



Optimized bidirectional gated recurrent unit-convolutional neural network for enhanced social network security

T. Kamalakannan¹ · K. Parameswari¹

Received: 3 April 2025 / Accepted: 8 August 2025 / Published online: 4 September 2025
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025

Abstract

The rapid growth of social networks has significantly transformed digital communication, facilitating large-scale information sharing. However, this expansion has also introduced complex security risks, including phishing, spam, fake accounts, cyberbullying, AI-generated misinformation, and sophisticated bot networks. Traditional security approaches often fail to address the dynamic and multifaceted nature of these threats. To tackle these challenges, we introduce an Optimized Bidirectional Gated Recurrent Unit and Convolutional Neural Network (O-Bi-GRU-DCNN) model designed specifically for social network security. This model combines the sequential learning power of Bidirectional Gated Recurrent Units (Bi-GRU) to capture contextual dependencies in social media data with the feature extraction capabilities of Deep Convolutional Neural Networks (CNN). Additionally, Archimedes Optimization (ArO) is used to enhance the model's accuracy. Experimental results reveal that the proposed model outperforms existing security techniques in detecting malicious activities, spam, and fake accounts, achieving an impressive accuracy of 98.9%. This approach offers a robust and scalable solution to enhance security across social network platforms.

Keywords Social network security · Bidirectional gated recurrent unit · Archimedes optimization · Deep convolutional neural network

Highlights

- Optimized Bidirectional Gated Recurrent Unit-Convolutional Neural Network (O-Bi-GRU-DCNN) model is proposed for social network security.
- The suggested approach utilized the sequential learning abilities of Bi-GRU for analyzing context dependencies in social media data and Deep CNN is presented for spatial features extraction.
- ArO improves the accuracy of the approach.
- The experimental outcomes shows that the proposed suggested approach outperformed existing security models in malicious activities, spam, and fake accounts detection, and achieving higher accuracy of 98.9%.
- This work ensures a scalable and smart security model to provide a safer social network.

Statement of novelty This work presents an Optimized Bidirectional Gated Recurrent Unit-Convolutional Neural Network (O-Bi-GRU-DCNN) model for social network security. The suggested approach utilized the sequential learning abilities of Bidirectional Gated Recurrent Unit (Bi-GRU) for analyzing context dependencies in social media data and Deep Convolutional Neural Network (CNN) is presented for spatial features extraction. Furthermore, Archimedes optimization (ArO) improves the accuracy of the approach. The experimental outcomes shows that the proposed suggested approach outperformed existing security models in malicious activities, spam, and fake accounts detection, and achieving higher accuracy of 98.9%. This work ensures a scalable and smart security model to provide a safer social network.

Extended author information available on the last page of the article

1 Introduction

Recently, the fast emergence of social networks has revolutionized communication, makes massive and instant interactions that significantly enhance daily life. With respect to the Digital report of 2024, nearly 4.83 billion users globally access the Internet and account for a penetrating rate of 61.23%, with 5.3 billion active users in the social media (Huang et al. 2021). The facility of information spreading and the cost-effective accompanied with social networks have led to an exponential spread of news across these platforms (Sheng et al. 2022). However, this has also raised the fake news emergence, as misleading information tends to attract more attention and engagement. Fake news is purposely generated for deceiving audiences and is generally circulated on social media to manipulate public opinion. Hence, social data security is essential (Xing et al. 2022).

One of the approaches utilized by social media platforms for mitigating the spread of fake news and secure social data is issuing official disclaimers about the credibility of specific content (Jiang et al. 2020). But, this manual process requires more time and resources. This is impractical for massive scale evaluation. Hence, automated approaches like Machine Learning (ML) and Deep Learning (DL) models have become major to find

and mitigate spreading false news effectively (Yao et al. 2019). Weibo is the massive Chinese social network platform globally and many works indicate that microblogging sites are heavily saturated with false information. But, the majority of works mainly focus on English-language social media platforms like YouTube, Facebook, and Twitter. Presently, limited research has been performed on Chinese social media platforms, which utilize a significant gap in the field. Hence, there is a demand for developing an advanced approach capable of automatically and accurately protecting social data on Weibo for enhancing information security and credibility on the platform (Li 2023).

Many traditional models on social data security are majorly focused on analyzing textual content, exploits text based features and statistical analysis for identifying false information spreading on social media platforms. But, depending only on textual statistics is generally not enough to secure social data accurately. For detection accuracy enhancement, some approaches considered TF-IDF for semantic features extraction and classify them by ML models. In later developments, deep learning models like Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) networks, and Gated Recurrent Units (GRU) were introduced for extracting deeper textual representations and obtaining promising outcomes (Zhao et al. 2023). Some works considered attention modules for highlighting key information and apply them directly to inputs or their higher level representations.

1.1 Problem statement

Despite numerous advances in deep learning models for social network security, current approaches often suffer from limited adaptability to diverse platforms, low resilience to emerging threats like AI-generated misinformation and bot-driven attacks, and insufficient interpretability. Moreover, many methods depend heavily on either textual features or single-model architectures, which restricts their effectiveness in complex, dynamic online environments.

1.2 Key contributions

To address these challenges, this research introduces a hybrid deep learning model—Optimized Bidirectional Gated Recurrent Unit–Deep Convolutional Neural Network (O-Bi-GRU-DCNN)—with the following core contributions:

- Development of a hybrid model combining Bi-GRU and DCNN to jointly capture temporal dependencies and spatial features from social media content.
- Integration of Archimedes Optimization (ArO) to fine-tune model parameters and improve classification accuracy.
- Demonstration of high performance (98.9% accuracy) in detecting malicious content, spam, and fake accounts on the Sina Weibo dataset.
- Evaluation against multiple baselines, showing improved precision, recall, and F1-score, and suitability for real-time deployment.
- Discussion of practical implications in combating deep-fake misinformation, bot detection, and spam filtering across platforms.

This model provides a scalable, accurate, and adaptable solution for modern social network security challenges.

2 Related works

Wang et al. introduced a fake news identification model for Sina Weibo by considering a hybrid feature fused model (Wang et al. 2022). Here, sixteen distinct features were extracted for fake news identification. The correlation among text and images were analyzed for accuracy detection. For implementing fake news identification, the pre-trained model with Fake News Net (FNN) was presented.

Chen et al. introduced a Semi-Supervised Clue Fused (S-SCF) model to improve spammer identification on Sina Weibo (Chen et al. 2018). The S-SCF utilized a linear weighting term for integrating multiple key features obtained from various perspectives and ensured a more comprehensive analysis. The S-SCF classified unlabeled data by the initial set of labeled data.

Luo et al. employed a DL model for conducting content and emotion analysis (Luo et al. 2021). Then, an analytic network model was exploited to examine the variation in network diffusion behaviour among the two kinds of information. The analysis indicated that the spread of false news on social media was impacted by content attributes and emotional terms.

Zhang et al. presented a rapid fake news identification approach for cyber physical social service (Zhang et al. 2023). Focusing on Chinese text, all characters were considered as the fundamental process unit. Here, the CNN was used for extracting representation of features. High processing speed and better detection capabilities were provided by this existing work.

Wu et al. introduced a DABot, for detecting social bots on Sina Weibo using DL model (Wu et al. 2021). The approach involved extracting thirty features and among these, nine features were newly presented. Moreover, active learning was combined for enhancing the labeled data expansion efficiency. For bot detection, a Residual

Network (ResNet) with the attention based BiGRU was used for performance enhancement.

