False News Recognition Model Based on Attention Mechanism and Multiple Features

Qiongyao Suo[1]*, Hongzhen Chang[2]
School of Media and Design, Xi'an Peihua University, Xi'an, 710125, China[1]
China Unicom Software Research Institute, Xi'an, 710000, China[2]

***Abstract*—As the prevalence of social media continues to grow, the rapid and wide dissemination of false news has become a critical societal challenge, undermining public trust, creating social unrest, and distorting political discourse. Traditional fake news detection methods often rely solely on linguistic cues or shallow semantic analysis, which leads to limited accuracy and poor robustness, particularly when addressing emotionally biased or contextually complex content. To overcome these limitations, this study proposes a novel fake news recognition model based on a bidirectional gated recurrent unit combined with a self-attention mechanism, further enhanced by integrating sentiment polarity, textual metadata, and contextual semantic features. Experimental results show that the proposed model achieves a recognition accuracy of ninety-seven per cent and an F1 score of ninety-seven. In addition, it demonstrates the lowest mean absolute error, which is zero point one nine, and the shortest recognition time, requiring only zero point eight seconds after eighty iterations. The model also maintains over ninety-three per cent accuracy across news content with active, negative, and neutral emotional tones. The model offers a scalable and reliable framework for detecting false news, with strong adaptability to diverse content types and emotional expressions, thereby contributing to the advancement of automated misinformation identification in real-world applications.**

Keywords*—*Fake news; attention mechanism; multiple features; bidirectional gated recurrent unit

I. Introduction

With the rapid expansion of digital communication platforms, especially social media, information dissemination has become faster and more decentralized than ever. While this shift enables real-time access to global news, it also creates fertile ground for the proliferation of fake news. False information can quickly gain traction online, misleading the public, disrupting social order, and even influencing political and economic landscapes [1-3]. The urgency of addressing fake news is no longer just a journalistic concern—it has evolved into a multidisciplinary research challenge in the fields of artificial intelligence, communication studies, and cybersecurity. Existing methods for detecting fake news primarily rely on either text-based analysis or behavioral pattern mining. While content-based methods evaluate linguistic cues, writing style, and emotional tone, they often fall short when dealing with complex semantic relationships and context-sensitive misinformation. Similarly, user behavior-based models, though useful for identifying propagation anomalies, lack robustness in adversarial environments, where malicious actors may mimic normal user behavior. These limitations underscore the need for a more comprehensive and adaptive detection framework. To address these challenges, this research proposes a fake news detection model that integrates a bidirectional gated recurrent unit (Bi-GRU) with a self-attention mechanism (SAM) and multi-feature fusion, including sentiment polarity, metadata, and contextual cues. The core idea is to capture both global and local semantic dependencies while emphasizing the most informative features through attention weighting. Unlike traditional shallow models, the proposed approach leverages deep contextual understanding to improve generalization across different news types and emotional tones [4-6]. The objective of this research is to enhance detection accuracy and robustness in identifying fake news by bridging the gap between single-modality feature extraction and deep multimodal representation. By explicitly linking emotional features and context-dependent signals, the model is expected to outperform existing methods, particularly in scenarios involving ambiguous or emotionally charged content. This study aims to contribute both a novel algorithmic architecture and empirical evidence supporting its efficacy, providing practical value for real-world applications in social media monitoring and content verification. The research content is further divided into five sections. The first section is a summary of other scholars' current research topics. The second section is a brief description of the algorithm used in this study. The third section presents the model results obtained by using the algorithm and analyzes the results. The fourth section is a discussion of the results obtained by the model. The fifth section is a summary of all the above studies and prospects for future research.

II. Related Works

Deep learning technology has significantly advanced Natural Language Processing. Zhao Y et al. suggested that advancements in interactive processing necessitate machines to identify human emotions. To examine latent emotional states, a weighted kernel strategy was proposed on, which effectively recognizes emotions expressed through human actions [7]. Ahmed M R et al. demonstrated that accurate recognition of emotions in speech signals enhances human-computer interaction. While most studies focus on extracting localized speech features, they often overlook global, long-term contextual representations. To address low recognition performance in current speech emotion systems, they introduced an LSTM-based ensemble method to enhance prediction accuracy, exhibiting promising accuracy across various datasets [8]. Chen L et al. developed a K-means clustering-based model to capture facial expressions and emotions from speech, selecting and reducing the dimensionality of multi-modal features. This model enhanced heterogeneity among modalities, promoting more accurate multi-modal emotion recognition and outperforming non-K-means approaches in recognition rate [9].

