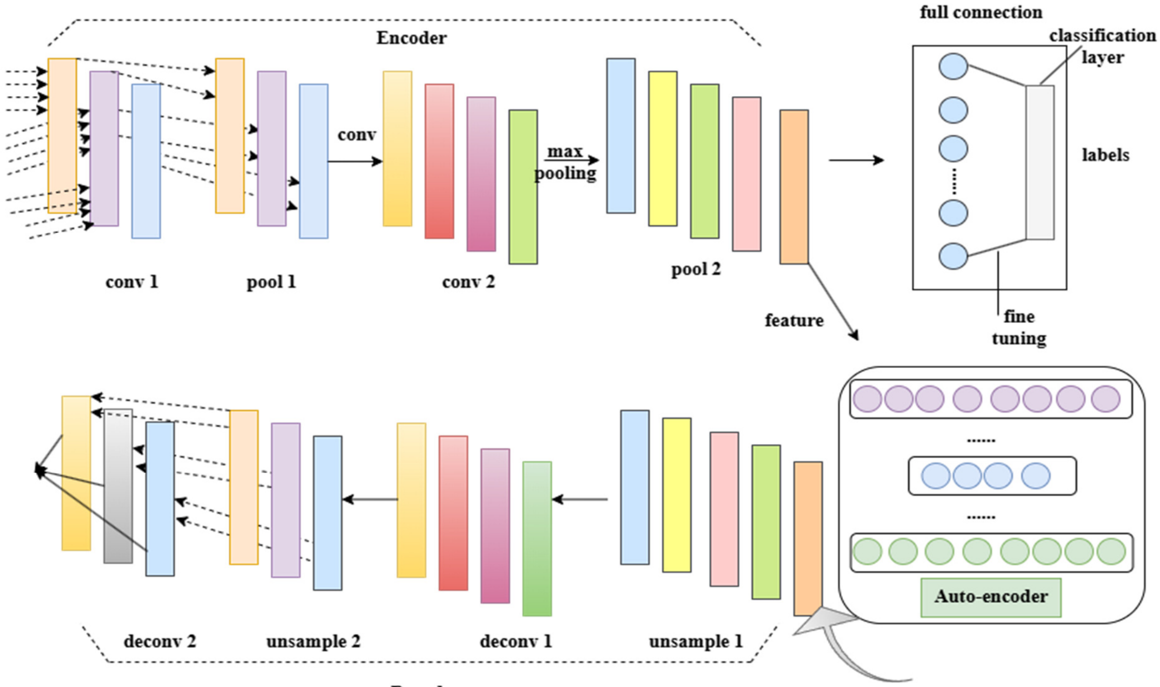


Fig. 2. Structure of 1D-CAE.

A particular convolution kernel produces each output feature. To construct a classification job using a cascaded feature map, the output containing T neurons is placed on top of a 1D-CAE f_v starting the earlier layers as a contribution, and the output is obtained by:

$$K = f(z_0 + c_0 f_v) \quad (2)$$

where z_0 and c_0 are, respectively, the weight and bias vectors. The criteria for learning k_{ij}^p, z_j^p, and z_0 are enhanced during the back propagation phase by gradient descent. The convolutional layer's weight sharing minimizes the amount of network parameters compared to the generic feed-forward back-propagation network, making it easier to prevent gradient vanishing.

3) Combined TF-IDF and 1-D CAE Features

For each document, the TF-IDF vector was concatenated with the 1-D CAE vector to form a combined feature vector:

$$combined_i = [TF-IDF(term), K] \quad (3)$$

The combined vector ($combined_i$) has dimensionality representing both the term relevance (from TF-IDF) and latent patterns (from 1-D CAE). When using TF-IDF, each document is represented as a vector in a high-dimensional space, where each dimension corresponds to a unique term in the vocabulary.

Direct concatenation of TF-IDF and 1D-CAE features with larger magnitudes could dominate the learning process, biasing the model toward one type of feature. To avoid this, each feature vector is normalized individually before fusion. After normalization, the vectors are combined horizontally to form a single unified vector.

This combined vector captures both the semantic information from TF-IDF and the latent patterns from 1D-CAE, providing a richer representation for modeling.