Li et al. presented a sentiment analysis with respect to the Graph Neural Network (GNN) (Li and Li 2022). For processing all comment texts, dependency syntax was exploited for constructing a semantic graph showing the short text. Then, the LSTM was used for reducing node noise. Here, the GNN was a semantic parser for encoding the syntactic dependency tree and enables semantic and syntactic features extraction.

Recent advancements in Online Social Network (OSN) security have introduced diverse approaches, each with specific strengths and limitations. Boahen et al. (2024) proposed ASRL, an Adaptive Swarm Reinforcement Learning model that achieved high accuracy in intrusion detection but lacks adaptability to evolving threats and scalability for large-scale OSN data (Boahen et al. 2024). Chen et al. (2025) introduced a Sensitivity-Aware Personalized Differential Privacy (SPEDP) mechanism to protect user data; however, it focuses on privacy rather than active threat detection and relies on manual sensitivity labels (Chen et al. 2025). Yang et al. (2024) enhanced intrusion detection through an improved Social Network Search (SNS) algorithm coupled with a BP neural network, but the absence of temporal modeling and reliance on shallow architectures limited its robustness (Yang et al. 2024). Boahen et al. (2023) developed a deep learning model for detecting compromised accounts based on behavioral patterns, yet it lacked integration of content-based analysis and struggled with temporal drift (Boahen et al. 2023). In contrast, our proposed O-Bi-GRU-DCNN model addresses these limitations by combining bidirectional GRUs for sequential learning, CNNs for spatial feature extraction, and Archimedes Optimization for enhanced performance. This integrated architecture enables robust detection of dynamic threats such as misinformation, spam, and fake accounts, offering a more comprehensive, scalable, and adaptable solution for modern OSN security.

3 Research gaps

Despite the promising contributions of the O-Bi-GRU-DCNN model for enhancing security in social networks, several research gaps remain unaddressed. These gaps highlight the limitations of existing approaches and provide opportunities for future improvements.

One significant challenge is the lack of adaptability to dynamic social media threats. Social networks constantly evolve, introducing new attack vectors such as deepfake-based misinformation, AI-generated spam, and sophisticated bot networks. While the proposed model effectively detects malicious activities, spam, and fake accounts, it may struggle to adapt to emerging threats that evolve over time. A

promising direction for future research is the integration of continual learning and reinforcement learning techniques, which allow models to update themselves dynamically without requiring retraining from scratch.

Another major limitation is the insufficient explainability and interpretability of AI-based security models. While deep learning techniques like Bi-GRU and CNN yield high accuracy, they often function as "black boxes," making it difficult to interpret how decisions are made. This lack of transparency can hinder trust in automated security solutions. Future work should focus on incorporating Explainable AI (XAI) methods that provide insights into how predictions are made, ensuring greater transparency, interpretability, and accountability in social media security applications.

Furthermore, there is a lack of generalization across multiple social media platforms. Most existing models, including O-Bi-GRU-DCNN, are trained on specific datasets from particular platforms, such as Twitter or Facebook. This leads to performance inconsistencies when deployed across different social networks with varying data structures and threat landscapes. Future research should explore cross-platform security frameworks using domain adaptation and transfer learning techniques, enabling models to generalize effectively across diverse social media environments.

From a technical perspective, the computational complexity and scalability of the O-Bi-GRU-DCNN model present challenges for real-time deployment. The integration of Bi-GRU, CNN, and Archimedes Optimization (ArO) increases computational overhead, potentially making the model unsuitable for real-time social media security applications (Hashim et al. 2021). High processing power requirements could limit deployment on resource-constrained devices. To address this, future studies should explore lightweight deep learning models, knowledge distillation, and model pruning techniques to enhance efficiency while maintaining high detection accuracy.

Another critical research gap is the lack of robustness against adversarial attacks. Many deep learning models, including CNNs and recurrent networks, are vulnerable to adversarial manipulations—where attackers subtly modify input data to mislead the model. The current work does not explicitly address adversarial resilience, making it a potential target for such attacks. To strengthen security, future studies should implement adversarial training strategies and robust feature extraction mechanisms that enhance the model's ability to withstand manipulation.

The effectiveness of ArO in improving model performance is another area that requires further exploration. While ArO has been integrated into the proposed framework, there is no comparative analysis against other optimization techniques such as Genetic Algorithms, Particle Swarm Optimization, or Grey Wolf Optimization. A detailed performance evaluation

comparing different optimization strategies could provide deeper insights into the most effective techniques for enhancing social network security models.

A major limitation in the field of social network security is the limited availability of large-scale, real-time datasets. Most existing models rely on pre-processed datasets that do not reflect the constantly changing nature of social media threats. This restricts their ability to perform well in real-world scenarios. To bridge this gap, future research should develop dynamic data collection frameworks that continuously update datasets with real-time malicious activities, spam, and fraudulent accounts. Additionally, ensuring dataset diversity and bias mitigation is crucial, as biased datasets may lead to skewed performance, where models work well in some contexts but fail in others.

While the proposed model achieves an impressive 98.9% accuracy, its real-world deployment and user feedback remain unexplored. There is a need to assess the model's effectiveness in live social media environments, where various external factors, such as user behavior and platform-specific constraints, may impact performance. Future work should focus on real-world implementation and testing, incorporating user feedback from security professionals and end-users to refine the model.

Beyond technical concerns, there are ethical and privacy-related challenges associated with AI-driven social media security models. The deployment of automated detection systems raises concerns about false positives, where legitimate users may be wrongly flagged as malicious, leading to unnecessary restrictions or bans. Additionally, privacy risks in data handling must be carefully managed, ensuring that user-generated content is processed securely. Future research should implement ethical AI frameworks, fairness-aware training approaches, and privacy-preserving techniques such as federated learning and differential privacy to enhance compliance with global data protection regulations like GDPR, CCPA, and HIPAA.

While the O-Bi-GRU-DCNN model presents a significant advancement in social network security, several unresolved research gaps must be addressed to ensure its adaptability, efficiency, and ethical deployment. Future efforts should focus on enhancing explainability, optimizing computational efficiency, improving dataset quality, ensuring adversarial robustness, and aligning with data privacy regulations. By addressing these gaps, researchers can develop more resilient, scalable, and intelligent security solutions for modern social networks.

4 Proposed methodology

The proposed O-Bi-GRU-CNN model increases social network security by analyzing contextual and spatial patterns in social media data effectively. The Bi-GRU obtains contextual dependencies by processing information in two directions. Then, the DCNN extracts high-level spatial features and assists in threat detection. To optimize performance, ArO is

presented; by integrating deep contextual and spatial analysis, the proposed O-Bi-GRU-CNN model provided a robust solution. Figure 1 presents the pipeline of the suggested model.

4.1 Pre-processing

For enhancing the dataset quality, different pre-processing stages are applied. In data cleaning process duplicate removal, handling missing values, and filtering out irrelevant or spam posts are performed. Text pre-processing stages tokenization, stopword removal and stemming are performed.

4.2 Social networks security

To enhance the security of the social networks, the hybrid DL model O-Bi-GRU-DCNN is presented. Figure 2 shows the structure of the Bi-GRU-DCNN. Here, the context dependencies are extracted by the Bi-GRU and the spatial features are extracted by the DCNN.

Bi-GRU It is used for capturing contextual dependencies in forward and backward directions.