Pushpalatha M N et al. believed that visual impairment significantly impacted information understanding. A social assistance application targeting visually impaired individuals was built to address this issue. They developed a transfer learning strategy for facial expression recognition that effectively recognized emotions in facial expressions [10]. Zhang Y et al. highlighted joint emotion recognition in human-computer dialogue as a key topic. A multi-modal, multi-task learning model was proposed based on an encoder-decoder architecture, which demonstrated robust performance across various datasets [11]. Wen et al. believe that multimodal data is crucial for emotion recognition, and textual data can help analyze emotions. However, how to effectively integrate multimodal features to capture complex contextual nuances in verbal communication is a huge challenge. To address this issue, researchers have proposed a model based on dynamic and multi-perspective memory, which integrates information for emotion recognition and demonstrates high accuracy through multimodal data [12].

To sum up, in the domain of fake news recognition, traditional text-based detection models focus primarily on linguistic patterns or stylistic features, yet often lack the ability to capture long-range dependencies and multi-level contextual signals. Similarly, models relying on user behavior analysis—such as propagation patterns or user credibility—face difficulties in adversarial settings, where users may intentionally distort interactions. Moreover, many existing frameworks treat news content as a flat structure, ignoring the dynamic emotional and semantic interplay embedded within. This narrow focus limits their ability to generalize across domains and emotion-laden content. In response, this study introduces a novel fake news detection framework that integrates Bi-GRU with a SAM and multi-feature fusion. Unlike existing models, the proposed architecture emphasizes salient textual patterns while capturing both forward and backward dependencies, thereby offering improved sensitivity to emotional nuances and semantic inconsistencies in news texts. This research contributes to the existing body of knowledge by bridging the gap between unimodal content analysis and deep multimodal learning, providing a scalable and robust solution to fake news detection in diverse digital environments.

III. METHODS

The first section at first encodes news text using a pre-trained BERT model, then extracts text features using an LSTM network, and captures key information in the text through SAM. The second section combines multiple features to identify fake news. The extracted text features are combined with various features, such as news sentiment and input into the classifier for false news recognition.

A. False News Recognition Model Based on Text Extraction

The difference between fake news and real news mainly lies in the authenticity of content, sources of information, purpose, and influence. Real news is based on verified facts, collected by professional journalists and editors through reliable sources, and undergoes strict editing processes to ensure the accuracy and fairness of information. Its purpose is to objectively report events and provide valuable information to the public. On the contrary, fake news often contains unverified or even fabricated content, often misleading readers by exaggerating, distorting facts, or completely fabricating them to advance specific agendas or interests. Fake news often lacks reliable sources and attracts clicks with provocative headlines, with the aim of manipulating public opinion, creating chaos, or gaining economic benefits. Readers need to improve their media literacy and learn to discern the credibility of news sources to avoid being misled by false information. The identification of fake news should comprehensively consider various characteristics such as text content, user behavior, and news sources [13]. The existing false news detection process is shown in Fig. 1.

As shown in Fig. 1, in false news detection, the first step is to extract characteristic information, including news text, images, etc. Then, the extracted data are preprocessed. Machine learning models are applied to represent the features of the news. Finally, a classification model is established [14]. The news text modality extraction is carried out using the BERT model, which is based on the Transformer architecture used for NLP tasks. Its structure is shown in Fig. 2.

In Fig. 2, the BERT language model involves input representation, encoder, SAM, feedforward neural network, and output layer. The input representation consists of Token Embeddings, Segment Embeddings, and Position Embeddings, which are used to map words to vectors, distinguish sentences, and represent word positions, respectively. The encoder is a multi-layer Transformer, with 12 layers for BERT-base and 24 for BERT-large, each layer containing a self-attention head and a feedforward neural network. The SAM captures contextual information by calculating the correlation between query, key, and value matrices, with each self-attention head updating the input representation. After the self-attention layer, the feedforward neural network contains two linear transformations and a ReLU activation function [15]. The output layer is the last layer's output of the encoder stack, used to add task specific output layers for specific tasks.

