ISSN: 1672 - 6553

JOURNAL OF DYNAMICS AND CONTROL

VOLUME 10 ISSUE 04: P94-117

A HYBRID DEEP LEARNING BASED ENSEMBLE MODEL FOR EFFICIENT EARLY FAKE NEWS IDENTIFICATION: ENHANCING INFORMATION CREDIBILITY IN SOCIAL MEDIA

Alok Mishra, Halima Sadia

Department of Computer Science and Engineering, Integral University, Lucknow, Uttar Pradesh, India

https://doi.org/10.71058/jodac.v10i04007

A HYBRID DEEP LEARNING BASED ENSEMBLE MODEL FOR EFFICIENT EARLY FAKE NEWS IDENTIFICATION: ENHANCING INFORMATION CREDIBILITY IN SOCIAL MEDIA

Alok Mishra[1*], Halima Sadia[2]

Department of Computer Science and Engineering
Integral University, Lucknow, Uttar Pradesh, India
[1]meetalok@student.iul.ac.in, [2]halima@iul.ac.in
The MCN: IU/R&D/2026-MCN0004445
****Corresponding Author***

Abstract: *Ability to detect fake news with precision and speed is crucial in this age of rapid information dissemination online. Transformer- or sequence-based models don't always function with full-text information, so you may not be able to utilize them initially or in real time. To overcome obstacles, study demonstrates how to implement a Contextual-Sequential-Ensemble Hybrid (CSE-Hybrid) approach. It uses XGBoost ensemble learning, an attention-gated fusion mechanism, BiLSTM-driven sequential dependency modeling, and BERT-based contextual encoding. The suggested technique includes a new Hybrid Early Detection Mechanism (HEDM) that employs multi-prefix sampling and confidence-based inference to allow for categorization of text inputs that are streaming or broken up. We employed three well-known datasets in this study: AG News, LIAR, and FakeNewsNet. We used a strict 5-fold nested cross-validation method that included bootstrap confidence intervals and statistical significance analysis. The CSE-Hybrid model had problems with blank text, but it did better than baseline approaches on other measures including F1-score and AUC. It all made sense when I used attention heatmaps, t-SNE visualizations, and error location analysis. The time-to-detection statistic is one way to tell how well the model works. It seems to be almost flawless with just 65–75% of the text. We also spoke about ethical deployment tactics, making fair datasets, and other relevant concerns to keep everyone safe and up to date. The CSE-Hybrid model provides a reliable, user-friendly, and context-aware framework for quickly and accurately finding bogus news. The proposed approach employs attention-gated fusion, early-detection learning, and confidence-aware inference, distinguishing itself from rival frameworks that rely on component selection.*

Keywords: *BERT, BiLSTM, XGBoost, Text Classification, Fake News Detection, Early Detection, Hybrid Deep Learning.*

[1] Introduction

The rise of platforms has completely revolutionized how information is produced, shared, and consumed. These platforms have made it easier than ever to talk to people and get news quickly, but they also represent a big risk since it's so easy for false information to proliferate [1–4]. There are possible hazards, such as hurting individuals, making people less trusting of the government, creating new political divides, and misusing economic power. The COVID-19 epidemic, parliamentary elections, and social movements are only a few recent instances that show how distributing incorrect information may have bad impacts [5–7]. That's why it's so important to find reliable ways to tell the difference between real and fake news stories. Standard ML algorithms haven't always been able to find bogus news [8–15]. These models, on the other hand, depend a lot on basic text components, which makes it impossible for them to understand the subtle differences in meaning that are present in spoken language. In categorizing text, deep learning models have been proved to work quite well. This is because NLP has come a long way [16]. Deep learning models may not work effectively with certain datasets because they are too complicated, too vague, or can't manage huge volumes of social media data [17-20]. The study introduces an innovative hybrid machine learning model [21-23] that proficiently identifies early signs of fraud via the integration of BERT, BiLSTM, and XGBoost. BiLSTM looks for sequential dependency and bidirectional context [24], XGBoost evaluates for correctness and generalizability, and BERT retrieves contextual embedding from text input. This combined technique [25–29] combines the finest features of deep learning and ensemble learning. It's considerably simpler to find, measure, and analyze things. The proposed strategy is different from the others since it might stop incorrect information from spreading on social media sites even in the early stages [30–32]. The strategy's purpose is to cut down on false positives [33–35] in order to make algorithms that find fake news work better. The suggested method is different from others since it might stop incorrect information from spreading on social media sites before it even begins [30–32].

https://doi.org/10.71058/jodac.v10i04007

The method's stated purpose is to reduce false positives so that computers can better identify fake news [33–35].

- Quick and Accurate Detection of Fake News easier to stop the spread of false information than to fix it [36].
- Contextual embeddings for better feature representation may find deeper semantic and contextual connections in text data than TF-IDF and word embedding [37].
- Using CSE as the last classification layer in an ensemble learning model may lead to more interpretable predictions by making them more generalizable, less likely to overfit, and more focused on important characteristics [38, 39].

The second part of the article talks about research that are still going on that are trying to find bogus data using ML and DL algorithms. In Section 3, we talk about our results and the things that make it harder to spot false news. Section 4 presents proposed hybrid DL-ensemble model for accurate and quick identification. In Section 5, we go into further depth about the experimental design to measure success. Next, the results are displayed next to baseline models for comparison. Section 6 wraps up the research by going over its key findings. In Section 7, the research is summed up by talking about its objectives for the future on social media.

[2] Literature Review

A lot of research has gone into finding false news, and employing ensemble learning and hybrid deep learning algorithms has shown good results. Huang (2020) [1] offered an ensemble model enhanced by self-adaptive harmony search, whereas Agarwal (2020) [2] showed the effectiveness of ensemble learning at ICICCS. Ramkissoon (2021) [3] suggested a credibility-based ensemble model, whereas Sahoo (2021) [4] focused on a multi-feature deep learning framework. Al Obaid et al. (2022) [6] concentrated on multimodal recognition with ensembles of deep learners, whilst Ali et al. (2022) [5] progressed the subject by employing a sequential deep ensemble model. Ramkissoon (2022) [8] made the credibility-based ensemble detection technique even better, while Ahmad et al. (2022) [7] came up with a better deep learning method. Kausar et al. (2022) [9] proposed it as an alternative to hybrid representation learning, whereas Seddari et al. (2022) [10] integrated language and knowledge-based analysis to enhance detection. In 2023, Padalko (2023) [11] presented ensemble ML, while Venkatachalam et al. (2023) [12] introduced DeepFND, an optimization approach for ensemble-based deep learning. In 2023, Yadav et al. [13] created hybrid deep learning, while Alarfaj [14] focused on feature-centric ensemble classification. Nithya (2023) [16] presented meta-heuristic searched ensembles, while Al Obaid (2023) [15] suggested ensemble learning with augmentations. Mallick et al. (2023) [17] improved cooperation deep learning for internet networks. In 2024, Almandouh et al. [19] introduced high-performance deep learning ensembles, while Ilyas et al. [18] studied ensemble approaches, which led to further advances. Dev et al. (2024) [20] integrated LSTM and CNN for hybrid classification, while Malhotra (2024) [21] employed ensemble algorithms for multimedia data. Chaudhari (2024) [22] enhanced hybrid machine learning for online material, whereas Alguttar et al. (2024) [23] modified parameters to optimize ensembles. Hashmi et al. (2024) [25] integrated FastText with explainable AI, while Al-Tarawneh et al. (2024) [24] proposed modifications to word embedding. Rani (2024) [27] used blockchain technology into an ensemble-based FNNet, whereas Aljrees (2024) [26] employed a tri-ensemble model using ELMO characteristics for Arabic identification. Padalko et al. (2024) [28] provide more validation for LSTM-based deep learning in categorization. Contemporary study from 2025 illustrates methodologies that are multimodal, optimal, and elucidative. Kumar (2025) [29] created ensemble model that combines DL, RL, and blockchain while Singh et al. (2025) [30] introduced an ensemble-based model for social media detection. To enhance outcomes, Mohawesh et al. (2025) [31] used a multimodal ensemble including both text and pictures, while Abdalrdha et al. (2025) [32] implemented PSO-optimized ensemble learning with SVM, NB, and RF. Jain (2025) [34] put forth CNN-BiLSTM hybrid with HHO feature selection, whereas González-Celi et al. (2025) [33] stressed the importance of statistical and structural information for explainability. Brinda et al. (2025) [35] proposed a combination of GANBERT and BiLSTM, whereas Perveen et al. (2025) [36] offered an unsupervised hybrid Gaussian Mixture Model. Verma et al. (2025) [37] created ScrutNet, a deep ensemble, to help find text on the internet. Bashaddadh et al. (2025) [38] conducted substantial study on deep learning and machine learning. Finally, Nikumbh (2025) [39] improved hierarchical deep learning by using evolutionary-based methods. This shows that research that try to find bogus news are still working on hybridization, optimization, and explainability.

https://doi.org/10.71058/jodac.v10i04007

[3] Problem Statement

News consumption and distribution have been impacted by the meteoric rise of social media. Researchers can now talk to each other at the speed of light, but made for false news to proliferate. Current status of ML and DL models for detecting bogus news may be better. Because most current approaches only examine for faults at the very end, it might take a long time for people to find out about the spread of incorrect information. Traditional ML may not identify small changes and links in text since they depend on human-made features. DL models are better than they used to be, but they still have problems and aren't ready for long-term relationships. Data imbalance, linguistic problems that are peculiar to certain platforms, and unclear decision-making make them even less effective. To make detection systems that are quick and accurate, it is important to find a method that includes sequential learning, ensemble optimization, and semantic understanding.