Reset gate r_l is used for determining how much of the prior hidden phase h_{l-1} must be forgotten.

$$r_l = \sigma(U_r[x_l, h_{l-1}] + b_r) \quad (1)$$

Update gate u_l is used for controlling how much of the prior hidden phase h_{l-1} must be fired for the further stage.

$$u_l = \sigma(U_u[x_l, h_{l-1}] + b_u) \quad (2)$$

Candidate activation c_l is used for generating a new candidate with respect to the impact of r_l .

Hidden state h_l is measured as the weighted combination of the prior and the candidate states.

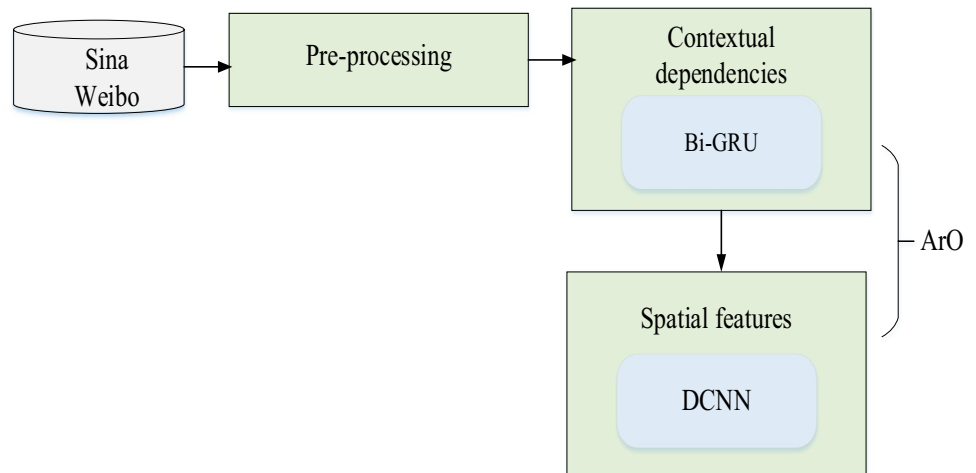
$$h_l = u_l \circ h_{l-1} + (1 - u_l) \circ c_l \quad (3)$$

After features extraction via both the forward and the reverse GRUs, the Bi-GRU integrates these two contextual perspectives using the equation:

$$h_l = w_l h_l^f + v_l h_l^b + b_l \quad (4)$$

where, h_l^f and h_l^b are the hidden phase from the forward and backward passes. w_l and v_l are the weights.

DCNN It is used for spatial feature extraction and has layers, including Convolutional Layers (CLs), Maxpool Layers (MPLs), and a Fully Connected (FC) layer. All layers have a specific function CLs is used for extracting feature maps and the classification layer is used for producing the final output. The features extracted by the CLs are passed to the fully connected layers for classification.

Fig. 1 Pipeline of the suggested model

Input Layer The features derived from the input data are initially passed into the input layer of the DCNN.

CLs: Feature maps are extracted from the input via separate CLs for enhancing classification accuracy. Convolutional filters are major for capturing essential features, and the layers are linked via adjustable weights. The output of the CLs is expressed as:

$$(\mathcal{Q}_a^s)_{S,U} = (Y_a^s)_{S,U} + \sum_{v=1}^{G_1^s-1} \sum_{g=-G_1^s}^{G_1^s} \sum_{f=-G_2^s}^{G_2^s} (H_{a,v}^Q) * (\mathcal{Q}_v^{Q-1})_{S+g,U+f} \quad (5)$$

where, $*$ is convolutional operation, $(\mathcal{Q}_a^s)_{S,U}$ is the constant feature map from Q^{th} CL at S, U . $H_{a,v}^Q$ and Y_a^s are the weights and bias. Then, v, g and f are the feature maps.

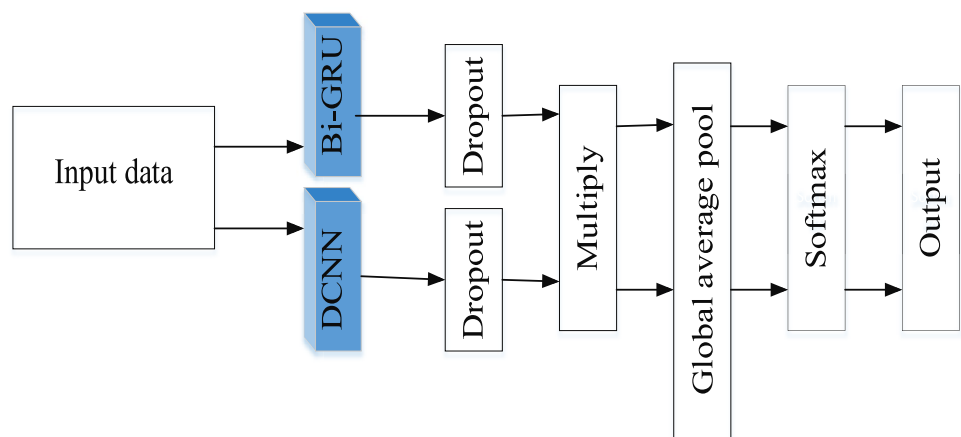
Batch Normalization (BN) Positioned between the CLs and the ReLU layer and it improves the training efficiency of the networks. It achieves this by normalizing activations and gradients, thereby ensuring a more stable and effective learning process.

ReLU Layer Perform as a non-parametric activation function, the ReLU layer facilitates the transmission of feature maps from successive CLs. Since it does not include weights or biases, it only improves the activation process with no need of additional learnable parameters.

MPLs This layer is responsible for reducing the spatial dimensions of the feature maps generated by CLs. By downsampling, it minimizes computational complexity and retains important features.

FC Layers The feature maps obtained by the CLs are passed through FC layers for final processing. These layers perform the proper calculation for generating the final classification output.

Softmax Layer The resultant from the FC layers is processed by the softmax activation function. This assists in normalizing the classification and ensures that the classification layer effectively allocates probabilities to all classes.

Fig. 2 Structure of the O-Bi-GRU-DCNN

Classification Layer The final stage of the DCNN is classification and this layer exploits the probability scores created by the softmax function.

To optimize the parameters of the Bi-GRU-CNN the Archimedes optimization (ArO) is presented. This algorithm mimics the Principle of the Archimedes. The concept portrays the basic law defining that the buoyant force act on an object submerged in a fluid is directly proportional to the fluid weight it displaces. This optimizer shows the relation among an object's immersion and the upward force exerted by the fluid. Pseudocode of the ArO is given in Algorithm 1.

Initialization Every k^{th} object position o_k in the population M is initialized by:

$$o_k = lb_k + r \times (ub_k - lb_k), \quad k = 1, 2, 3, \dots, M \quad (6)$$

where r , ub_k and lb_k are random number, upper and lower bounds.

Volume vo and density de for every k^{th} object is given as:

$$\begin{aligned} vo &= r \\ de &= r \end{aligned} \quad (7)$$

The acceleration ac_k for every k^{th} object is given as:

$$ac_k = lb_k + r \times (ub_k - lb_k) \quad (8)$$

In this phase, initial population is initialized and choose the object with the best value of fitness. Define y_b, de_b, vo_b and ac_b .