4) Feature Selection Using Pearson Correlation

The combined features from TF-IDF and 1-D CAE were carefully refined to retain only the most relevant information. One common method for feature selection is Pearson correlation, which measures the linear relationship between two variables. The correlation coefficient is defined as:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (5)$$

where r is the Pearson correlation coefficient, which measures the strength and direction of a linear relationship between two variables $\bar{x}$ and $\bar{y}$. The $\bar{r}$ value ranges between -1 and 1.

Pearson correlation is useful in feature selection because it identifies highly correlated features, which often contain redundant information.

5) Classification

The selected features were then fed into a strong ensemble model that combines the strengths of XGBoost, DF, and ALB.

a) XGBoost

XGBoost was utilized to identify false information on social media and was defined in terms of prediction. For binary classification (fake versus real news), XGBoost output probabilities. The final prediction for an instance *i* can be computed as:

$$\hat{y}_i = \sigma\left(\sum_{t=1}^{T} f_t(X_i)\right) \quad (6)$$

where $\sigma(z) = \frac{1}{1+e^{-z}}$ is the sigmoid function that converts raw scores into probabilities, T is the total number of trees, and $F_t(x)$ is the prediction of the t^{th} tree for input x. So, the decision rule is given by:

$$\hat{Y} = \begin{cases} 1, & if \ \hat{y}_i \geq 0.5 \\ 0, & otherwise \end{cases} \quad (7)$$

b) Deep Forest

The chosen characteristics of social media data, which are word sequences, are the inputs of the DF classifier. When creating tweet-level representations, word embedding is crucial. Figure 3 shows the DF structure.

Each tweet or document was converted to a sequence of embedding vectors:

$$Emb_{tweet} = Emb_{word1} + \ldots + Emb_{wordN} \quad (8)$$

where + represents the concatenation operation and Emb_i represents the embedding vector of the i^{th} word.

Similar to other deep neural networks, the Cascade Forest Component (CFC) has a multi-layer structure. Every cascade layer is an ensemble of decision trees that receive data processed by the level before it:

$$PV_{j+1} = f(PV_j) \quad (9)$$

To forecast the final class label, the last cascade layer gets the PV class from every forest in the preceding layer.

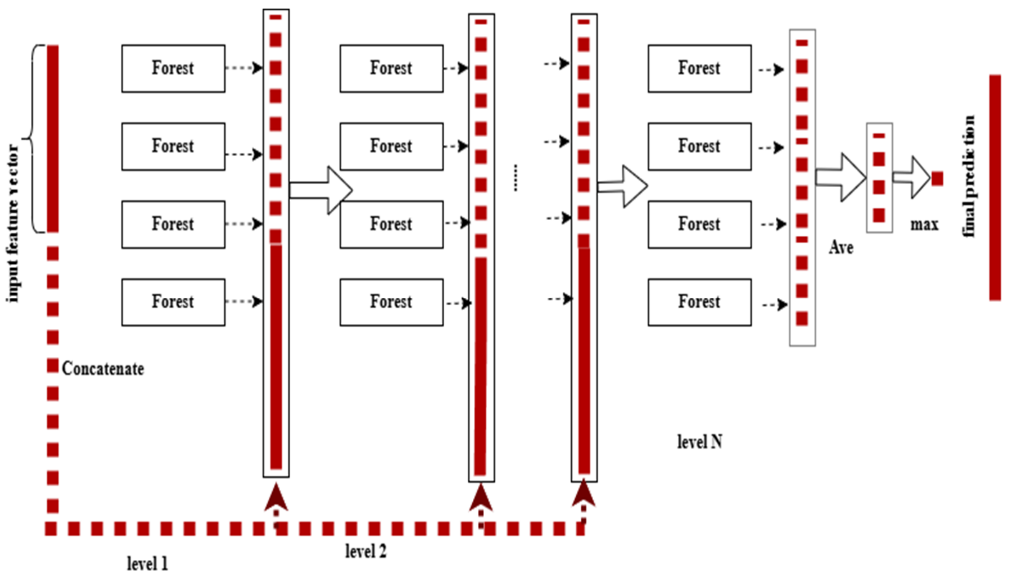


Fig. 3. Structure of DF.

c) Adaptive Layer Based BERT

The digital encoder of bi-directional transformers was used as the central layer for gathering information in the ALB model, which inherits the benefits of the BERT form:

$$U_h = U_{token\_emb} + U_{seg\_emb} + U_{pos\_emb} \quad (10)$$

Every character's encoding representation is denoted by U_h, its token embedding is represented by U_{token_emb}, its segment embedding is represented by U_{seg_emb}, and its place in the embedding is indicated by U_{pos_emb}.