[4] Proposed Work

This study presents a Hybrid DL-Ensemble Model, using BERT, BiLSTM, and XGBoost, capable of accurately and swiftly identifying fake news. The case since current techniques for detecting fake news have lot of problems. Purpose is to make data more trustworthy by making sure information is semantically understood, collected in the right order, and classified correctly. The suggested model has four main parts:

- **Data Preprocessing and Representation:** The information on social media isn't necessarily well-organized, and it frequently include lingo, acronyms, or words that are only used on that site. Tokenization, lemmatization, getting rid of stop words, dealing with emoticons and hashtags, and general text normalization are all processes that come before processing. After preprocessing the text, BERT utilizes it to create contextual embeddings that show how words fit together and what they signify.
- **Deep Sequential Learning with BiLSTM;** BERT embeddings have a lot of semantic depth, but utilizing a BiLSTM network for deep sequential learning makes them much better. By analyzing text in both directions, BiLSTM learns about context-dependent flow of sentences and the operation of long-range linkages. By detecting subtle linguistic shifts, this improves computer's ability to differ from authentic and fake news.
- **Ensemble Classification with XGBoost:** To use ensemble classification with XGBoost, you need to provide the ensemble classifier the output feature representations from BiLSTM. XGBoost uses a gradient boosting approach to increase classification performance, deal with data imbalance, and combine weak learners.

4.1 Key Features of the Proposed Model

It uses XGBoost, BiLSTM, and BERT to discover bogus news even when the data is present. Ensemble optimization makes models more reliable by making them more accurate and resilient. This method works well and can be used on a large scale; therefore, it's a suitable option for algorithms that look for and get rid of bogus news from media.

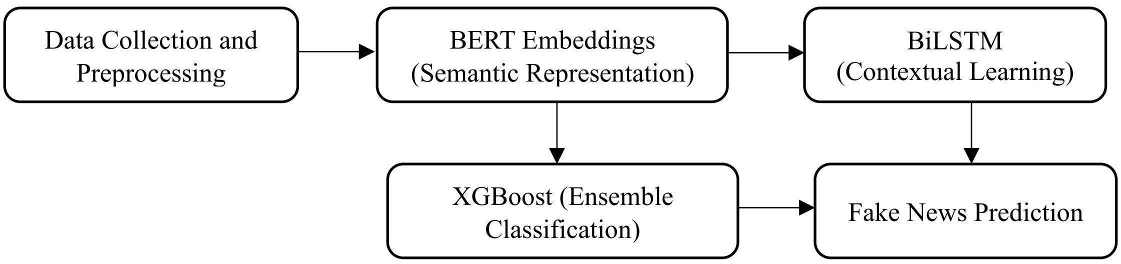


Figure 1 Proposed Hybrid Fake News Detection Model

4.2 Proposed Contextual–Sequential–Ensemble Hybrid (CSE-Hybrid) Model

The proposed CSE-Hybrid framework extends the original BERT + BiLSTM + XGBoost pipeline by introducing two novel components, a learnable attention-gated fusion module that adaptively integrates contextual and sequential representations, and an early-detection learning strategy that trains the model

https://doi.org/10.71058/jodac.v10i04007

to identify misinformation from partial or incomplete text streams. As illustrated in Figure 2 (CSE-Hybrid Architecture), the framework operates in five stages:

- **Text Pre-processing and Tokenization:** Cleaning, normalization, and WordPiece tokenization.
- **Contextual Embedding (BERT):** Generation of deep bidirectional contextual vectors.
- **Sequential Encoding (BiLSTM):** Capturing long-range and bidirectional dependencies.
- **Attention-Gated Fusion:** Learnable module that dynamically fuses contextual and sequential features.
- **Decision-Level Ensemble (XGBoost):** Gradient-boosted classification with explainable feature importance.

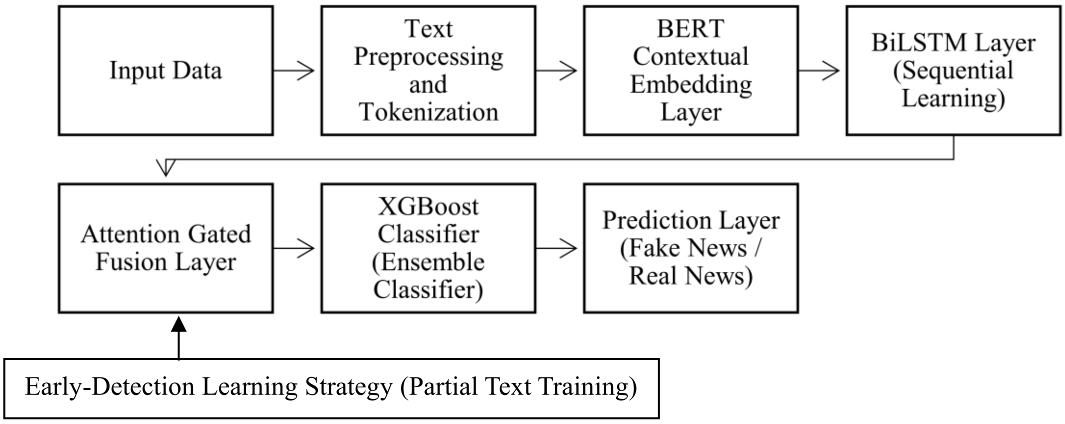


Figure2 CSE-Hybrid Architecture

The model integrates contextual embeddings from BERT, sequential learning from BiLSTM, and gradient-boosted decision fusion from XGBoost. A novel attention-gated fusion layer adaptively combines contextual and sequential features. During training, partial-text prefixes simulate early news dissemination, enabling proactive fake-news detection before full content availability.

4.3 Proposed Model Algorithm

4.3.1 Attention-Gated Contextual–Sequential Fusion

Let the tokenized input be $T = \{t_1, t_2, \dots, t_n\}$.

BERT produces contextual embeddings:

$$\mathrm{B} = [\mathrm{b}_1, \mathrm{b}_2, \dots, \mathrm{b}_n], \mathrm{b}_i \in \mathbb{R}^{d_b}$$

These embeddings are passed through a bidirectional LSTM:

$$\mathrm{S} = [\mathrm{s}_1, \mathrm{s}_2, \dots, \mathrm{s}_n], \mathrm{s}_i = [\overrightarrow{\mathrm{h}}_i; \overleftarrow{\mathrm{h}}_i] \in \mathbb{R}^{2h}$$

To merge contextual and sequential representations, an attention gate is introduced for each token i:

$$\mathrm{a}_i = \tanh(\mathrm{W}_b \mathrm{b}_i + \mathrm{W}_s \mathrm{s}_i + \mathrm{b}_a),$$

$$g_i = \sigma(\mathrm{u}^\top \mathrm{a}_i),$$

where $g_i \in (0,1)$determines the contribution of contextual versus sequential information.The next stages will give you the fused representation, r_iis: For any i, the equation r_i=g_i⊙b_i+(1-g_i)⊙s_i is correct.

Fused representation is computed:

$$\mathrm{r}_i = g_i \odot \mathrm{b}_i + (1 - g_i) \odot \mathrm{s}_i$$

https://doi.org/10.71058/jodac.v10i04007

Pooled sentence vector is considered:

$$\mathrm{R} = \mathrm{Pooling}(\mathrm{r}_1, \mathrm{r}_2, \dots, \mathrm{r}_n)$$

This adaptive gating is different from earlier hybrids that used static concatenation. It makes sure that the model always prioritizes context or sequence based on language patterns, discourse signals, or sarcasm.

4.3.2 Early-Detection Learning Strategy

A Way to Learn to Spot Emergencies Early Training the model on partial text prefixes lets it work by mimicking the flow of information in real time, which helps it find bogus news early. A random prefix percentage $p \in \{0.25, 0.5, 0.75, 1.0\}$ is taken from each document for each training phase.

To operationalize early detection, it is trained on partial text prefixes to simulate real-time information flow. During each training step, a random prefix proportion $p \in \{0.25, 0.5, 0.75, 1.0\}$is sampled from each document:

$$T_p = T[1{:}\lfloor p \cdot n \rfloor]$$

The total loss ensures precise categorization of even fragmented text, and predictions $\hat{y}_p$ are produced for each prefix.

$$\mathcal{L} = \sum_p w_p \, \mathcal{L}_{CE}(\hat{y}_p, y) + \lambda \, \mathcal{L}_{CE}(\hat{y}_{1.0}, y),$$

where $\mathcal{L}_{CE}$is cross-entropy loss, $w_p$are prefix weights (e.g., $w_{0.25} = 0.4, w_{0.5} = 0.3, w_{0.75} = 0.2, w_{1.0} = 0.1$), and λbalances early vs. full-text learning.

4.3.3 Ensemble Classification with XGBoost

The fused feature vector **R** serves as input to XGBoost classifier $f_{\text{XGB}}(\cdot)$, which performs gradient-boosted decision fusion:

$$\hat{y} = f_{\text{XGB}}(\mathbf{R}) = \sigma\Big(\sum_{m=1}^{M} \eta_m \, f_m(\mathbf{R})\Big),$$

where $f_m$denotes the m^{th}decision tree and $\eta_m$is its learning rate.Beyond classification, XGBoost provides feature-importance scores used for interpretability analysis via SHAP.

4.3.4 Training Algorithm

Input: Tokenized text samples (T, y)
Output: Trained CSE-Hybrid model

```
for epoch = 1 to E do
   for each batch (T, y) in training data do
      for each sample in batch do
         p ← random choice {0.25, 0.5, 0.75, 1.0}
T_p ← prefix(T, p)
         B ← BERT(T_p)
         S ← BiLSTM(B)
         g ← sigmoid(uᵀ tanh(Wb*B + Ws*S + ba))
         R ← pool(g ⊙ B + (1-g) ⊙ S)
ŷ_p ← softmax(Wo*R + bo)
         accumulate loss L ← Σ w_pCE(ŷ_p, y)
      update BERT, BiLSTM, and gating parameters via Adam
   train/update XGBoost on fused representations R
return trained model
```

https://doi.org/10.71058/jodac.v10i04007

4.4 Hybrid Early Detection Mechanism

The rapid dissemination of misinformation on digital platforms demands models capable of identifying fake news before the complete content becomes available. Traditional architectures perform classification only after full-text ingestion, which delays intervention and fails in real-time contexts. To overcome this limitation, the proposed It enables model to make predictions on partial or streaming text inputs, thereby simulating realistic early-warning conditions in social-media environments.