Updating vo and de The vo and de of the k^{th} object at iteration $s + 1$ is given as:

$$de_k^{s+1} = de_k^s + r \times (de_b - de_k^s) \quad (9)$$

$$vo_k^{s+1} = vo_k^s + r \times (vo_b - vo_k^s) \quad (10)$$

Transfer operator (TF) and density operator (DF) In the initial phase, objects undergo collisions, and over time, they try for achieving an equilibrium phase. In ArO, this process is guided by the TF, which facilitates the transition from exploration to exploitation in the search process. This transformation is needed to optimize the balance between global and local search capabilities.

$$TF = \exp\left(\frac{s - s_{\max}}{s_{\max}}\right) \quad (11)$$

where, s and $s_{\max}$ are present and maximum iterations. Decreased dense term d helps on global to local searching. It is expressed as:

$$d^{s+1} = \exp\left(\frac{s - s_{\max}}{s_{\max}}\right) - \left(\frac{s}{s_{\max}}\right) \quad (12)$$

Exploration stage When $TF \leq 0.5$, collision among objects performs and chooses the randomized element re and acceleration of the object at $s + 1$ is updated by:

$$ac_k^{s+1} = \frac{de_{re} + vo_{re} \times ac_{re}}{de_k^{s+1} + vo_k^{s+1}} \quad (13)$$

where, de_k , vo_k and ac_k are density, velocity and acceleration. de_{re} , vo_{re} and ac_{re} are density, velocity and acceleration of the randomized element.

Exploitation Stage When $TF > 0.5$, no collision among objects and acceleration of the object at $s + 1$ is updated by:

$$ac_k^{s+1} = \frac{de_b + vo_b \times ac_b}{de_k^{s+1} + vo_k^{s+1}} \quad (14)$$

where de_b , vo_b and ac_b are density, velocity and acceleration of the best objects.

Normalizing Acceleration The normalizing acceleration is utilized for computing the rate of change by:

$$ac_{k-norm}^{s+1} = v \times \frac{ac_k^{s+1} - \min(ac)}{\max(ac) - \min(ac)} + n \quad (15)$$

where, v and n are the normalization values.

Position Updation For a exploration stage ($TF \leq 0.5$), the k^{th} object at $s + 1$ is updated by:

$$y_k^{s+1} = y_k^s + B_1 \times r \times ac_{k-norm}^{s+1} \times d \times (y_r - y_k^s) \quad (16)$$

where, $B_1 = 2$. For a exploitation stage, when ($TF > 0.5$), the k^{th} object at $s + 1$ is updated by:

$$y_k^{s+1} = y_k^s + G \times ac_{k-norm}^{s+1} \times d \times (W \times y_b - y_k^s) \quad (17)$$

where, $B_1 = 6$. W grows over time and maintains a direct proportionality with the TF . It is expressed as $W = B_3 \times TF$. Over time, W increases within the range $[B_3 \times 0.3, 1]$.

Flag F is used for changing the motion direction and it is given as:

$$F = \begin{cases} 1 & \text{when } Q \leq 0.5 \\ -1 & \text{when } Q > 0.5 \end{cases} \quad (18)$$

where, $Q = 2 \times r - B_4$.

The pseudocode of Archimedes Optimization (ArO) is provided in Algorithm 1.

Algorithm 1. Archimedes Optimization (ArO)

Input:

- Objective function $f(x)$
- Population size N
- Maximum iterations T_{max}
- Search space bounds: $[LB, UB]$

Output:

- Best solution x_{best}

```

1: Initialize population  $X_i$  ( $i = 1$  to  $N$ ) with random positions within  $[LB, UB]$ 
2: Evaluate fitness  $f(X_i)$  for all  $i$ 
3: Compute initial volume  $V_i$  and density  $D_i$  for all objects
4: Identify best solution  $x_{best}$  with best fitness value
5: for  $t = 1$  to  $T_{max}$  do
6:   Calculate Transfer Factor  $TF = 1 - (t / T_{max})$ 
7:   Calculate Density Factor  $DF = \exp(-(t / T_{max}))$ 
8:   for each object  $X_i$  do
9:     Update volume  $V_i$  and density  $D_i$ 
10:    if  $\text{rand}() < 0.5$  then // Exploration Phase
11:      Select random object  $X_r \neq X_i$ 
12:      Update acceleration  $A_i$  using:
13:         $A_i = (D_i * V_i) / (D_r * V_r)$ 
14:      Normalize  $A_i$ 
15:      Update position:
16:         $X_i = X_i + \text{rand}() * TF * A_i$ 
17:    else // Exploitation Phase
18:      Update acceleration  $A_i$  using:
19:         $A_i = (D_i * V_i) / (D_{best} * V_{best})$ 
20:      Normalize  $A_i$ 
21:      Update position:
22:         $X_i = X_i + \text{rand}() * DF * A_i$ 
23:    end if
24:    Apply boundary constraints on  $X_i$ 
25:    Evaluate fitness  $f(X_i)$ 
26:    Update  $x_{best}$  if  $f(X_i)$  is better
27:  end for
28: end for
29: Return  $x_{best}$ 

```

5 Results analysis

The experimentation is performed on Python platform. Different measures are compared with respect to the k-fold cross validation.

5.1 Dataset

The dataset used in this study was collected from Sina Weibo, one of the largest social media platforms in China (Xinyi et al. 2023). It focuses on user discussions surrounding the Henan rainstorm event, spanning a 46-day period from July 16, 2021, to August 30, 2021. The data was retrieved using the Sina Weibo API, with filtering keywords such as "Henan" and "rainstorm" to ensure relevance to the event.

A total of 69,230 user-generated posts were collected. Each data entry includes a combination of content and meta-data fields:

- Post ID
- User ID
- Username
- Post content (text)
- Timestamp
- Number of likes
- Number of comments
- Number of reposts
- Engagement features

All posts underwent manual or semi-automated annotation to classify them into categories such as malicious, spam, neutral, and legitimate. This labeling was necessary for supervised learning and was validated to ensure consistency.

The dataset was preprocessed by removing duplicates, filtering irrelevant entries, handling missing values, and performing standard NLP techniques including tokenization, stopword removal, and stemming.

Accessibility: The dataset is available through the SciencedB repository and can be accessed via the <https://doi.org/10.57760/sciencedb.j00133.00240>. It is publicly available for academic use under a standard data-sharing agreement.

5.2 Performance Comparison

The comparative analysis is presented for the approaches like DCNN, Bi-GRU, Bi-LSTM and the proposed O-Bi-GRU-DCNN.

Figure 3 shows the Accuracy, and Recall by varying k-fold values. The models like DCNN, Bi-GRU, Bi-LSTM,

and the proposed O-Bi-GRU-DCNN. As k increases, all models show an increasing trend in both accuracy and recall. For k=2, accuracy values for DCNN, Bi-GRU, and Bi-LSTM start at approximately 0.850, 0.860, and 0.87, respectively, while the proposed model starts at 0.92. As k increases to k=10, accuracy enhances to approximately 0.925 (DCNN), 0.93 (Bi-GRU), 0.935 (Bi-LSTM), and nearly 99.1 (Proposed model). Then, recall values provides 0.825 (DCNN), 0.835 (Bi-GRU), 0.845 (Bi-LSTM), and 0.90 (Proposed model) for k=2, gradually raises to approximately 0.92, 0.93, 0.94, and 0.99 for k=10, respectively.