Self-attention was computed through multiple attention heads:

$$head_i = Attention\ (IC_i^I, KC_i^K, FC_i^F) \quad (11)$$

The multi-head attention output was subsequently obtained by using an exponential transformation:

$$MultiHead(I, K, F) = Concat(head_1, head_2, \ldots, head_n)C^O \quad (12)$$

where IC_i^I, KC_i^K, FC_i^F and C^O represent the attention mechanism's weight matrices, and I, K, and F stand for the mechanism's input vectors.

This mechanism effectively captures the syntactic and semantic content of words, which are essential for achieving accurate classification in various NLP tasks.

6) Proposed Ensemble Soft Voting Classifier

The proposed classifier is a meta-classifier that combines conceptually different or identical ML models to make predictions via the vote of the majority. Both hard and soft voting strategies were used by voting classifiers. A minority vote was employed in hard voting to make the final forecast, and the aggregator chose the class prediction that consistently appeared among the base models.

- $P_{XGBoost}\left(y=k|x\right)$ is the probability of class k predicted by XGBoost for instance x
- $P_{DF}\left(y=k|x\right)$ is the probability of class k predicted by DF
- $P_{ALB}\left(y=k|x\right)$ is the probability of class k predicted by ALB

The final probability of class k for an instance x $P\left(y=k|x\right)$ is computed by averaging the probabilities from each model:

$$P\left(y=k|x\right)=\frac{1}{3}\begin{pmatrix}P_{XGBoost}\left(y=k|x\right)\\ +P_{DF}\left(y=k|x\right)+P_{ALB}\left(y=k|x\right)\end{pmatrix} \quad (13)$$

The class with the highest average probability is chosen as the final prediction for the instance x:

$$\hat{y}=\arg\max_k P\left(y=k|x\right) \quad (14)$$

So, if k represents potential classes, like "fake" or "real," the formula chooses the class with the highest averaged probability from the ensemble model as the final prediction.

7) Algorithm 1: Hybrid Ensemble Model for Fake News Detection

The following algorithm presents the pseudocode for the proposed work:

```
Input: Raw text dataset D
Output: Prediction labels (Fake/Real)
```

a) Preprocessing:

```
Clean text (remove punctuation, stopwords, lowercasing).
Tokenize text into words.
Normalize text (lemmatization/stemming).
Represent text using TF-IDF vectors.
```

b) Feature Extraction:

```
Pass TF-IDF features into 1-D Convolutional Autoencoder (1-D CAE).
Train CAE to learn compressed latent representation of text.
Extract encoded features from CAE.
```

c) Feature Selection:

```
Compute Pearson correlation between features and target labels.
Remove weakly correlated features below threshold τ.
Retain refined feature set F.
```

d) Classification:

```
Train three classifiers using F:XGBoost, DF, and ALB.
Obtain prediction probabilities from classifiers.
```

e) Weighted Voting:

```
Assign weights based on validation performance.
Compute final score
If S ≥ 0.5, classify as "Fake", else "Real".
```

f) Return Final Predictions.

D. Experimental Setup

The experimental setup consisted of two data centers, four hosts, and eight gigabytes of RAM, with a bandwidth of 2800 Mbps. The datasets were split into 70% training, 15% validation, and 15% testing. The validation set was utilized for hyperparameter tuning and weight optimization in ensemble voting. Table I shows the hyperparameters of the proposed model.

TABLE I. HYPERPARAMETERS OF THE PROPOSED MODEL

Model	Hyperparameters (key)
XGBoost	learning_rate = 0.05, max_depth = 6, n_estimators = 500, subsample = 0.8, colsample_bytree = 0.8, random_state = 42
DF	n_layers = 6, n_forests_per_layer = 4, n_estimators_per_forest = 200, criterion = "gini", max_depth = None, random_state = 42
ALB	pretrained model = "bert-base-uncased", max_seq_length = 256, batch_size = 32, learning_rate = 2e-5, epochs = 5, dropout_rate = 0.3, optimizer = AdamW, random_state = 42
1-D CAE	conv_layers = 3, kernel_size = 5, stride = 1, latent_dim = 128, activation = ReLU, optimizer = Adam, learning_rate = 1e-3, batch_size = 64, epochs = 50

III. RESULTS AND DISCUSSION

The performance of the proposed Hybrid Ensemble Model was compared with Transformer+T5, XLNet+GPT-3, and BERT+GPT-2. The evaluation metrics included accuracy, precision, recall, F1-score, and accuracy-loss curves.