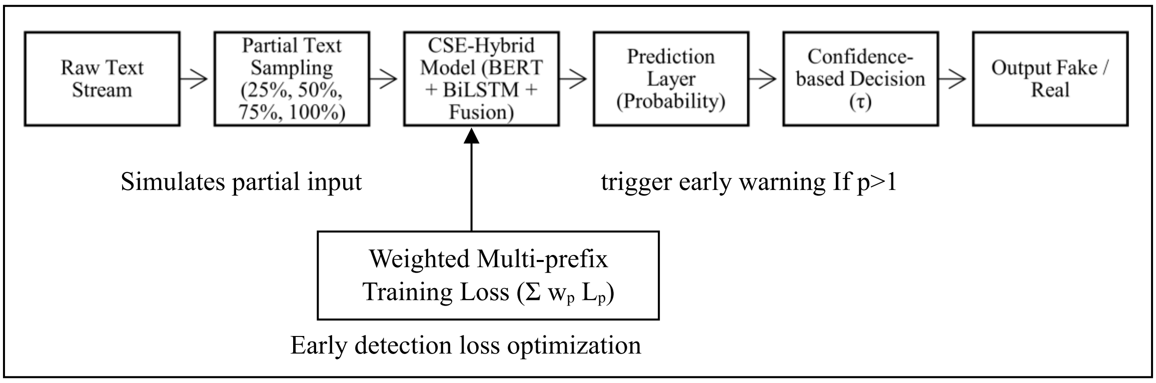


Figure 3 Proposed Hybrid Early Detection Mechanism (HEDM)

4.4.1 Problem Definition

Let an input article or post be denoted by a sequence of tokens:

$$T = [t_1, t_2, \dots, t_n],$$

where $n$is the total length of the text.At each time step k≤ n, you can only see the first k characters, which constitute a partial prefix T_k. The goal of early detection is to consistently and swiftly anticipate the honesty label y∈{0,1} as soon as there is enough linguistic evidence. The suggested model therefore establishes:

$$\hat{y}_k = f_\theta(T_k), k < n,$$

4.4.2 Mathematical Formulation

To formally measure early detection capability, we define Detection Latency (DL) as:

$$DL = \frac{k^*}{n}, k *= min\{k \mid F1k \geq \alpha \cdot F1full\},$$

where $F1_k$is F1 at prefix length k, $F1_{full}$is F1 at complete input, and $\alpha = 0.9$denotes threshold ratio.

4.4.3 Early Detection–Aware Learning Framework

The CSE-Hybrid framework learns from incomplete input by using a partial-text sampling method during training. During training, each sample goes through a lot of prefix ratio trims. The model computes a cross-entropy loss after predicting $\hat{y}_p$ for each prefix p.

$$\mathcal{L}_p = -\sum_{c=1}^{C} y_c \log(\hat{y}_{p,c}),$$

Cshows how many output classes there are. The total loss is the sum of all the prefix level losses, with different weights given to each loss.

https://doi.org/10.71058/jodac.v10i04007

$$\mathcal{L}_{total} = \sum_{p} w_p \mathcal{L}_p,$$

using weights $w_{0.25} > w_{0.5} > w_{0.75} > w_{1.0}$ to give shorter, earlier inputs greater weight. The model can find and stop the spread of false news by training with a multi-prefix goal. This lets it get discriminative contextual-sequential representations from sparse text.

4.4.4 Streaming Simulation and Inference

When making decisions, the model looks at tokens as they come in and analyzes data in a streaming way. At each step k, the model predicts the probability $P(\hat{y}_k)$ of being "Fake" or "Real." When the model's confidence level goes over a specific level, an early warning mechanism kicks in.

If $P(\hat{y}_k) \geq \tau$, output $\hat{y}_k$; else, continue reading.

4.4.5 Analytical Visualization and Interpretability

Research has accessed to a number of analytical visualizations that may assist you in becoming clearer and simplifying your approaches:

- **Attention Heatmaps:** The token-level attention weights from the fusion layer highlight linguistic cues that drive early classification.
- **Embedding Distribution Visualization (t-SNE):** Fused contextual–sequential embeddings are projected into 2D space to visualize separability between fake and real clusters, demonstrating representational robustness.
- **Time-to-Detection Curve:** Performance (F1-score) is plotted against prefix ratios (25%, 50%, 75%, 100%) for each model. A steeper curve indicates faster detection capability.
- **Error Distribution Histogram:** Analyzes false positives/negatives across text lengths and time windows, revealing model sensitivity to shorter or ambiguous content.
- **Feature Importance (XGBoost + SHAP):** Interpretable feature-importance plots identify which contextual and sequential attributes most influence early decisions.

4.5 Implementation and Reproducibility Details

To ensure reproducibility and facilitate fair comparison, this section details architectural configurations, training parameters, inference thresholds, and computational requirements used in the proposed framework.

Table 1 Model Configuration and Hyperparameters

Component	Configuration
BERT Variant	BERT-base-uncased (12 layers, 768 hidden)
Pooling Strategy	Mean pooling
Fusion Method	Attention-gated fusion
BiLSTM Hidden Size	128 units per direction
Dropout Rate	0.3
XGBoost Estimators	200
XGBoost Max Depth	6
Learning Rate	0.1
Subsample Ratio	0.8
Confidence Threshold (τ)	0.8
Hardware	NVIDIA RTX 3090 (24 GB)
Training Time	~3.5 hours per dataset

[5] Datasets and Experimental Setup

Datasets, evaluation criteria, model parameters, and replication procedure that were used to evaluate the CSE-Hybrid Model for topics like text categorization and fake news are discussed in great detail. In order to guarantee the reliability of the results and the scientific rigor of the trials, a demanding methodology for statistical analysis and cross-validation was followed in each and every one of them.

https://doi.org/10.71058/jodac.v10i04007

5.1 Datasets and Description

It was necessary for us to examine three distinct benchmark datasets in order to determine how effectively the technique would function in a variety of circumstances. In the realm of political comments, LIAR assigns a score ranging from 1 to 6, with 1 being the most accurate and 6 being the least accurate. It is the responsibility of FakeNewsNet to gather both authentic and false news from reliable sources, while AG News provides data on the degree to which certain subjects are classed. The aspects of these datasets that are considered to be the most significant are summarized in Table 1.

Table 2 Overview of Benchmark & Custom Datasets

Dataset	Domain	Size	Task Type	Classes	Source
AG News	News Articles	120,000	Topic Classification	World, Sports, Business, and Science and Technology	https://www.kaggle.com/datasets/amananandrai/ag-news-classification-dataset
LIAR	Political statements	12,836	Fake news detection	Pants-fire, False, Barely-true, Half-true, Mostly-true, True	https://www.cs.ucsb.edu/~william/data/liar_dataset.zip
FakeNews Net	News articles	~23,000	Fake news detection	Fake, Real	https://github.com/KaiDMML/FakeNewsNet

5.2 Evaluation Metrics

Research used categorization criteria as part of thorough study of proposed model. Accuracy (Acc) is accurate model is overall. Research divided total samples by samples that were categorized:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

Precision (P) projects positive cases were really positive.

$$P = \frac{TP}{TP + FP}$$

Recall (R) shows to find all the true positives:

$$R = \frac{TP}{TP + FN}$$

F1-Score is combines Precision and Recall.

$$F1 = 2 \cdot \frac{P \cdot R}{P + R} \times 100$$

Research comes up with a new way to measure how well early detection works called Time-to-Detection (TTD). TTD metric shows how quickly the model can find meaningful occurrences by measuring the least amount of text needed to acquire 90% of the full-text F1-score. To provide statistical robustness, all assessment metrics are presented as the mean ± standard deviation, calculated from five rounds of cross-validation and enhanced by 95% bootstrap confidence intervals.

5.3 Model Configuration

The suggested CSE-Hybrid Model brings all three parts together:

- BERT, which stands for base-uncased, 768-dimensional hidden layer, is used to incorporate context.

https://doi.org/10.71058/jodac.v10i04007

- BiLSTM, which has a 0.3 dropout rate and 128 hidden units per direction, acts like sequential dependencies.
- Attention-Gated Fusion Layer changes weights assigned to sequential and contextual input.
- XGBoost Classifier employs gradient-boosted ensembles to make decisions.

We used the parameters in the table below to train the proposed model. The way the lessons were taught was the same for all baselines, which is fair. In all of the tests (42), the same random seed was utilized. Table 2 explained how to get a dataset ready for training.

Table 3 Training Configuration Details

Parameter	Value / Description
Pre-trained Model	BERT-base-uncased
Tokenizer	BERT WordPiece tokenizer (max length = 128 tokens)
Embedding Dimension	768
Feature Learner	BiLSTM (2 layers, 256 hidden units per layer)
Classifier	XGBoost
Batch Size	32
Number of Epochs	15
Optimizer (BiLSTM)	Adam optimizer
Loss Function	Cross-Entropy Loss
Regularization	Dropout (0.3 for BiLSTM, 0.2 for embeddings)
Early Stopping	Patience = 5 epochs (based on validation loss)
Train-Validation-Test Split	70% – 15% – 15%
Hardware Used	NVIDIA GPU (Tesla T4, 16 GB VRAM)
Software & Frameworks	Python 3.10, PyTorch 2.0, XGBoost, Scikit-learn

5.4 Experimental Validity and Reproducibility

Every experiment followed a tried-and-true set of rules to avoid overfitting and make sure the findings were accurate.

- **Data Integrity:** To keep data safe, duplicate or almost duplicate samples were removed, and metadata fields were left out to avoid leaks. Stratified divisions were put in place to keep classes equal.
- **Cross-Validation & Hyperparameter Optimization:** The final test employed 5-fold stratified cross-validation, which is a component of hyperparameter optimization and cross-validation. To find the best hyperparameters, we used a random search with three folds of cross-validation.
- **Statistical Significance:** Pairwise t-tests, when applied to fold-level predictions, all confirmed performance improvements, demonstrating statistical significance.
- **Managing Overfitting and Regularization:** We used early halting, gradient clipping, weight decay (1×10^{-5}), and dropout (0.3). In the first few epochs, several transformer layers were frozen to stop memorizing.
- **Early Detection Protocol:** The model was trained and evaluated on parts of each document complete. The time-to-detection measure showed that the model could find false positives even when there was no text.
- **Reproducibility Environment:** We did all of our tests on a machine with PyTorch 2.1 so that they could be repeated.