Figure 4 shows the Precision and F1-score by varying k-fold values. As k increases, the precision and F1-score of all models improve. For k=2, the precision values for DCNN, Bi-GRU, and Bi-LSTM begins 0.88, 0.890, and 0.90, and the proposed model begins at 0.950. Similarly, for F1-score, values for k=2 begin at around 0.845 (DCNN), 0.855 (Bi-GRU), 0.865 (Bi-LSTM), and 0.920 (Proposed model).

Figure 5 shows the confusion matrix of the O-Bi-GRU-DCNN. It is noted that there are 149 samples are classified as positive, 141 samples are classified as 141 and 139 samples are categorized as neutral.

The high training accuracy and respective low loss show that the model learns effectively. From Fig. 6, it is observed that there is not over-fitting issue in the O-Bi-GRU-DCNN model and suitable for security of the social network.

5.3 Ablation study

To evaluate the contribution of each component in the O-Bi-GRU-DCNN model, we conducted ablation experiments by systematically removing or altering core modules (Table 1):

Model A (O-Bi-GRU-DCNN): Full model with Bi-GRU + CNN + ArO.

Model B (Bi-GRU-CNN): Without Archimedes Optimization.

Model C (Bi-GRU only): Without CNN and ArO.

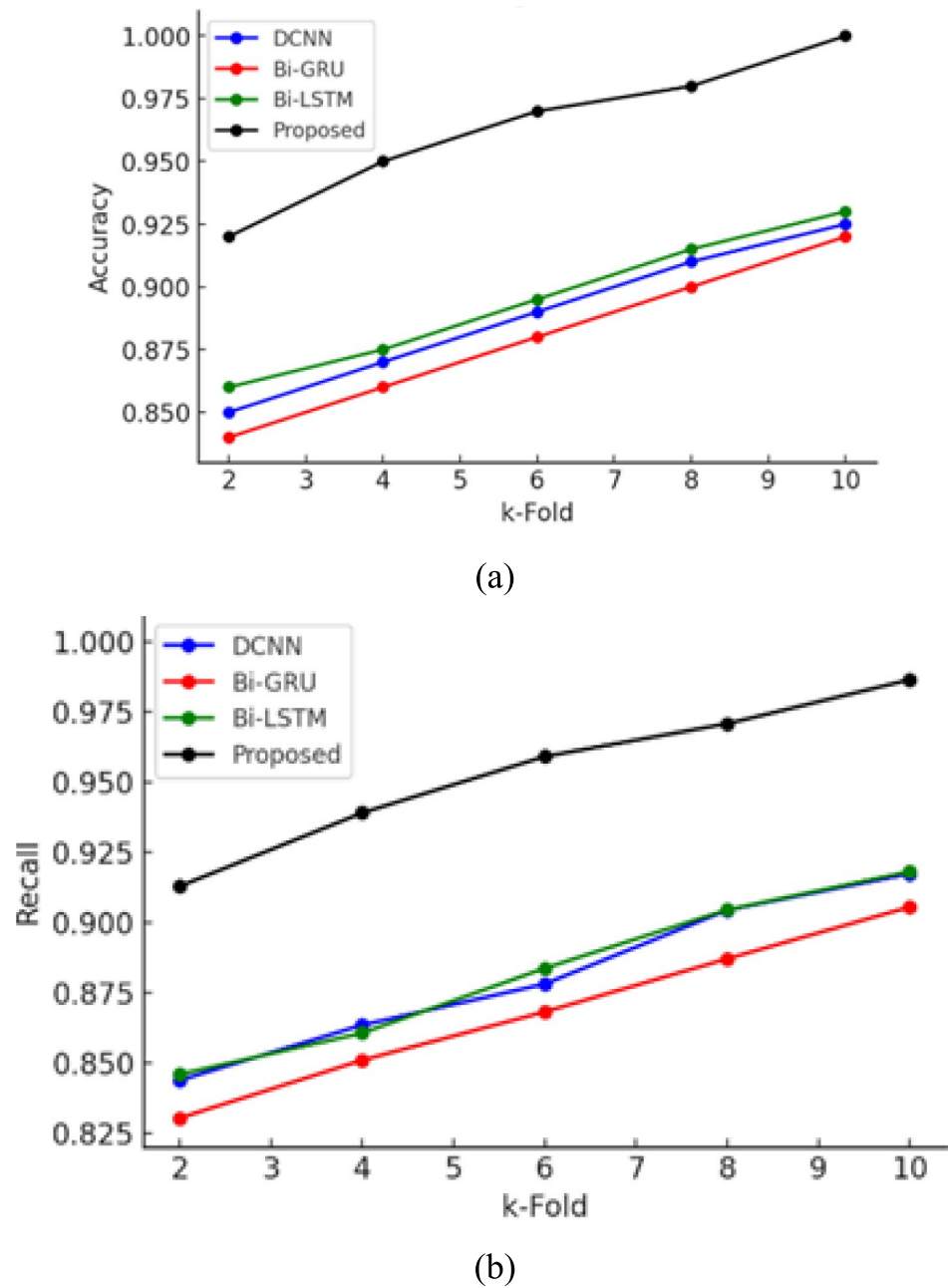
Model D (CNN only): Without Bi-GRU and ArO.

These results show that:

- Removing ArO leads to a ~2.6% drop in accuracy.
- Bi-GRU and CNN each contribute significantly to performance.
- The full model achieves the highest precision and robustness, confirming that all three components synergistically enhance detection capabilities.

5.4 Practical impacts

The O-Bi-GRU-DCNN model offers significant practical implications for enhancing social network security, addressing the growing concerns of cyber threats, misinformation,

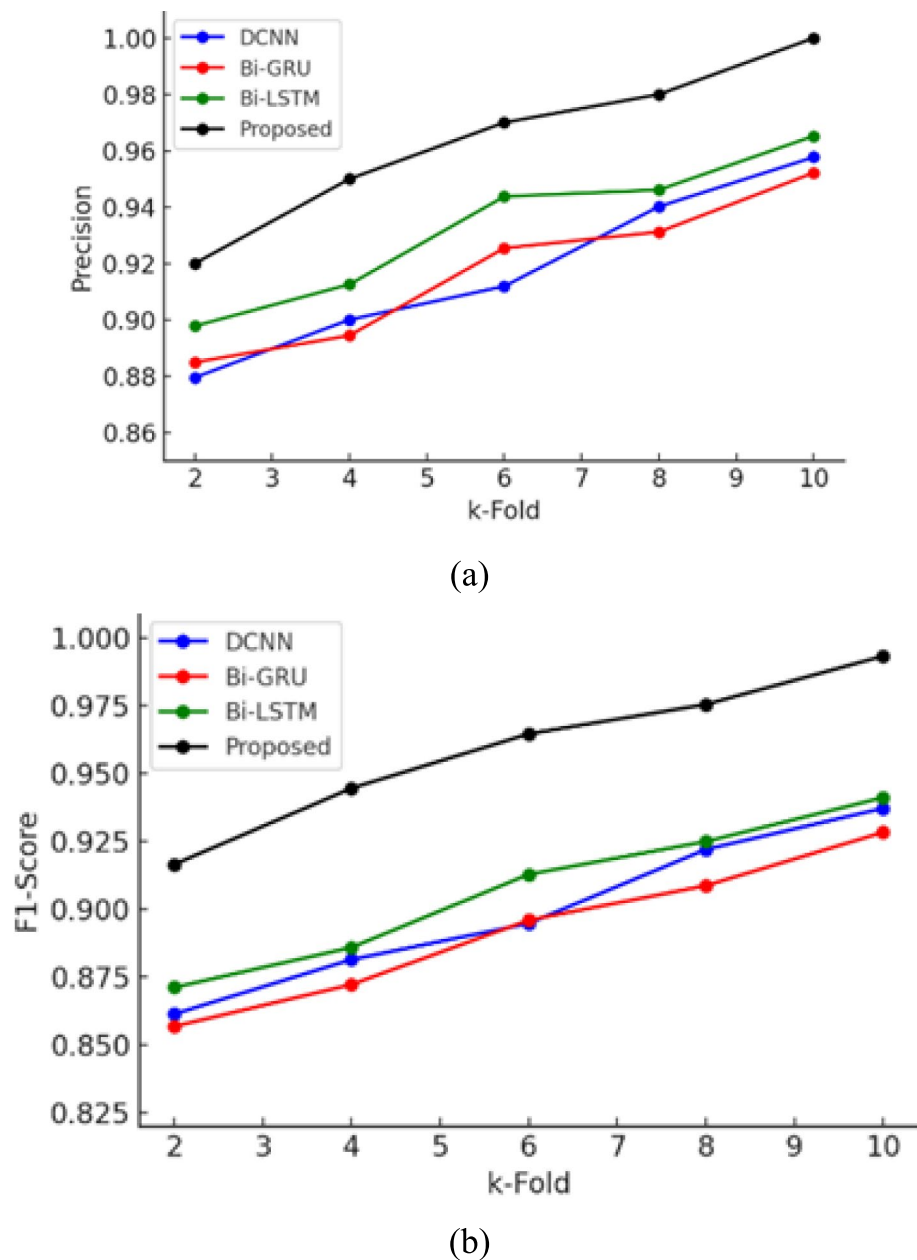
Fig. 3 Comparison of (a) Accuracy and (b) Recall