A. FA-KES Dataset

The FA-KES dataset is a widely used resource for training and evaluating fake news detection models, containing labeled instances of fake and real news. Figure 4 illustrates the confusion matrix of the FA-KES database, providing a visual representation of the classification performance. For the FA-KES dataset, a total of 160 samples were tested, consisting of 75 real and 85 fake news instances. Specifically, the model correctly identified 73 true negatives and 85 true positives, while it misclassified 2 instances as false positives and had no false negatives.

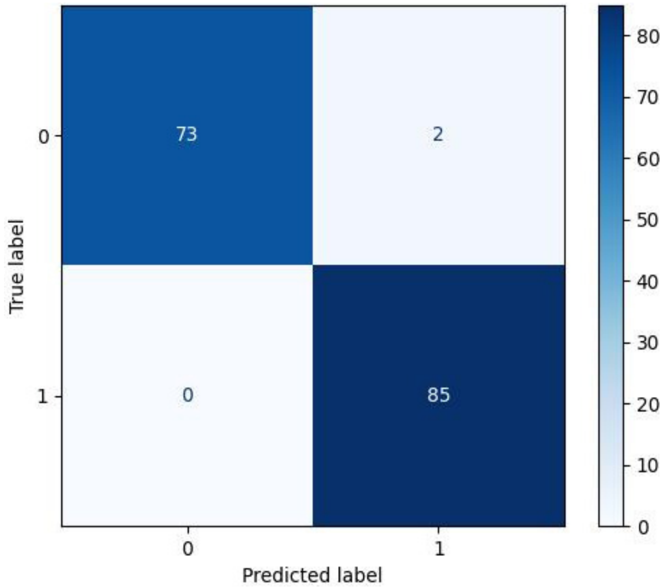
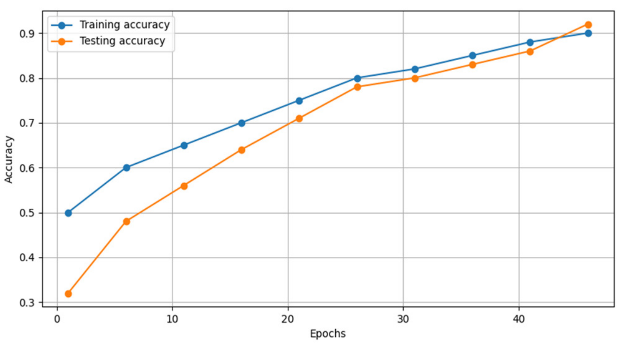


Fig. 4. Confusion matrix for the FA-KES dataset.

Figure 5 depicts the training as well as the testing accuracy and loss curves for the FA-KES dataset. The training accuracy demonstrated a steady increase, indicating improved model performance over epochs.

(a)



(b)

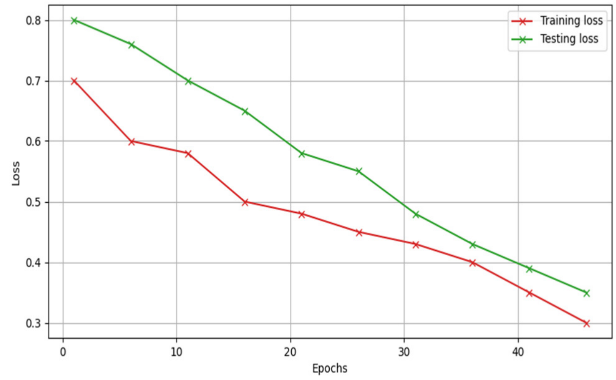


Fig. 5. (a) Accuracy versus epochs curve, (b) Loss versus epochs curve for the FA-KES dataset.

A comparative analysis based on fusion techniques was conducted using the FA-KES dataset, including Graph Fusion, Tensor Fusion, Late Fusion, and Early Fusion.

Figure 6 displays the AUC curve comparison for various fusion techniques, relevant to the FA-KES dataset. This curve analysis underscores the strength of the proposed approach in maximizing true positives while minimizing false positives.