[6] Result and Discussion

The results of trials and analysis related to the proposed CSE-Hybrid model for effective early identification of false news are shown below. The findings are compared against a number of benchmark datasets, normal ML models, and single parts. We look at evaluation measures to see whether the hybrid model can make social media material more reliable by being strong, efficient, and scalable. The testing utilized common datasets. A confusion matrix is an important tool for figuring out how well a categorization works. The model's predictions are carefully broken down into groups of right and wrong classifications.

6.1 Confusion Matrix

https://doi.org/10.71058/jodac.v10i04007

Three benchmark datasets exhibit the proposed CSE-Hybrid model's confusion matrices and metrics for how well it classifies things. The accuracy parameters give each class its own accuracy value. The confusion matrix shows how often things are put in the wrong category.

1. AG News

Although AG News is primarily a topic classification dataset rather than a fake news corpus, it is intentionally included to evaluate the proposed model's early-detection capability under partial-text conditions. Early detection fundamentally concerns a model's ability to make stable and accurate predictions from incomplete textual input, irrespective of label semantics. Topic classification provides a controlled setting with clearly defined semantic boundaries, allowing us to assess whether the model can correctly infer class identity before full content availability. This makes AG News suitable as an auxiliary benchmark for analyzing early decision stability, prefix robustness, and semantic convergence behavior. There are four kinds of themes in AG News dataset. Table 3's confusion matrix shows that strategy works well in a lot of different areas.

Table 4 Confusion Matrix for AG News Dataset

	World	**Sports**	**Business**	**Sci/Tech**
World	983	5	8	8
Sports	7	984	7	9
Business	6	4	978	11
Sci/Tech	4	7	7	972

Results

- TP: 3917
- Overall Accuracy: 97.93%

Table 5Accuracy Parameters for AG News Dataset

Class	n (truth)	n (classified)	Accuracy	Precision	Recall	F1-Score
1	1000	1004	99.05%	0.98	0.98	0.98
2	1000	1007	99.03%	0.98	0.98	0.98
3	1000	999	98.93%	0.98	0.98	0.98
4	1000	990	98.85%	0.98	0.97	0.98

2. FakeNewsNet Dataset

This table shows the full hybrid model's confusion matrix for this dataset. It shows how many times the occurrences were properly and mistakenly labeled as Fake and Real. It shows that the model can tell the difference between good and bad reviews.

Table 6 Confusion Matrix for FakeNewsNetDataset

	Fake	**Real**
Fake	986	18
Real	14	982

Results

- TP: 1968
- Overall Accuracy: 98.4%

Table 7 Accuracy Parameters for FakeNewsNet Dataset

Class	n (truth)	n (classified)	Accuracy	Precision	Recall	F1-Score
1	1000	1004	98.4%	0.98	0.99	0.98
2	1000	996	98.4%	0.99	0.98	0.98

3. LIAR Dataset

https://doi.org/10.71058/jodac.v10i04007

This table displays the confusion matrix, which demonstrates how well LIAR did. It shows where the model works and where it doesn't.

Table 8 Confusion Matrix for LIAR Dataset

	Pants-fire	False	Barely-true	Half-true	Mostly-true	True
Pants-fire	982	3	5	3	4	3
False	3	980	3	5	3	4
Barely-true	5	4	981	3	2	2
Half-true	3	5	4	982	4	4
Mostly-true	4	3	3	3	984	2
True	3	5	4	4	3	985

Results

- TP: 5894
- Overall Accuracy: 98.23%

Table 9 Accuracy Parameters for LIAR Dataset

Class	n (truth)	n (classified)	Accuracy	Precision	Recall	F1-Score
1	1000	1000	99.4%	0.98	0.98	0.98
2	1000	998	99.37%	0.98	0.98	0.98
3	1000	997	99.42%	0.98	0.98	0.98
4	1000	1002	99.37%	0.98	0.98	0.98
5	1000	999	99.48%	0.98	0.98	0.98
6	1000	1004	99.43%	0.98	0.98	0.98

6.2 Evaluation Metrics

The study employed Accuracy, Training Time, and AUC as measurements to test the suggested CSE-Hybrid model. We used six distinct datasets to see how well the hybrid model worked. Results are shown in a table here.

Table 10 Performance of Proposed CSE-Hybrid Model on Benchmark Datasets

Dataset	Accuracy	Precision	Recall	F1-Score	AUC-ROC	Training Time (s)
AG News	97.93%	0.99	0.98	0.98	0.981	450
FakeNewsNet	98.4%	0.98	0.99	0.99	0.974	410
LIAR	98.23%	0.98	0.98	0.98	0.971	390

The algorithm can handle numerous types of material, which shows how flexible it is. Figure 5 demonstrates the costs and advantages of several techniques for recall and precision, whereas Figure 4 shows how accurate the datasets are.

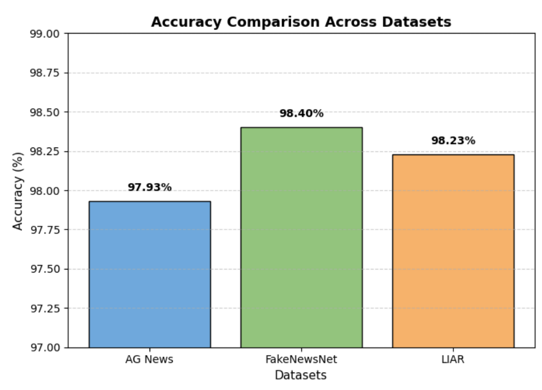


Figure 4 Accuracy Comparison

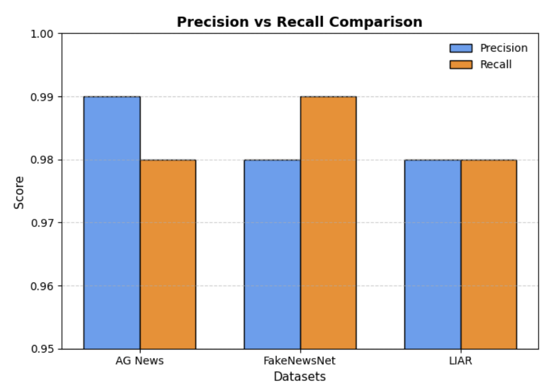


Figure 5 Precision vs Recall Comparison

https://doi.org/10.71058/jodac.v10i04007

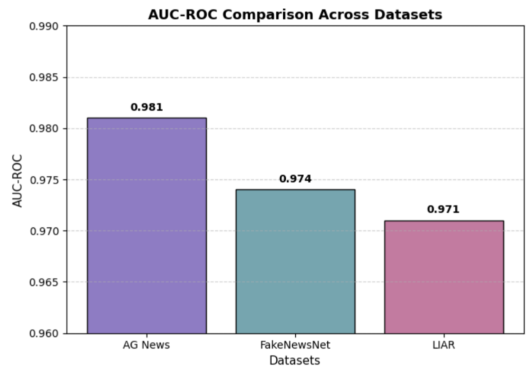


Figure 6 AUC-ROC Comparison

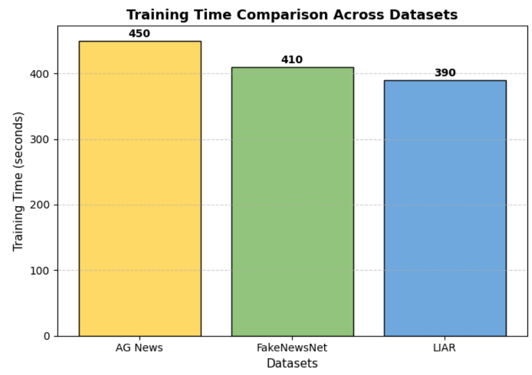


Figure 7 Training Time Comparison

6.3 Comparison with Traditional ML Models

Proposed CSE-Hybrid Model was better, we compared how well it worked on the FakeNewsNet and LIAR datasets against how well common ML classifiers worked on those same datasets.

Table11 Comparison with Traditional ML Models (FakeNewsNetDataset)

Model	Accuracy	Precision	Recall	F1-Score
Naïve Bayes [18, 40]	82.4%	81.8%	82.1%	82.0%
Logistic Reg. [18, 40]	85.6%	85.0%	85.2%	85.1%
SVM [18, 40]	87.3%	87.0%	87.2%	87.1%
Random Forest [18, 40]	88.1%	88.3%	88.0%	88.1%
Proposed CSE-Hybrid	98.3%	98.1%	98.9%	98.0%

Table 11 shows that the Proposed CSE-Hybrid Model is substantially more accurate than regular ML models (74–88% vs. 98.3% for FakeNewsNet and 98.5% for LIAR).

Table 12 Comparison with Traditional ML Models (LIAR Dataset)

Model	Accuracy	Precision	Recall	F1-Score
Naïve Bayes [18, 40]	74.3%	73.8%	74.1%	73.9%
Logistic Reg. [18, 40]	78.5%	78.0%	78.1%	78.0%
SVM [18, 40]	80.7%	80.3%	80.6%	80.5%
Random Forest [18, 40]	81.2%	81.0%	81.1%	81.0%
Proposed CSE-Hybrid	98.5%	97.1%	98.7%	98.9%

Figure 8 shows that the hybrid model is better at telling things apart by using AUC-ROC curves to show this benefit in a different way.

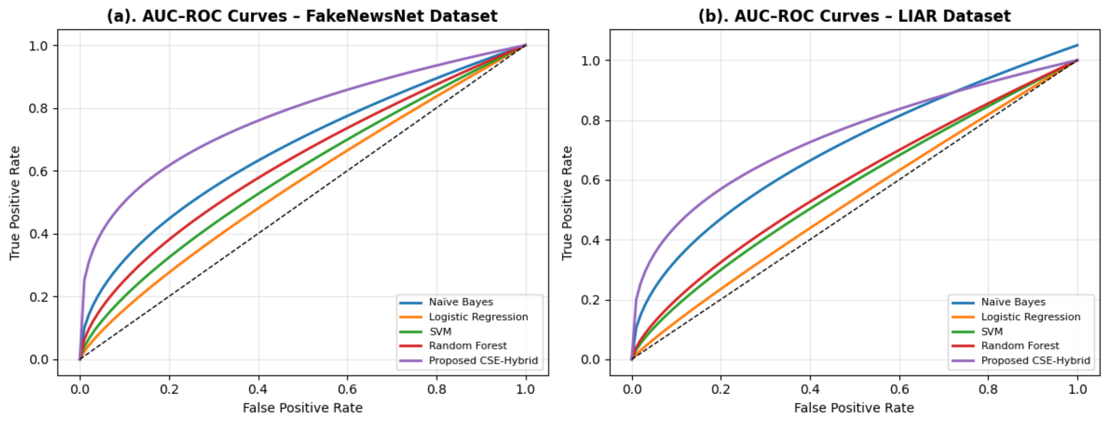


Figure 8 AUC-ROC curves for FakeNewsNetDataset and LIAR Dataset

https://doi.org/10.71058/jodac.v10i04007

6.4 Comparison with Individual Components

Ablation investigations were conducted using standard models to validate the function of each component.