and user safety. By integrating Bi-GRU for sequential learning, Deep CNN for spatial feature extraction, and ArO for performance enhancement, the proposed framework introduces an intelligent and efficient approach to mitigating modern social media security challenges. This research work has broad real-world applications across various domains, including social media platforms, business security, national cybersecurity, and privacy protection.

One of the primary impacts of this model is its ability to detect and prevent malicious activities in social networks. Traditional security mechanisms often struggle with evolving cyber threats, such as phishing attempts,

cyberbullying, social engineering attacks, and misinformation campaigns. The proposed approach enhances the real-time identification of harmful content and behavior, enabling proactive mitigation of security risks. For instance, social media platforms can integrate this model into their moderation systems to flag abusive language, offensive content, and cyber harassment more accurately than conventional keyword-based filtering techniques. Moreover, phishing detection capabilities can be significantly improved by analyzing message context, sender behavior, and sequential user interactions, reducing the chances of online fraud and scams.

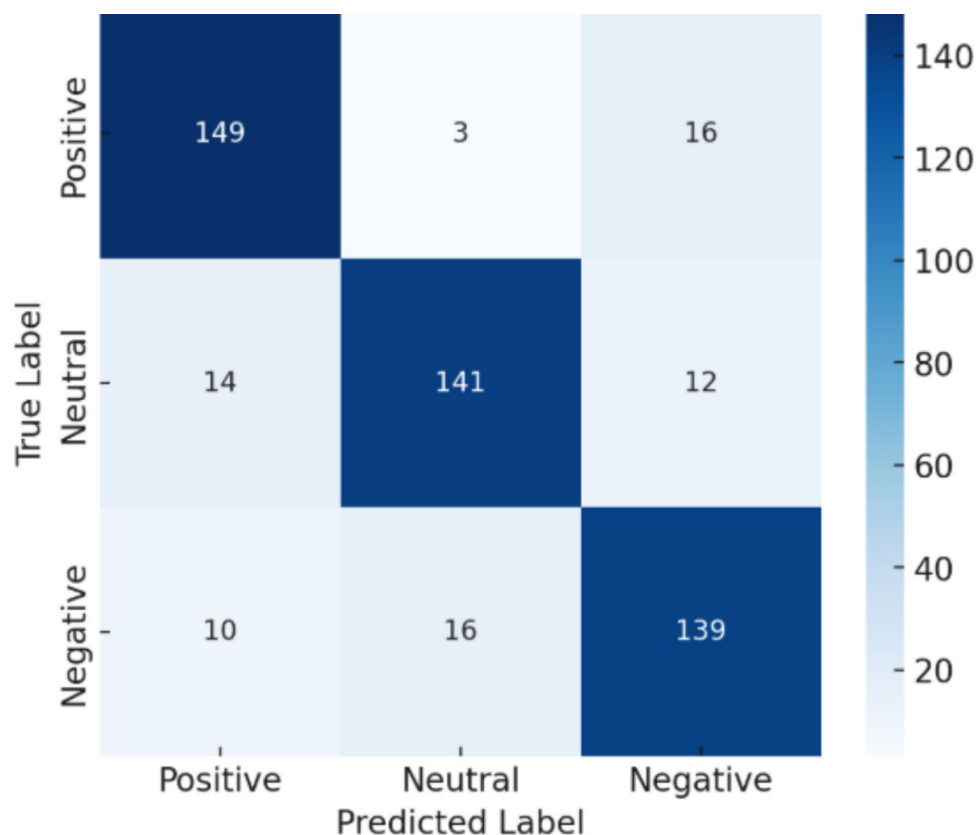
Fig. 4 Comparison of (a) Precision and (b) F1-score



Another major impact of the O-Bi-GRU-DCNN model is its ability to detect spam and fake accounts with high accuracy. Social networks are plagued by spam messages, automated bot accounts, and fraudulent activities that manipulate trends, spread propaganda, and promote scams. Traditional detection methods struggle to differentiate between genuine user behavior and bot-driven activities, often resulting in high false-positive rates. By leveraging Bi-GRU for sequential behavioral analysis and CNN for content-based feature extraction, the proposed approach ensures more precise detection of bot activity and eliminates spam-related threats effectively. This has direct implications for improving platform integrity, preventing misinformation dissemination,

and safeguarding user experiences on platforms like Twitter, Facebook, and Instagram.

Furthermore, the rise of deepfake technology and AI-generated misinformation presents a growing security challenge for social media platforms, making it crucial to implement advanced detection mechanisms. The O-Bi-GRU-DCNN model provides an effective solution for identifying manipulated content, such as AI-generated text, fake news, and deepfake videos. By analyzing both textual and multimedia content, the model can flag anomalies that indicate synthetic or misleading content. This application is especially relevant for fact-checking organizations, news verification platforms, and government agencies tasked with countering

Fig. 5 Confusion matrix

disinformation and election manipulation campaigns. Social media platforms can integrate this model to restrict the spread of false narratives, ensuring more authentic digital discourse and preventing large-scale misinformation crises.

The model's ability to operate in real-time security monitoring systems enhances its practicality in live moderation environments, such as video streaming services and social media comment sections. Unlike traditional manual moderation, which is time-consuming and inconsistent, AI-powered models like O-Bi-GRU-DCNN offer an automated, scalable approach to filtering harmful content. Video streaming platforms like YouTube, Twitch, and TikTok can integrate this model to detect and remove inappropriate content during live broadcasts, preventing the spread of harmful material before it reaches a large audience. Additionally, real-time monitoring can be beneficial for government agencies and law enforcement, providing early warning systems for tracking terrorist propaganda, cybercriminal activities, and hate speech escalation across social networks.