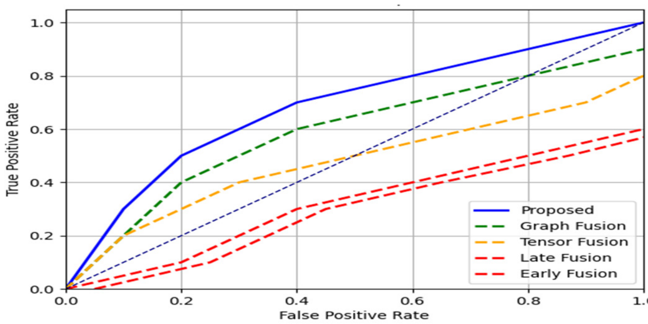


Fig. 6. AUC curve comparison in the fusion stage for the FA-KES dataset.

In the comparative analysis based on classification performance for the FA-KES dataset, the proposed model was evaluated against several existing methods, including Transformer+T5, XLNet+GPT-3, and BERT+GPT-2. This comparison highlighted the effectiveness of each approach in accurately detecting fake news. Figure 7 portrays the AUC curve, showcasing the performance of various models, with the proposed model outperforming.

Table II presents a classification performance comparison on the FA-KES dataset between the proposed model and existing methods were calculated on an independent test set. XLNet+GPT-3 and BERT+GPT-2 showed solid performance but were lower than the proposed approach in all metrics, highlighting the effectiveness of the proposed fusion strategy for fake news detection.

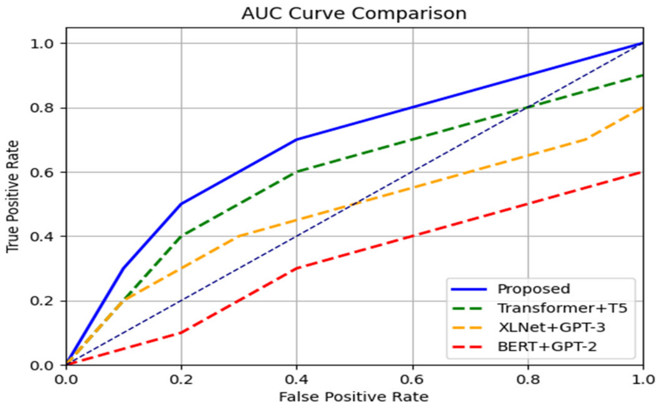


Fig. 7. AUC curve comparison in the classification stage for the FA-KES dataset.

TABLE II. CLASSIFICATION COMPARISON WITH EXISTING METHODS FOR THE FA-KES DATASET

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Transformer+T5	97.5	96.55	98.82	97.67
XLNet+GPT-3	95.62	95.34	96.47	95.9
BERT+GPT-2	93.75	94.11	94.11	94.11
Proposed	98.75	97.7	100	98.83

B. *LIAR Dataset*

The LIAR dataset is a widely used resource for fake news detection research, featuring over 12,000 labeled statements with rich metadata, including speaker information, context, and credibility ratings.

Figure 8 presents the confusion matrix for the LIAR dataset, providing insights into the model's classification performance. A total of 2,048 samples were considered, including 1,041 real and 1,007 fake news items. The model correctly classified 1,024 real and 996 fake samples, demonstrating strong generalization capability. Overall, the confusion matrix illustrates the model's ability to accurately classify the majority of samples while highlighting areas for potential improvement.

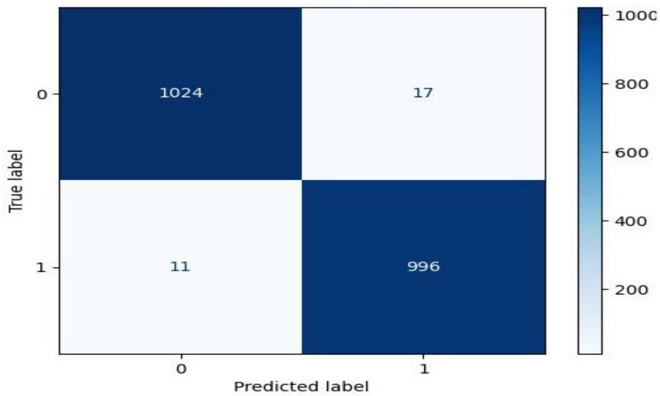


Fig. 8. Confusion matrix for the LIAR datasets.