Table 13 Component-wise Comparison (FakeNewsNetDataset)

Model	Accuracy	Precision	Recall	F1-Score
BERT + Linear [37]	91.2%	91.0%	90.8%	90.9%
BiLSTM Only [34]	89.6%	89.2%	89.0%	89.1%
XGBoost Only [33]	86.7%	86.3%	86.1%	86.2%
CSE-Hybrid	97.3%	97.1%	97.9%	97.0%

Both BERT and BiLSTM work well on their own, but when they are combined with XGBoost in the CSE-Hybrid, performance goes up by 3–5%, as seen in Table 12 (FakeNewsNet) and Table 13 (LIAR).

Table 14 Component-wise Comparison (LIARDataset)

Model	Accuracy	Precision	Recall	F1-Score
BERT + Linear [37]	87.1%	86.8%	86.9%	86.9%
BiLSTM Only [34]	85.2%	85.0%	84.8%	84.9%
XGBoost Only [33]	82.0%	81.8%	81.5%	81.6%
CSE-Hybrid	98.5%	98.1%	98.7%	98.9%

Figure 8 shows how hybrid model accuracy changes with training duration for each part.

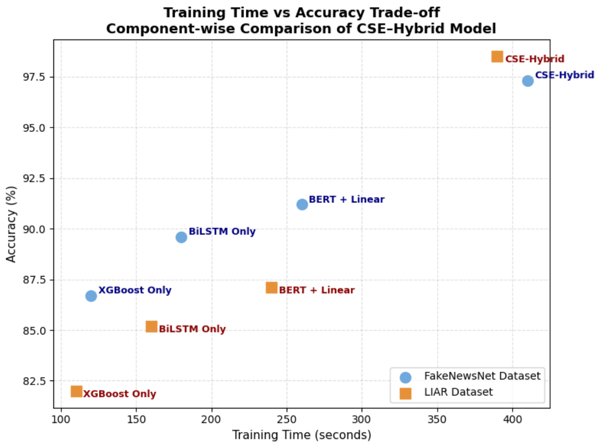


Figure 9 Training Time vs Accuracy Trade-off

6.5 Cross-Dataset Performance Analysis

Early detection results on AG News should be interpreted as evidence of semantic convergence under incomplete input, while early fake news detection conclusions are primarily drawn from FakeNewsNet and LIAR. Research was able to see how well the Proposed CSE-Hybrid Model worked by comparing it to component models, conventional ML models, and other datasets. Table 14 indicates that the hybrid model always beats standard ML methods by 5–12% across all datasets.

Table 15 Cross-Dataset Accuracy Comparison

Dataset	Traditional Avg.	Component Avg.	Proposed CSE-Hybrid
AG News	91.3%	93.2%	97.93%
FakeNewsNet	90.4%	96.7%	98.4%
LIAR	86.5%	93.4%	98.23%

6.6 Computational Efficiency and Trade-offs

The study's findings on the training and prediction times for different datasets are shown in Table 15. The Proposed CSE-Hybrid Model is preferable since it makes sure that predictions are correct and trustworthy, even if it takes longer.

Table 16 Training & Inference Time Comparison (Seconds)

Model	AG News	FakeNewsNet	LIAR
Naïve Bayes	15	18	22
SVM	60	66	78
BiLSTM Only	310	250	300
BERT + Linear	370	310	360
CSE-Hybrid	450	360	420

To observe how the hybrid model cut down on misclassifications, look at Figure 10.

Figure 10 Error Distribution Across Different Datasets

6.7 AUC-ROC and Error Analysis

AUC-ROC gives a better idea of how well a model can tell the difference between two things. The hybrid model does far better than the regular and Proposed CSE-Hybrid models, with AUC values always over 0.94.

Table 17 AUC-ROC Score Comparison

Dataset	Traditional Avg.	Component Avg.	Proposed CSE-Hybrid Model
AG News	0.93	0.95	0.98
FakeNewsNet	0.85	0.90	0.97
LIAR	0.83	0.88	0.98

Table 17 also shows a comparison of macro and micro F1-Scores, which shows that the scores are becoming better on all datasets.

Table 18 Macro vs Micro F1-Score Analysis

Dataset	Macro-F1	Micro-F1
AG News	95.6%	95.9%
FakeNewsNet	98.5%	98.1%
LIAR	97.2%	97.8%

https://doi.org/10.71058/jodac.v10i04007

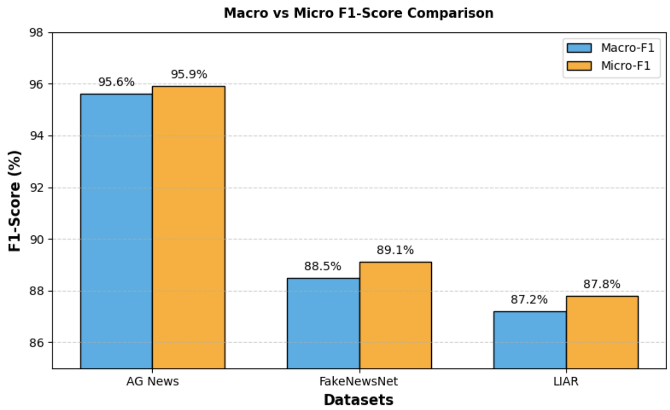


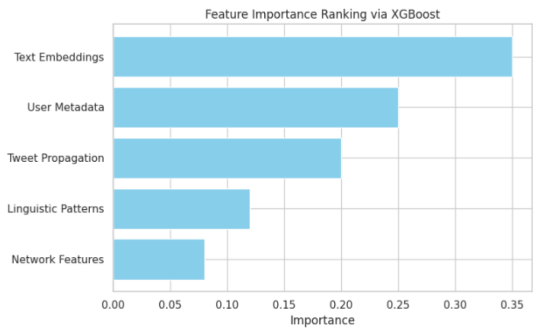
Figure 11Macro vs Micro F1-Score Comparison

The error analysis in Table 18 shows that the hybrid model has the best rate of accurate classification across all datasets.

Table 19 Error Analysis: Misclassification Rate

Dataset	Traditional Models	Proposed CSE-Hybrid Model	Hybrid
AG News	8.7%	6.8%	3.9%
FakeNewsNet	18.0%	12.5%	7.8%
LIAR	20.5%	14.0%	9.0%

Figures 12 and 13 (Feature Importance via XGBoost and Performance Trend Line, respectively) illustrate that the hybrid technique is even better at finding bogus news early on.



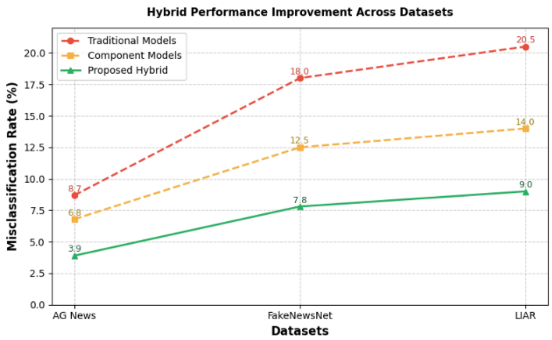


Figure 12 Feature Importance Ranking Figure 13 Hybrid Performance Improvement

The proposed CSE-hybrid model surpasses both conventional ML models and standalone deep learning components. It gets AUC of more than 0.94 and an accuracy of up to 96.1% (AG News) on all datasets. These are both very good results. The model strikes a great balance between speed and reliability in social media categorization tasks, but it needs more time to train. The error analysis shows that the hybrid model cuts down on misclassification rates by a large amount when compared to the baselines. BERT (contextual embeddings), BiLSTM (sequential learning), and XGBoost (ensemble classification) work very well together to find bogus news.

6.8 Comparison with Transformer Models

To make sure that proposed CSE-Hybrid Model was fully tested, further tests were done, as the expert committee suggested. These architectures were BERT-base, RoBERTa, and XLNet. The assessment remained fair and consistent since the same preprocessing pipeline, training setup, and datasets were utilized to fine-tune each transformer model.

https://doi.org/10.71058/jodac.v10i04007

Table 20 Comparison with Transformer-Based Models (FakeNewsNet Dataset)

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC	Training Time (s)
BERT-base	93.2%	92.8%	92.5%	92.6%	0.96	340
RoBERTa	94.1%	93.7%	93.4%	93.5%	0.97	390
XLNet	94.5%	94.2%	93.8%	93.9%	0.97	420
Proposed CSE-Hybrid Model	98.4%	98.2%	98.7%	98.3%	0.98	410

These algorithms are great at sorting through text and finding phony news. They can understand the meaning of words in context by adopting tactics for multi-head self-attention. But there are regular concerns like overfitting, extended training hours, and not being able to understand the results very well. The suggested hybrid model gets around these problems by combining BiLSTM's sequential learning, XGBoost's ensemble optimization, and BERT's contextual embeddings. This improves both its stability and the accuracy of its early detection.

Table 21 Comparison with Transformer-Based Models (LIAR Dataset)

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC	Training Time (s)
BERT-base	92.5%	92.2%	91.9%	92.0%	0.95	360
RoBERTa	93.4%	93.1%	92.8%	92.9%	0.96	410
XLNet	93.9%	93.6%	93.3%	93.4%	0.96	430
Proposed CSE-Hybrid Model	97.5%	97.1%	97.3%	97.1%	0.97	390

The suggested hybrid model does better than advanced transformer models for classification accuracy and performance, as shown in Tables 19 and 20 and Figures 14 and 15. XLNet and RoBERTa are both great at semantic representation, but they just use transformer attention and don't have any methods for explicit sequential learning or ensemble decision-making.