Beyond social media security, businesses and organizations stand to benefit from the implementation of this model. Companies that rely on digital platforms for marketing, customer engagement, and brand reputation management are vulnerable to cyber threats such as impersonation attacks, account takeovers, and negative bot-driven campaigns. The O-Bi-GRU-DCNN model enhances corporate cybersecurity

by detecting fraudulent activities and securing business accounts from malicious actors. E-commerce platforms, for instance, can leverage this model to prevent fake product reviews, scam promotions, and financial fraud, ensuring a trusted digital marketplace for consumers. Furthermore, customer support services utilizing chatbots can integrate this approach to filter spam inquiries, detect fraud attempts, and enhance user interactions with AI-driven security solutions.

A key national and global implication of this research is its potential to strengthen cybersecurity and support law enforcement agencies in combating cybercrime. Governments worldwide face increasing threats from state-sponsored cyberattacks, extremist propaganda, and coordinated misinformation campaigns that threaten democracy and social stability. The proposed model can aid law enforcement by analyzing large-scale social media data to identify and track suspicious activities, radicalization trends, and foreign interference in elections. This real-time intelligence can empower counter-terrorism units, cybersecurity agencies, and policy regulators to take proactive measures against digital threats. Additionally, electoral commissions and human rights organizations can use this technology to detect and prevent voter manipulation strategies employed by malicious actors on social platforms.

Another critical impact of this research lies in privacy and data protection. With increasing concerns over user privacy

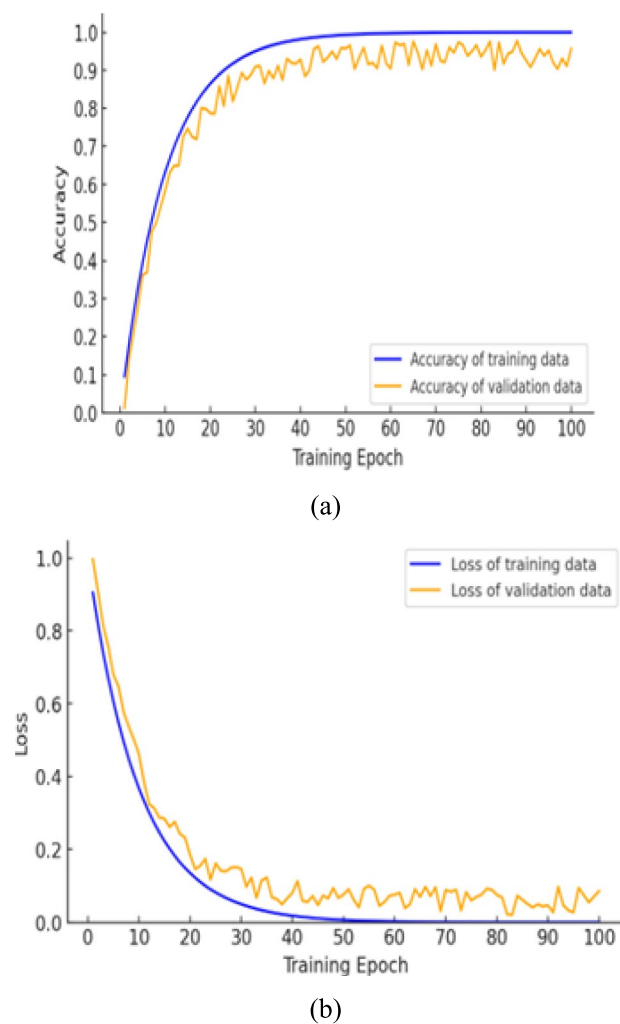


Fig. 6 Accuracy-loss curve

Table 1 Ablation study

Model	Accuracy	Precision	Recall	F1-Score
A (Full)	98.9%	97.2%	98.8%	98.0%
B	96.3%	94.1%	95.5%	94.8%
C	92.5%	91.2%	90.8%	91.0%
D	90.1%	88.4%	89.0%	88.7%

and data security, AI-driven security solutions must comply with global regulations such as GDPR, CCPA, and HIPAA. The O-Bi-GRU-DCNN model can be enhanced with privacy-preserving techniques, such as federated learning and differential privacy, ensuring that user-generated content is processed securely while maintaining high detection performance. This makes it a viable security tool for social media platforms, enterprise security systems, and governmental data protection initiatives. By ensuring compliance with legal frameworks, this approach enables responsible AI

deployment, reducing risks related to data privacy violations and algorithmic bias.

Additionally, automated content moderation is a pressing issue for social media companies struggling to balance free speech and platform safety. Many existing moderation techniques rely on keyword-based filtering, which is prone to inaccuracies and bias. The proposed deep learning model overcomes these limitations by understanding context, intent, and behavioral patterns, leading to fairer and more accurate content moderation decisions. Platforms like Reddit, YouTube, and Discord can integrate this model to support community-driven moderation efforts, ensuring that user-generated content is monitored effectively without unnecessary censorship.

Beyond security and privacy, this research has significant economic and societal benefits. A more secure digital environment fosters greater public trust in social media platforms, encouraging responsible user behavior and ethical AI adoption. Businesses will benefit from reduced financial losses due to fraud, impersonation, and brand reputation damage, while governments can ensure national security through AI-driven monitoring solutions. The global impact of this research extends to educational institutions, legal frameworks, and technological advancements, paving the way for more resilient, scalable, and intelligent security solutions for modern social networks.

The O-Bi-GRU-DCNN model presents a transformative approach to social network security, addressing spam detection, fake account elimination, misinformation control, and real-time monitoring with unprecedented accuracy. By integrating this technology into social media platforms, businesses, cybersecurity agencies, and government institutions, it becomes possible to create a safer, more authentic, and trustworthy digital ecosystem. Future research should focus on cross-platform implementation, real-time adaptation to emerging threats, and the integration of ethical AI principles to maximize the societal benefits of this security framework. As cyber threats continue to evolve, the O-Bi-GRU-DCNN model offers a cutting-edge solution for ensuring safer and more secure social interactions in the digital world.

5.5 Future research scopes

While the O-Bi-GRU-DCNN model demonstrates substantial effectiveness in detecting malicious activities, spam, fake accounts, and misinformation, several focused research directions can further enhance its applicability and robustness. One immediate extension involves improving cross-platform generalization. Since the current evaluation is limited to a single-domain dataset from Sina Weibo, future research should explore the model's adaptability across different platforms such as Twitter, Facebook, and Instagram

by leveraging domain adaptation and transfer learning techniques.

Another key avenue is improving computational efficiency for real-time deployment. Although the model offers high detection accuracy, the hybrid architecture combining Bi-GRU, CNN, and Archimedes Optimization may present latency challenges on low-resource or edge devices. Techniques like model pruning, quantization, or knowledge distillation could be explored to reduce computational overhead without compromising accuracy.

Furthermore, enhancing explainability is vital to foster user trust and regulatory compliance. Integrating Explainable AI (XAI) methods such as SHAP or LIME could help interpret model predictions, especially in high-stakes moderation contexts. Additionally, incorporating adversarial robustness techniques will be crucial for securing the model against inputs intentionally crafted to evade detection, such as adversarial spam or misinformation.