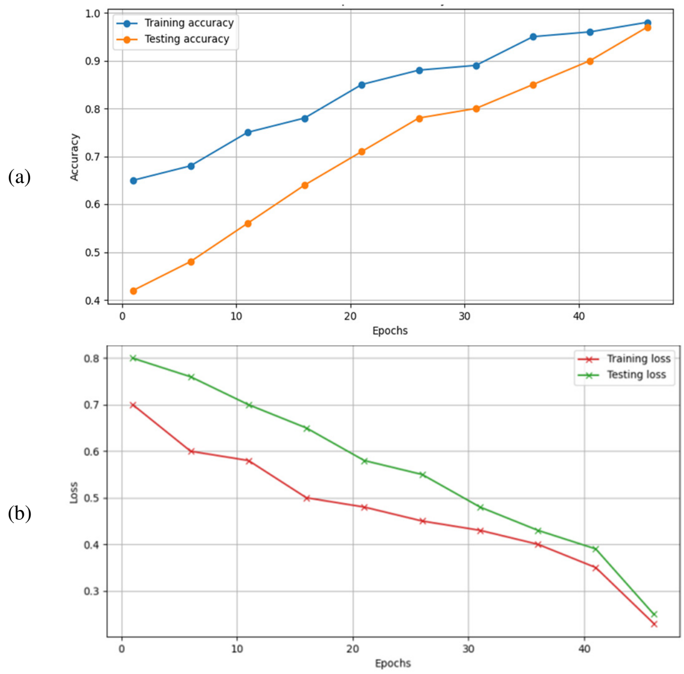


Fig. 9. Accuracy versus Loss for the LIAR dataset.

Figure 9 shows the accuracy and loss curves for the LIAR dataset. The training accuracy displayed a significant upward trend, indicating effective learning as epoch's progress, while loss decreased. This combined trend of increasing accuracy and decreasing loss reflected successful model training and validation.

In the comparative analysis, based on classification performance for the LIAR dataset, the proposed model was evaluated against several existing methods, including Transformer+T5, XLNet+GPT-3, and BERT+GPT-2. Figure 10 illustrates the AUC curve comparison for various classification methods useful to the LIAR dataset. Transformer+T5, XLNet+GPT-3, and BERT+GPT-2 -2 demonstrated progressively lower AUC values, reflecting their reduced effectiveness.

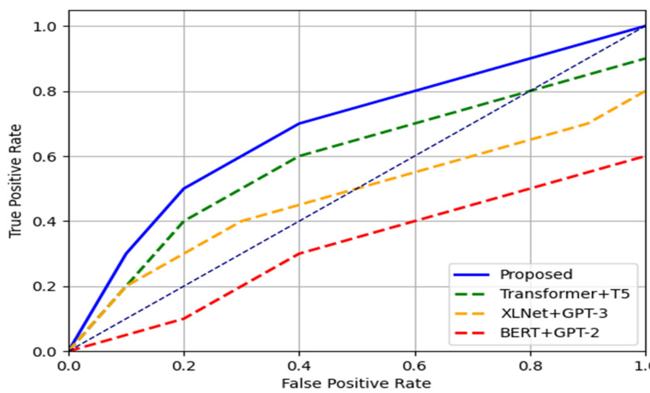


Fig. 10. AUC curve comparison in the classification stage for the LIAR dataset.

Table III compares the classification performance of the proposed method against existing models on the LIAR dataset, using accuracy, precision, recall, and F1-score as evaluation metrics.

TABLE III. CLASSIFICATION COMPARISON WITH EXISTING METHODS FOR THE LIAR DATASET

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Transformer+T5	96.72	96.25	97.12	96.68
XLNet+GPT-3	94.82	94.47	95.03	94.75
BERT+GPT-2	92.52	91.94	92.94	92.44
Proposed	98.63	98.32	98.90	98.61

C. Comparison with Previous Works

Previous literature has explored various approaches to the problem, often highlighting limitations such as inefficiency or limited scalability. This provides a significant contribution to advancing the field.

Table IV presents a performance comparison between the proposed Hybrid Ensemble Model and previous state-of-the-art methods across various fake news datasets.

The use of weighted voting aggregates their complementary strengths, ensuring a high-performing model. To assess the performance improvements of the proposed Hybrid Ensemble Model, paired t-tests and Wilcoxon signed-rank tests were conducted across multiple runs on the FA-KES and LIAR datasets. The results revealed consistently low *p*-values ($p < 0.05$) for accuracy, precision, recall, and F1-score, confirming that the improvements over Transformer+T5, XLNet+GPT-3, and BERT+GPT-2 were statistically significant. These findings demonstrated that the superior performance of the proposed model is robust and unlikely due to random chance. The error analysis examined misclassified instances to reveal patterns, weaknesses, and areas for improvement in the model beyond overall performance metrics.