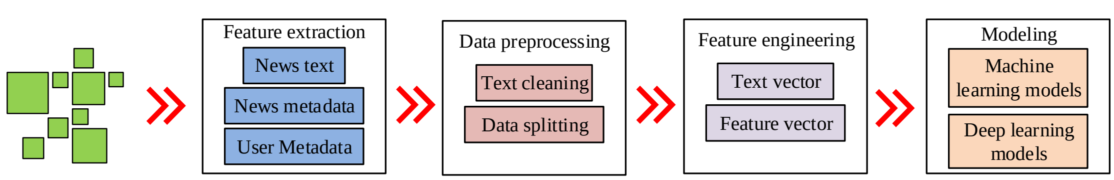


Fig. 1. False news detection process.

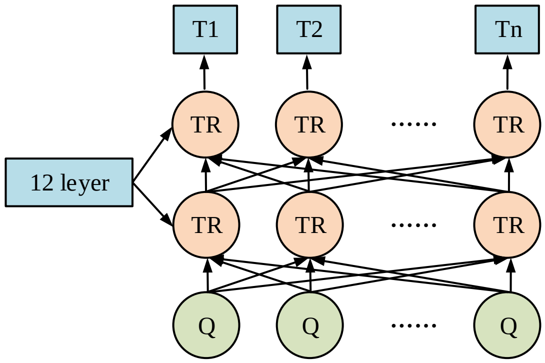


Fig. 2. BERT model.

BERT overcomes the limitation of context learning in directional methods by masking language models and predicting the next sentence. During training, the masking language model randomly masks some words in the sentence and requires the model to predict these masked words. Through this approach, the masked language model is able to observe both forward and backward in context, unlike traditional directed language models, which typically can only make predictions unidirectionally from left to right or from right to left. The next sentence prediction will also randomly select some sentence pairs during training, some of which are continuous sentence pairs and others are discontinuous sentence pairs. The model needs to determine whether the given sentence pair is an actual adjacent sentence before and after. Through this task, the next sentence prediction can learn the relationships between sentences, thereby better understanding the context and coherence of the text.

BERT first inputs a sentence through Token, Segment, and Position Embeddings, then enters a multi-layer Transformer encoder, and finally outputs a representation. BERT captures rich contextual information through bidirectional training and performs well in various NLP tasks [16-17]. The BERT model has two stages: pre-training and fine-tuning. Firstly, the input text is segmented into lexical units using the Word Piece algorithm, and special markers are added at the beginning and end of the sentence. Then add positional encoding information to enable the model to capture the order of vocabulary. In the model, BERT uses multi-layer Transformer encoders to obtain contextual information through bidirectional encoding and a self-attention mechanism, which can fully utilize context for bidirectional understanding. In the pre-training stage, BERT uses a masked language model to randomly mask some vocabulary to predict the masked words; it simultaneously predicts the next sentence to train inter-sentence relationships. During the fine-tuning phase, BERT adds task layers based on specific tasks, such as classification, named entity recognition, or question answering, output layers, and trains the entire model while retaining pre-trained parameters [18]. When outputting, BERT provides corresponding results based on different tasks: vectors are used in classification tasks, results are provided for each token in sequence annotation tasks, and the starting and ending positions of predicted answers are predicted in question answering tasks [19]. The processed data is extracted through a Long Short-Term Memory (LSTM), as shown in Fig. 3.

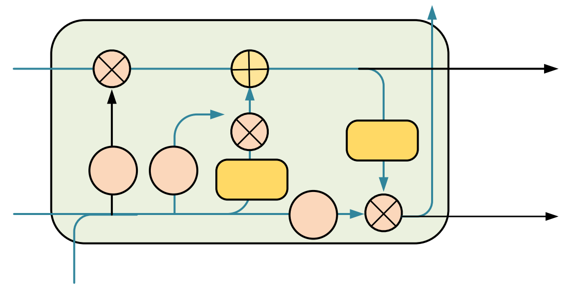


Fig. 3. LSTM structure.