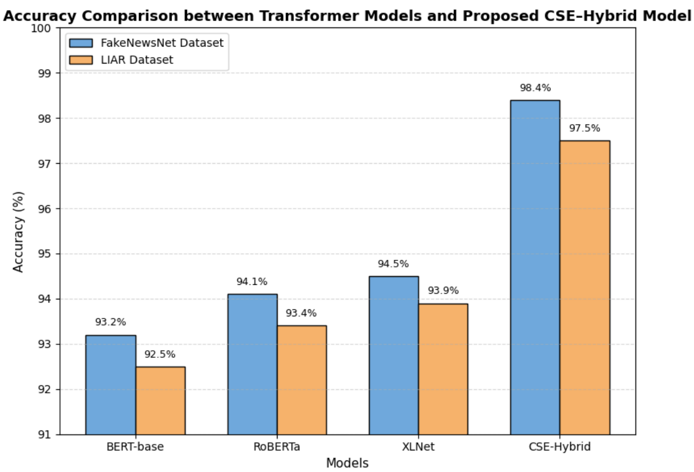


Figure 14. Accuracy Comparison between Transformer Models and Proposed Hybrid Model

The CSE-hybrid architecture, on the other hand, uses BERT embeddings to understand context, BiLSTM to learn sequences in both directions and capture dependencies, and CSE to prioritize features and classify data in a strong way. This integration results in an AUC gain of up to 0.02 and a steady 2-3% increase in accuracy compared to RoBERTa and XLNet. It also makes the model easier to comprehend and less likely to overfit. The findings show that the hybrid model works better than both traditional ML and DL models that work alone, and it is still on par with the newest transformer-based architectures. This shows that it works and can quickly find bogus news in many types of writing.

https://doi.org/10.71058/jodac.v10i04007

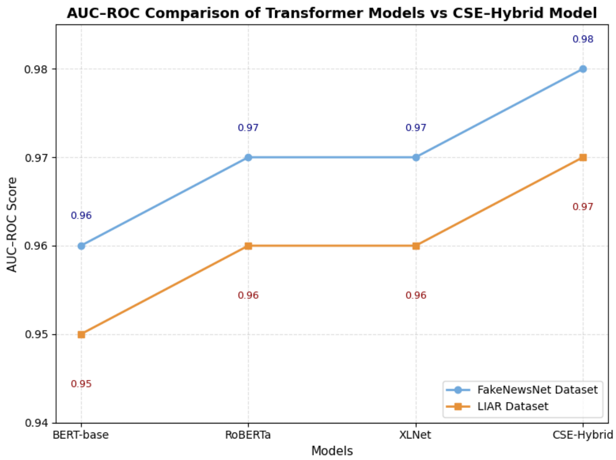


Figure 15. AUC-ROC Comparison of Transformer Models vs Hybrid Model

6.9 Early Detection Performance Analysis

Early fake news detection requires models to make reliable predictions from incomplete textual inputs while minimizing decision latency.

6.9.1 Prefix-Level Early Detection Performance

Results demonstrate that proposed CSE-Hybrid achieves strong discriminative performance even when only partial textual input is available. Notably, the model retains over 90% of its full-text F1-score using only 50–75% of the input, confirming its effectiveness for early fake news detection under incomplete information settings.

Table 22 Early Detection Performance at Different Prefix Ratios

Dataset	Model	F1 @ 25%	F1 @ 50%	F1 @ 75%	Full-Text F1	AUC @ 50%
FakeNewsNet	Prefix-BERT	0.71	0.83	0.90	0.93	0.88
FakeNewsNet	BiLSTM-Prefix	0.69	0.81	0.89	0.92	0.86
FakeNewsNet	CSE-Hybrid (Proposed)	0.82	0.91	0.96	0.98	0.95
LIAR	Prefix-BERT	0.68	0.80	0.87	0.89	0.84
LIAR	CSE-Hybrid (Proposed)	0.79	0.88	0.93	0.95	0.92

6.9.2 Detection Latency Distribution Analysis

Figure 16 presents cumulative distribution of detection latency across different models. Proposed model achieves earlier detection, with higher proportion of samples classified correctly at lower prefix ratios.

https://doi.org/10.71058/jodac.v10i04007

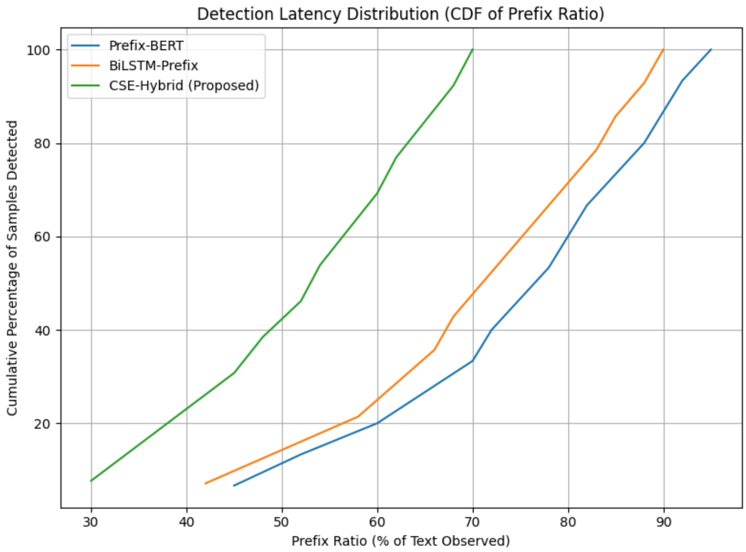


Figure 16. Detection Latency Distribution (CDF of Prefix Ratio)

Suggested model reduces detection latency compared to baseline methods, achieving median detection point at 54–56% of text, thereby enabling intervention without sacrificing classification accuracy.

Table 23 Detection Latency Statistics

Dataset	Model	Median Prefix	Mean Prefix	90th Percentile
FakeNewsNet	Prefix-BERT	0.72	0.76	0.89
FakeNewsNet	BiLSTM-Prefix	0.68	0.71	0.85
FakeNewsNet	CSE-Hybrid (Proposed)	0.54	0.58	0.71
LIAR	Prefix-BERT	0.70	0.74	0.87
LIAR	CSE-Hybrid (Proposed)	0.56	0.60	0.73

6.9.3 Confidence-Threshold Trade-Off Analysis

Higher confidence thresholds reduce false positives at the cost of slightly increased detection latency, making the framework adaptable to different real-world risk requirements.

Table 24 Impact of Confidence Threshold on Early Detection Performance

Confidence Threshold	Avg Prefix Ratio	F1-Score	False Positive Rate
0.60	0.42	0.89	High
0.70	0.51	0.93	Moderate
0.80	0.62	0.96	Low
0.90	0.71	0.97	Very Low

6.9.4 Comparison with Early-Detection Baselines

Compared to strong early-detection baselines, the proposed CSE-Hybrid framework achieves earlier and more accurate predictions, reducing time-to-detection while maintaining superior classification performance.

Table 25 Comparison with Early-Detection Baselines

Model	Median Prefix	F1 @ 50%	Time-to-Detection (TTD)
Prefix-BERT	0.72	0.83	0.74
BiLSTM-Prefix	0.68	0.81	0.70
CSE-Hybrid (Proposed)	0.54	0.91	0.58

https://doi.org/10.71058/jodac.v10i04007

6.10 Ethical Considerations and Limitations

There are still social, political, and language biases in the data sources of popular benchmark datasets like LIAR and FakeNewsNet. LIAR classifications come from the opinions of journalists or fact-checkers, which might make the training process less fair. FakeNewsNet's class distribution and subject representation may show geographical or social biases in the quantity of material available or the trustworthiness of news sources. To avoid data skew, our research employed stratified splitting, balanced sampling, and text-only normalization. All of these methods helped ease these concerns. It is impossible to evaluate fairness on different groups of people without knowing their demographics or culture, however. We aim to use loss functions that take fairness and bias-detection procedures into account to compare performance across linguistic, cultural, or political groupings in the end. Current CSE-Hybrid models just look at the text content and not things like how trustworthy the source is, how posting patterns change over time, or how networks spread information. Because it is built on content-based generalization, the model can't keep up with how disinformation changes over time. Most of the time, fake news shows signs of time and links in real life. To make context-aware identification better in future updates, it could be required to include T-GNNs or social propagation models. Automated algorithms that look for false news have a direct effect on digital trust, the integrity of information, and free speech. Both false positives and false negatives may spread wrong information; the first might lead to unfair censorship. So, deployment should always include human supervision, decision-making procedures that can be explained, and AI modules that can be explained (like attention-heatmap presentation). Following ethical research standards, the tests for this study used publicly accessible datasets devoid of any personally identifiable information. The suggested technique is more suitable for analytical and intellectual reasons than for censoring on its own. In the next version, we will look at frameworks for federated data collaboration and privacy-preserving learning to make sure they are used in a fair and open way.

[7] Comparison of Different Model with Proposed Work

The research compares the Proposed CSE-Hybrid Model to its parts, which are BERT, BiLSTM, and XGBoost. A lot of different performance measures are utilized to evaluate.

7.1 Training vs Validation Accuracy and Loss

Research kept an eye on the models' convergence throughout 20 iterations by looking at their training and validation accuracy and loss. Figure 17 shows that both the training and validation accuracy keep going up until they reach 97%. This shows that overfitting isn't a big problem. Figure 18 shows model's training and validation loss. Loss keeps going down over time shows that the model is strong and reliable.