TABLE IV. PERFORMANCE COMPARISON WITH PREVIOUS SOTA METHODS

Reference	Dataset	Methods	Metrics			
			Accuracy (%)	Precision (%)	Recall (%)	F1-score(%)
[12]	fake news	Geometric deep learning	92.7	-	-	-
[14]	BuzzFeed	UFD	67.9	66.7	71.4	69
Proposed	FA-KES	Hybrid Ensemble Model	98.75	97.7	100	98.83
Proposed	LIAR	Hybrid Ensemble Model	98.63	98.32	98.90	98.61

D. Limitations

Several factors influenced the performance of the proposed model. First, the quality of input attributes derived by TF-IDF and 1D-CAE is crucial to the model performance; predictive accuracy can be negatively impacted by noisy, sparse, or low-representative data. Second, especially with large datasets or high-dimensional TF-IDF representations, the cost of training and fine-tuning the ensemble may be prohibitive, thereby restricting scalability in resource-limited settings. Third, the method is dependent on the choice of hyperparameters for individual component models; if this process is suboptimal, it may result in overfitting or underfitting, which would compromise reproducibility and reliability. Finally, there might be a difficulty in generalizing the model to new domains or datasets due to changes in data distribution, feature properties, or vocabulary that can make the model less effective and need retraining or adaptation. By recognizing these drawbacks, the paper offers a fair analysis that emphasizes the advantages and usefulness of the proposed methodology.

IV. CONCLUSION AND FUTURE SCOPE

This research proposed a Hybrid Ensemble Model for fake news detection on social media. By combining powerful feature extraction techniques, such as Term Frequency-Inverse Document Frequency (TF-IDF) and One-Dimensional Convolutional Autoencoder (1-D CAE), with a robust ensemble of Extreme Gradient Boosting (XGBoost), Deep Forest (DF), and Adaptive Layer-Based BERT (ALB), the model not only enhanced the detection accuracy but also ensured adaptability in handling the dynamic nature of fake news. The results achieved on both the FA-KES and LIAR datasets showed the model’s ability to perform with an accuracy of 98.75%, precision of 97.7%, recall of 100%, and F1-score of 98.83% on FA-KES, and an accuracy of 98.63%, precision of 98.32%, recall of 98.90%, and F1-score of 98.61% on LIAR, proving its potential as a reliable tool for fake news detection.

Future work could explore the integration of real-time data processing and multimodal inputs to further enhance the model's adaptability and performance in detecting evolving fake news across various platforms.

CONFLICT OF INTERESTS

On behalf of all authors, the corresponding author states that there is no conflict of interest.

AVAILABILITY OF DATA AND MATERIAL

Data sharing is not applicable to this article due to its proprietary nature.

CODE AVAILABILITY

Code sharing is not applicable to this article due to its proprietary nature.

REFERENCES

[1] E. Tacchini, G. Ballarin, M. L. D. Vedova, S. Moret, and L. de Alfaro, "Some Like it Hoax: Automated Fake News Detection in Social Networks." arXiv, Apr. 25, 2017, https://doi.org/10.48550/arXiv.1704.07506.

[2] S. R. Sahoo and B. B. Gupta, "Multiple features based approach for automatic fake news detection on social networks using deep learning," *Applied Soft Computing*, vol. 100, Mar. 2021, Art. no. 106983, https://doi.org/10.1016/j.asoc.2020.106983.

[3] M. Balmas, "When Fake News Becomes Real: Combined Exposure to Multiple News Sources and Political Attitudes of Inefficacy, Alienation, and Cynicism," *Communication Research*, vol. 41, no. 3, pp. 430–454, Apr. 2014, https://doi.org/10.1177/0093650212453600.

[4] A. Kesarwani, S. S. Chauhan, and A. R. Nair, "Fake News Detection on Social Media using K-Nearest Neighbor Classifier," in *2020 International Conference on Advances in Computing and Communication Engineering (ICACCE)*, Las Vegas, NV, USA, June 2020, https://doi.org/10.1109/ICACCE49060.2020.9154997.

[5] Z. I. Mahid, S. Manickam, and S. Karuppayah, "Fake News on Social Media: Brief Review on Detection Techniques," in *2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA)*, Subang Jaya, Malaysia, July 2018, https://doi.org/10.1109/ICACCAF.2018.8776689.