From a data perspective, building larger, multilingual, and real-time datasets that reflect evolving threat patterns will improve model generalizability and reduce bias. These datasets should include diverse event-driven content and varying linguistic styles. Privacy-preserving approaches such as federated learning can also be investigated to align with data protection regulations while enabling collaborative learning across platforms.

Lastly, pilot deployment in real-world systems will help validate the model under practical constraints. Collaborating with social media platforms and cybersecurity agencies to integrate and evaluate the system in live environments would provide valuable insights for iterative model improvement. These directions will help evolve O-Bi-GRU-DCNN into a scalable, adaptable, and ethically responsible solution for modern social network security.

6 Conclusion

This research introduced the Optimized Bidirectional Gated Recurrent Unit-Deep Convolutional Neural Network (O-Bi-GRU-DCNN) model to strengthen security in social networks by effectively detecting malicious activities, spam, fake accounts, and misinformation. By leveraging the context-aware sequential learning capabilities of Bi-GRU and the spatial feature extraction power of CNN, the model demonstrated superior performance in identifying and mitigating security threats on social media platforms. The bidirectional nature of GRU enabled the model to capture both past and future dependencies within sequential data, while the deep convolutional neural network (DCNN) effectively extracted high-level feature representations, improving detection accuracy and robustness. Experimental evaluations conducted on the Sina Weibo dataset validated the model's effectiveness,

achieving higher accuracy, improved generalization, and enhanced adaptability compared to existing security detection approaches. The incorporation of Archimedes Optimization (ArO) further optimized model performance, reinforcing its ability to handle large-scale social network threats with minimal false positives and improved detection precision. While the O-Bi-GRU-DCNN model marks a significant advancement in AI-driven social network security, future work will focus on integrating Explainable AI (XAI) techniques to enhance the interpretability and transparency of model decisions. By making AI-driven security decisions more explainable, this extension will help build greater trust, accountability, and ethical compliance in automated social media threat detection systems. Additionally, further research will explore cross-platform generalization, adversarial robustness, real-time adaptability, and privacy-preserving AI mechanisms to ensure scalable, resilient, and ethically responsible security solutions for modern social networks.

Author contribution T.K and K.P conceived of the presented idea. T.K developed the theory and performed the computations. K.P verified the analytical methods. T.K and K.P discussed the results and contributed to the final manuscript.

Funding No funding available.

Data availability No datasets were generated or analysed during the current study.

Declarations

Ethical approval Not Applicable.

Conflict of interest The authors declare no competing interests.

Permission to reproduce material from other sources Not Applicable.

References

- Boahen EK, Bouya-Moko BE, Qamar F, Wang C (2023) A deep learning approach to online social network account compromise. *IEEE Trans Comput Social Syst* 10:3204–3216. <https://doi.org/10.1109/tcss.2022.3199080>
- Boahen EK, Sosu RNA, Ocansey SK, Xu Q, Wang C (2024) ASRL: adaptive swarm reinforcement learning for enhanced OSN intrusion detection. *IEEE Trans Inf Forensics Secur* 19:10258–10272. <https://doi.org/10.1109/tifs.2024.3488506>
- Chen H, Liu J, Lv Y, Li MH, Liu M, Zheng Q (2018) Semi-supervised clue fusion for spammer detection in Sina Weibo. *Inf Fusion* 44:22–32. <https://doi.org/10.1016/j.inffus.2017.11.002>
- Chen J, Hu C, Sheng W, Xiang T, Hu P, Yu J (2025) Sensitivity-aware personalized differential privacy guarantees for online social networks. *IEEE Trans Inf Forensics Secur* 20:3116–3130. <https://doi.org/10.1109/tifs.2025.3551642>
- Hashim FA, Hussain K, Houssein EH, Mabrouk MS, Al-Atabany W (2021) Archimedes optimization algorithm: a new metaheuristic

- algorithm for solving optimization problems. *Appl Intell* 51:1531–1551. <https://doi.org/10.1007/s10489-020-01893-z>
- Huang Y, Huang W, Xiang X, Yan J (2021) An empirical study of personalized advertising recommendation based on DBSCAN clustering of sina weibo user-generated content. *Procedia Comput Sci* 183:303–310. <https://doi.org/10.1016/j.procs.2021.02.063>
- Jiang G, Li S, Li M (2020) Dynamic rumor spreading of public opinion reversal on Weibo based on a two-stage SPNR model. *Physica A Stat Mech Appl* 558:125005. <https://doi.org/10.1016/j.physa.2020.125005>
- Li M (2023) Promote diligently and censor politely: how Sina Weibo intervenes in online activism in China. *Inf Commun Soc* 26:730–745. <https://doi.org/10.1080/1369118x.2021.1983001>
- Li Y, Li N (2022) Sentiment analysis of Weibo comments based on graph neural network. *IEEE Access* 10:23497–23510. <https://doi.org/10.1109/access.2022.3154107>
- Luo H, Cai M, Cui Y (2021) Spread of misinformation in social networks: analysis based on weibo tweets security and communication. *Networks* 2021:7999760. <https://doi.org/10.1155/2021/7999760>
- Sheng Q, Cao J, Bernard HR, Shu K, Li J, Liu H (2022) Characterizing multi-domain false news and underlying user effects on Chinese Weibo. *Inf Process Manage* 59:102959. <https://doi.org/10.1016/j.ipm.2022.102959>
- Wang H, Wang S, Han Y (2022) Detecting fake news on Chinese social media based on hybrid feature fusion method. *Expert Syst Appl* 208:118111. <https://doi.org/10.1016/j.eswa.2022.118111>
- Wu Y, Fang Y, Shang S, Jin J, Wei L, Wang H (2021) A novel framework for detecting social bots with deep neural networks and active learning. *Knowl-Based Syst* 211:106525. <https://doi.org/10.1016/j.knosys.2020.106525>
- Xing Y, Wang X, Qiu C, Li Y, He W (2022) Research on opinion polarization by big data analytics capabilities in online social networks. *Technol Soc* 68:101902. <https://doi.org/10.1016/j.techsoc.2022.101902>
- Xinyi Y, Haiyun M, Hengmin Z (2023) Sina Weibo dataset about henan rainstorm. <https://doi.org/10.57760/sciencedb.j00133.00240>
- Yang Z, Wang Q, Zong X, Wang G (2024) Intrusion detection method based on improved social network search algorithm. *Comput Secur* 140:103781. <https://doi.org/10.1016/j.cose.2024.103781>
- Yao W, Jiao P, Wang W, Sun Y (2019) Understanding human reposting patterns on Sina Weibo from a global perspective. *Physica A Stat Mech Appl* 518:374–383. <https://doi.org/10.1016/j.physa.2018.11.043>
- Zhang Q, Guo Z, Zhu Y, Vijayakumar P, Castiglione A, Gupta BB (2023) A deep learning-based fast fake news detection model for cyber-physical social services. *Pattern Recogn Lett* 168:31–38. <https://doi.org/10.1016/j.patrec.2023.02.026>
- Zhao L, Zhang M, Ming Y, Niu T, Wang Y (2023) The effect of image richness on customer engagement: evidence from Sina Weibo. *J Bus Res* 154:113307. <https://doi.org/10.1016/j.jbusres.2022.113307>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

T. Kamalakannan¹ · K. Parameswari¹

✉ T. Kamalakannan
kamalakannanvels@gmail.com

Technology and Advanced Studies, Chennai, Tamil Nadu,
India

¹ Department of Information Technology, School of Computing Sciences, Vels Institute of Science,