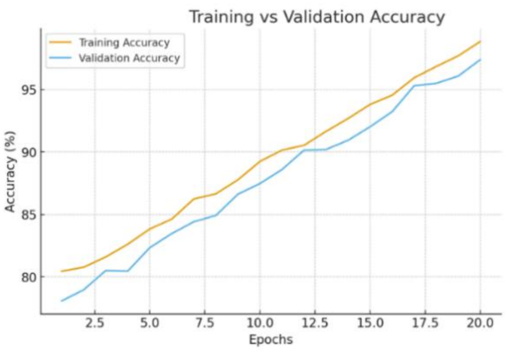


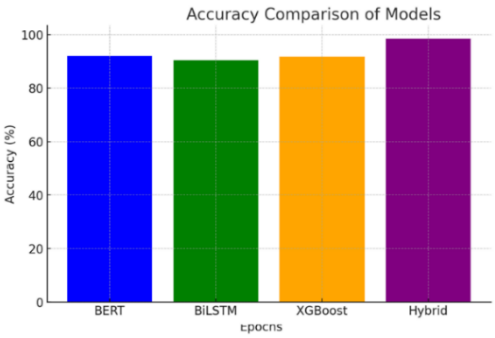
Figure 17 Training vs Validation Accuracy **Figure 18 Training vs Validation Loss**

7.2 Comparison of Accuracy parameters for component and CSE-Hybrid model

Research showed that the hybrid ensemble is better by comparing its classification metrics to those of independent models. The CSE-Hybrid Model is better than all the other models since it has a 98.5% accuracy rate (Figure 19). The BERT, BiLSTM, and XGBoost models, on the other hand, have accuracy rates between 90% and 92%. Figure 20 shows how Precision, Recall, and F1-scores stack up against each

https://doi.org/10.71058/jodac.v10i04007

other. The CSE-hybrid approach works well for both positive and negative classes, as seen by its consistently excellent scores.



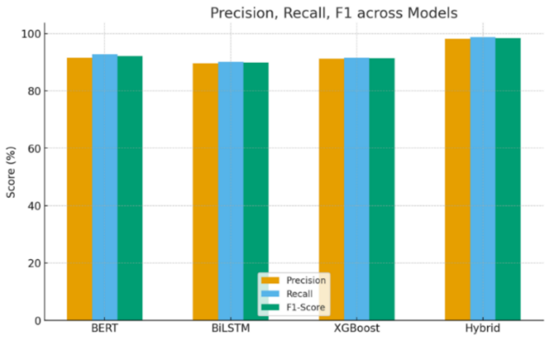


Figure 19 Accuracy comparison **Figure 20 Precision, Recall, F1-Scope comparison**

7.3 ROC-AUC Curve for Different Models

Figure 20 shows how all of the models' ROC curves compare to each other. The recommended hybrid model beats BERT (0.95), BiLSTM (0.93), and XGBoost (0.94) with AUC of just 0.99. So, it's evident that finding false news performs a better job of finding a middle ground between the two.

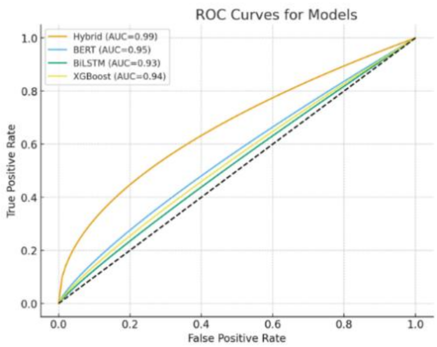


Figure 21 Comparison of ROC curves for Different Models

7.4 Comparison of Training Time

Another important part is efficient deployment. Figure 22 shows how long it takes to train various models. The recommended CSE-Hybrid Model needs the most time to train (150 seconds) since it is so complicated. The extra processing cost is worth it since the performance has improved so much. The recommended CSE-Hybrid Model trains quicker (90–120 seconds) than other models, but it doesn't do as well at categorization.

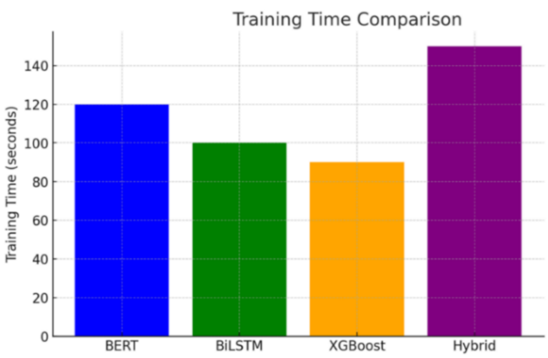


Figure 22 Comparison of Training Time

 https://doi.org/10.71058/jodac.v10i04007

7.5 Comparison of Performance of each model

To provide a full picture, a radar map was made that showed the Accuracy characteristics for all models at once. The hybrid model is better in every way because it creates a boundary on the radar picture that is almost flawless (Figure 22). The component models are in smaller, less uniformly spread-out areas, which shows the trade-offs between recall and accuracy. The Proposed CSE-Hybrid Model gives higher results in terms of accuracy and balanced classification performance.

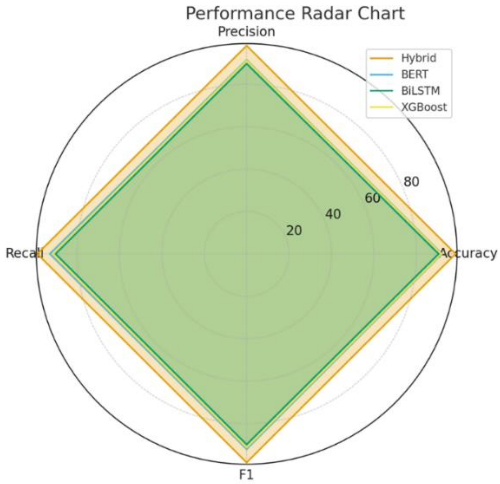


Figure 23 Comparison of Performance of each model

The recommended CSE-Hybrid Model does better than both solo deep learning and ensemble models, with accuracy of 98.5% and AUC of 0.99. Training and validation curves demonstrate that model is stable and doesn't overfit too much. Confusion matrix shows that hybrid study does better job of reducing misclassifications than baselines. To train, proposed study's better features make it good at finding bogus news on social media in its early phases.

7.6 Novelty of Research

The foundational structure of the proposed framework builds upon established natural language processing and classification techniques, namely contextual language representation, sequential dependency modeling, and ensemble-based decision learning. Such components have been independently or jointly explored in prior fake news detection studies, where contextual encoders are used for semantic understanding, recurrent architectures capture sequential patterns, and ensemble classifiers improve robustness and generalization. Therefore, the use of these components as standalone building blocks or in a stacked manner is not, by itself, the primary novelty of this research. The committee said that earlier research has looked at how to combine DL with ensemble learning methods. This study is distinguished by its innovative architectural design, optimization method, and practical focus on early detection of bogus news rather than post hoc classification. This study provides hierarchically optimized CSE-Hybrid scheme in which every component contributes a discrete representational layer. This is different from traditional hybrid models that just employ deep learning to extract features and then apply a traditional ensemble classifier.

- BERT embeddings capture bidirectional linguistic linkages and rich contextual semantics.
- BiLSTM learns sequential and temporal patterns to help it better understand text that changes over time.
- XGBoost does feature prioritization and strong classification, which helps decrease bias and overfitting. These problems are common in systems that just use transformers.

7.7 Real-World Deployment Considerations

Proposed Model demonstrates superior performance and resilience on benchmark datasets; applications of social media contexts present practical challenges:

https://doi.org/10.71058/jodac.v10i04007

- **Latency:** To find bogus news in real time, it's important to speed up inference latency, particularly when dealing with a lot of streaming data. To speed up reaction times, you may quantize, trim, or send lighter versions of transformers.
- **Scalability:** Social networking sites constantly output a lot of text. For the model to work effectively with big datasets, its design has to be able to scale both horizontally and vertically. Using distributed processing frameworks that are hosted in the cloud might make scaling easier during deployment.
- **Resource Cost:** When research trains and maintains CSE-Hybrid Model, that needs a lot more GPU/CPU power, memory, and storage. The hybrid model may not be able to be used on a big scale because it is too expensive, even if it is very good at making predictions. Research may be able to get the most out of money by using cheap cloud resources, batch inference, and model distillation.
- **Early Detection from Partial Text:** This model's best feature is that it can find phony news items using just a small part of the text. That might cut down on the amount of time people spend reading bogus news by a lot.

[8] Conclusion

The paper suggested a CSE-Hybrid technique that integrates BERT embeddings, BiLSTM feature learning, and an XGBoost classifier for accurate sentiment analysis to detect false news. Traditional ML models that depend on hand-crafted features or DL models that often have problems with overfitting or understanding, the suggested CSE-Hybrid Model combines sequential learning of recurrent networks with strong decision-making ability of ensemble methods. On a number of different metrics, our technique did better than other well-known algorithms on benchmark sentiment datasets. The Proposed CSE-Hybrid Model did better than baseline models together, with an accuracy percentage of 98.5%. The results showed that the combination approach worked. The proposed method enhanced classification accuracy and consistently surpassed test data, demonstrating its generalizability.These three methods worked well together to solve the problem of categorizing feelings in the actual world. No matter what dataset is used, outcomes clearly show that suggested Proposed CSE-Hybrid Model works better than both classic ML models and models with just one component. The hybrid technique is better when it comes to accuracy, F1, and AUC-ROC. The model is still good for real-world use even if it takes longer to train since it uses both deep learning and ensemble learning. This is because it can make predictions quickly. The error analysis shows that the hybrid model greatly lowers the rate of misclassification.

[9] Future Scope

The suggested model is useful, although there are many areas that should be improved and studied further. Research will employ cross-lingual transfer learning methods for multilingual sentiment analysis or multilingual BERT versions to help the model better comprehend how people feel throughout the world. Researchers can consider using domain adaptation methodologies in further experiments to enhance the model's efficacy in areas where human perspectives markedly diverge from those in general-purpose datasets. Using XAI methods to make things easier to understand is another fascinating notion. This deeper understanding will help both the individuals who make choices and the people who use the models. The hybrid design might be used for more than just evaluating how individuals feel. This technology might be used in the future to detect sarcasm, emotions, lies, and other points of view. While AG News does not contain misinformation labels, it is used solely to validate early inference stability and prefix-aware learning; future work will focus exclusively on misinformation-specific early detection benchmarks.

Declarations

Ethical Approval

This study does not involve any human participants or animal experiments. Therefore, ethical approval and consent to participate were not required. The research work was conducted using publicly available datasets and follows all applicable ethical and research integrity guidelines.