[6] J. Gottfried and E. Shearer, "News Use Across Social Media Platforms 2016," Pew Research Center, May 2016, https://www.pewresearch.org/journalism/2016/05/26/news-use-across-social-media-platforms-2016/.

[7] A. Willmore, "The Oscars Were Not Actually A Battle Against Trump's America," *BuzzFeed*, Feb. 27, 2017.

[8] Y. Liu and Y.-F. Wu, "Early Detection of Fake News on Social Media Through Propagation Path Classification with Recurrent and Convolutional Networks," in *Thirty-Second AAAI Conference on Artificial Intelligence*, New Orleans, Louisiana, USA, Feb. 2018, vol. 32, https://doi.org/10.1609/aaai.v32i1.11268.

[9] J. Ma, W. Gao, and K.-F. Wong, "Detect rumors in microblog posts using propagation structure via kernel learning," in *5th Annual Meeting of the Association for Computational Linguistics*, Vancouver, CA, July 2017, https://doi.org/10.18653/v1/P17-1066.

[10] K. Shu, D. Mahudeswaran, S. Wang, and H. Liu, "Hierarchical Propagation Networks for Fake News Detection: Investigation and Exploitation," in *Fourteenth International AAAI Conference on Web and Social Media*, June 2019, https://doi.org/10.1609/icwsm.v14i1.7329.

[11] F. Monti, F. Frasca, D. Eynard, D. Mannion, and M. M. Bronstein, "Fake News Detection on Social Media using Geometric Deep Learning." arXiv, Feb. 10, 2019, https://doi.org/10.48550/arXiv.1902.06673.

[12] R. K. Kaliyar, A. Goswami, and P. Narang, "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach," *Multimedia Tools and Applications*, vol. 80, no. 8, pp. 11765–11788, Jan. 2021, https://doi.org/10.1007/s11042-020-10183-2.

[13] S. Tschiatschek, A. Singla, M. Gomez Rodriguez, A. Merchant, and A. Krause, "Fake News Detection in Social Networks via Crowd Signals," in *The Web Conference 2018*, Lyon France, Apr. 2018, https://doi.org/10.1145/3184558.3188722.

[14] S. Yang, K. Shu, S. Wang, R. Gu, F. Wu, and H. Liu, "Unsupervised Fake News Detection on Social Media: A Generative Approach," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 5644–5651, July 2019, https://doi.org/10.1609/aaai.v33i01.33015644.

[15] F. Sadeghi, A. J. Bidgoly, and H. Amirkhani, "Fake news detection on social media using a natural language inference approach," *Multimedia Tools and Applications*, vol. 81, no. 23, pp. 33801–33821, Apr. 2022, https://doi.org/10.1007/s11042-022-12428-8.

[16] O. Ajao, D. Bhowmik, and S. Zargari, "Sentiment Aware Fake News Detection on Online Social Networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, May 2019, https://doi.org/10.1109/ICASSP.2019.8683170.

[17] X. Dong, U. Victor, and L. Qian, "Two-Path Deep Semisupervised Learning for Timely Fake News Detection," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 6, pp. 1386–1398, Dec. 2020, https://doi.org/10.1109/TCSS.2020.3027639.

[18] H. Reddy, N. Raj, M. Gala, and A. Basava, "Text-mining-based Fake News Detection Using Ensemble Methods," *International Journal of Automation and Computing*, vol. 17, no. 2, pp. 210–221, Apr. 2020, https://doi.org/10.1007/s11633-019-1216-5.

[19] K. Wu, S. Yang, and K. Q. Zhu, "False rumors detection on Sina Weibo by propagation structures," in *IEEE 31st International Conference on Data Engineering*, Seoul, Korea (South), Apr. 2015, pp. 651–662, https://doi.org/10.1109/ICDE.2015.7113322.

[20] L. Wu and H. Liu, "Tracing Fake-News Footprints: Characterizing Social Media Messages by How They Propagate," in *Eleventh ACM International Conference on Web Search and Data Mining*, New York, NY, USA, Oct. 2018, https://doi.org/10.1145/3159652.3159677.

[21] Y. Han, S. Karunasekera, and C. Leckie, "Graph Neural Networks with Continual Learning for Fake News Detection from Social Media." arXiv, Aug. 14, 2020, https://doi.org/10.48550/arXiv.2007.03316.