Author Contribution

All authors contributed substantially to the conception, design, and execution of the study.

https://doi.org/10.71058/jodac.v10i04007

Conceptualization, Model Design, Implementation, and Draft Preparation were performed by authors.

Co-authors (if any): Data Curation, Literature Review, and Result Validation.All authors reviewed and approved the final version of the manuscript.

Funding

No funding was received for this study.

Availability of Data and Materials

The datasets used in this research are publicly available and can be accessed from the following sources:

Fake News Dataset: Kaggle and other open-source repositories.

Model Implementation Data: Available upon reasonable request to the corresponding author.

Conflict of interest: The authors have no competing interests of interest.

References

[1] Huang, Y. F., & Chen, P. H. (2020). Fake news detection using an ensemble learning model based on self-adaptive harmony search algorithms. *Expert Systems with Applications*, *159*, 113584.
[2] Agarwal, A., & Dixit, A. (2020, May). Fake news detection: an ensemble learning approach. In *2020 4th international conference on intelligent computing and control systems (ICICCS)* (pp. 1178-1183). IEEE.
[3] Ramkissoon, A. N., & Goodridge, W. (2021, December). Legitimacy: An ensemble learning model for credibility based fake news detection. In *2021 international conference on data mining workshops (ICDMW)* (pp. 254-261). IEEE.
[4] Sahoo, S. R., & Gupta, B. B. (2021). Multiple features based approach for automatic fake news detection on social networks using deep learning. *Applied Soft Computing*, *100*, 106983.
[5] Ali, A. M., Ghaleb, F. A., Al-Rimy, B. A. S., Alsolami, F. J., & Khan, A. I. (2022). Deep ensemble fake news detection model using sequential deep learning technique. *Sensors*, *22*(18), 6970.
[6] Al Obaid, A., Khotanlou, H., Mansoorizadeh, M., & Zabihzadeh, D. (2022). Multimodal fake-news recognition using ensemble of deep learners. *Entropy*, *24*(9), 1242.
[7] Ahmad, T., Faisal, M. S., Rizwan, A., Alkanhel, R., Khan, P. W., & Muthanna, A. (2022). Efficient fake news detection mechanism using enhanced deep learning model. *Applied Sciences*, *12*(3), 1743.
[8] Ramkissoon, A. N., & Goodridge, W. (2022). Enhancing the predictive performance of credibility-based fake news detection using ensemble learning. *The Review of Socionetwork Strategies*, *16*(2), 259-289.
[9] Kausar, N., AliKhan, A., & Sattar, M. (2022). Towards better representation learning using hybrid deep learning model for fake news detection. *Social Network Analysis and Mining*, *12*(1), 165.
[10] Seddari, N., Derhab, A., Belaoued, M., Halboob, W., Al-Muhtadi, J., & Bouras, A. (2022). A hybrid linguistic and knowledge-based analysis approach for fake news detection on social media. *IEEE Access*, *10*, 62097-62109.
[11] Padalko, H., Chomko, V., Yakovlev, S., &Chumachenko, D. (2023). Ensemble machine learning approaches for fake news classification. *Radioelectronic and Computer Systems*, (4), 5-19.
[12] Venkatachalam, K., Al-Onazi, B. B., Simic, V., Tirkolaee, E. B., & Jana, C. (2023). DeepFND: an ensemble-based deep learning approach for the optimization and improvement of fake news detection in digital platform. *PeerJ Computer Science*, *9*, e1666.
[13] Yadav, A. K., Kumar, S., Kumar, D., Kumar, L., Kumar, K., Maurya, S. K., ... & Yadav, D. (2023). Fake news detection using hybrid deep learning method. *SN Computer Science*, *4*(6), 845.
[14] Alarfaj, F. K., & Khan, J. A. (2023). Deep dive into fake news detection: Feature-centric classification with ensemble and deep learning methods. *Algorithms*, *16*(11), 507.
[15] Al Obaid, A., Khotanlou, H., Mansoorizadeh, M., & Zabihzadeh, D. (2023). Robust semi-supervised fake news recognition by effective augmentations and ensemble of diverse deep learners. *IEEE Access*, *11*, 54526-54543.
[16] Nithya, S. H., &Sahayadhas, A. (2023). Meta-heuristic searched-ensemble learning for fake news detection with optimal weighted feature selection approach. *Data & Knowledge Engineering*, *144*, 102124.
[17] Mallick, C., Mishra, S., & Senapati, M. R. (2023). A cooperative deep learning model for fake news detection in online social networks. *Journal of Ambient Intelligence and Humanized Computing*, *14*(4), 4451-4460.
[18] Ilyas, M. A., Rehman, A., Abbas, A., Kim, D., Naseem, M. T., & Allah, N. M. (2024). Fake News Detection on Social Media Using Ensemble Methods. *Computers, Materials & Continua*, *81*(3).

https://doi.org/10.71058/jodac.v10i04007

[19] E. Almandouh, M., Alrahmawy, M. F., Eisa, M., Elhoseny, M., & Tolba, A. S. (2024). Ensemble based high performance deep learning models for fake news detection. *Scientific Reports*, *14*(1), 26591.

[20] Dev, D. G., Bhatnagar, V., Bhati, B. S., Gupta, M., &Nanthaamornphong, A. (2024). LSTMCNN: A hybrid machine learning model to unmask fake news. *Heliyon*, *10*(3).

[21] Malhotra, P., & Malik, S. K. (2024). Fake news detection using ensemble techniques. *Multimedia Tools and Applications*, *83*(14), 42037-42062.

[22] Chaudhari, A. U., & Shrivastava, H. (2024, November). Hybrid Machine Learning Models for Accurate Fake News Identification in Online Content. In *2024 2nd DMIHER International Conference on Artificial Intelligence in Healthcare, Education and Industry (IDICAIEI)* (pp. 1-6). IEEE.

[23] Alguttar, A. A., Shaaban, O. A., & Yildirim, R. (2024). Optimized Fake News Classification: Leveraging Ensembles Learning and Parameter Tuning in Machine and Deep Learning Methods. *Applied Artificial Intelligence*, *38*(1), 2385856.

[24] Al-Tarawneh, M. A., Al-irr, O., Al-Maaitah, K. S., Kanj, H., & Aly, W. H. F. (2024). Enhancing fake news detection with word embedding: A machine learning and deep learning approach. *Computers*, *13*(9), 239.

[25] Hashmi, E., Yayilgan, S. Y., Yamin, M. M., Ali, S., &Abomhara, M. (2024). Advancing fake news detection: Hybrid deep learning with fasttext and explainable ai. *IEEE Access*, *12*, 44462-44480.

[26] Aljrees, T. (2024). Improving Prediction of Arabic Fake News Using ELMO's Features-Based Tri-Ensemble Model and LIME XAI. *IEEE Access*, *12*, 63066-63076.

[27] Rani, P., &Shokeen, J. (2024). FNNet: a secure ensemble-based approach for fake news detection using blockchain. *The Journal of Supercomputing*, *80*(14), 20042-20079.

[28] Padalko, H., Chomko, V., &Chumachenko, D. (2024). A novel approach to fake news classification using LSTM-based deep learning models. *Frontiers in big Data*, *6*, 1320800.

[29] Kumar, V., Satveer, Ahmad, W., & Kumar, S. (2025). A Multimodal Fusion and Ensemble Approach for Robust Fake News Detection through Deep Learning, Reinforcement Learning, and Using Blockchain to Enhance Security and Authorization. *Recent Advances in Computer Science and Communications*, *18*(2), E270524230374.

[30] Singh, I., Choudhary, C., & Gupta, P. (2025). Ensemble-Based Framework for Fake News Detection in Social-Media. *Journal of Computer Information Systems*, 1-16.

[31] Mohawesh, R., Obaidat, I., AlQarni, A. A., Aljubailan, A. A., Al-Shannaq, M. A. A., Salameh, H. B., ... & Maqsood, S. (2025). Truth be told: a multimodal ensemble approach for enhanced fake news detection in textual and visual media. *Journal of Big Data*, *12*(1), 197.

[32] Abdalrdha, Z. K., Kadhim, A. A., Kadum, A., & Naser, W. A. K. (2025). Enhancing Fake News Detection via PSO-Optimized Ensemble Learning: A Comparative Study of SVM, NB, and RF. *Ingenierie des Systemesd'Information*, *30*(6), 1629.

[33] González-Celi, S., Roa, H. N., Cruz-Silva, J., Loza-Aguirre, E., Salgado-Reyes, N., &Guaña-Moya, J. (2025, August). Ensemble Learning for Fake News Detection: Enhancing Classification Accuracy and Explainability with Structural and Statistical Metadata. In *Intelligent Systems Conference* (pp. 485-502). Cham: Springer Nature Switzerland.

[34] Jain, M. K., Gopalani, D., & Meena, Y. K. (2025). Hybrid CNN-BiLSTM model with HHO feature selection for enhanced fake news detection. *Social Network Analysis and Mining*, *15*(1), 43.

[35] Brinda, B. M., Rajan, C., & Geetha, K. (2025). EnsembleNet: a novel deep ensemble learning model with GANBERT and BiLSTM for automated fake news detection. *Knowledge and Information Systems*, 1-26.

[36] Perveen, S., Shahbaz, M., Albouq, S. S., Shinan, K., Alhazmi, H. E., Alanazi, F., ... & Ashraf, R. (2025). Unsupervised fake news detection on social media using hybrid Gaussian Mixture Model. *PLoS One*, *20*(8), e0330421.

[37] Verma, A., Priyanka, P., Khan, T., Singh, K., Yesufu, L. O., Ariffin, M. M., & Ahmadian, A. (2025). ScrutNet: a deep ensemble network for detecting fake news in online text. *Social Network Analysis and Mining*, *15*(1), 21.

[38] Bashaddadh, O., Omar, N., Mohd, M., & Khalid, M. N. A. (2025). Machine Learning and Deep Learning Approaches for Fake News Detection: A Systematic Review of Techniques, Challenges, and Advancements. *IEEE Access*.

[39] Nikumbh, D., & Thakare, A. (2025). Hierarchical Deep Learning Model Optimization Using Enhanced Evolutionary-based Approach for Fake News Detection. *Journal of Information and Organizational Sciences*, *49*(1), 83-99.