In Fig. 3, first, the forget gate determines how much information to discard from the cell state, and gets the current input and the previous hidden state [20-21]. A value between 0 and 1 will be produced by the Sigmoid, representing the proportion of retained information. Its expression is shown in Eq. (1).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

In Eq. (1), f_t signifies the forget gate. σ represents the Sigmoid function. W_f signifies the linear relationship weight. h_{t-1} stands for the previous moment's hidden state. x_t signifies the data input at this moment. b_f signifies bias. If the output is close to 1, it means most information is retained. If the value is close to 0, it means discarding most of the information. The input gate controls new information input and determines what information is stored, as expressed in Eq. (2).

$$\begin{cases} i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t + b_C] \end{cases} \tag{2}$$

In Eq. (2), i_t represents the input gate which determines whether information passes through based on its importance. The output gate decides the amount of information that is output from the cell state to the hidden state. The information input is determined by the sigmoid layer and multiplied by the output of the Sigmoid gate, as shown in Eq. (3).

$$\begin{cases} o_t = \sigma(W_o [h_{t-1} . x_t] + b_o) \\ h_t = o_t \cdot \tanh(C_t) \end{cases} \tag{3}$$

In Eq. (3), o_t represents the output gate, which allows incoming information to affect the current time-step output. The sentence representation of contextual information in text is shown in Eq. (4).

$$X_L = BERT(L; \theta_L^{BERT}) \tag{4}$$

In Eq. (4), $BERT(\cdot)$ represents the BERT model. θ_L^{BERT} represents the relevant parameters of the BERT model.

B. *False News Recognition Model Based on Attention Mechanism and Multiple Features*

Due to the fact that a single text cannot fully judge fake news, multi-feature information are used to judge fake news, and emotional features in fake news are obtained through text to determine the authenticity of the news [22]. The study adopts the Bi-GRU structure to replace the LSTM. Bi-GRU is an improved recurrent neural network, similar to Bidirectional LSTM (Bi-LSTM), which processes input sequences through two GRU layers, forward and backward, to capture richer contextual information. Its structure is shown in Fig. 4.

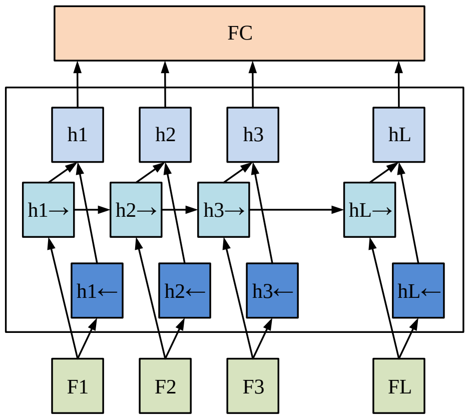


Fig. 4. Bi-GRU structure.

In Fig. 4, the Bi-GRU processes the input sequence through two GRU layers, one from front to back and the other from back to front, in order to capture richer contextual information [23]. Firstly, the input sequence is passed to both the forward GRU and backward GRU simultaneously. The former processes the original sequence and generates a forward hidden state sequence. The backward GRU processes the sequence in the reverse order and produces a backward sequence of hidden states. Then, the forward and backward hidden states are connected at each time step to form a bidirectional hidden state that contains contextual information before and after each time step. Finally, the bidirectional hidden state is passed to the output layer. In comparison to the LSTM, the GRU is simpler to set up and easier to use. To get more information from the text, the model is improved using a multi-head attention mechanism, as displayed in Fig. 5.

As shown in Fig. 5, this mechanism allows the model to simultaneously highlight various parts of the input sequence by calculating multiple attention heads in parallel, thereby capturing richer contextual information [24]. Each head undergoes linear transformation on queries, keys, and values, then calculates scaled dot product attention. Finally, the outputs are concatenated and linearly transformed to produce the result. Each head calculates the scaled dot product attention, as displayed in Eq. (5).

$$Attention(Q,K,V) = soft\max\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (5)$$

In Eq. (5), Q, K, and V refer to the query matrix, key matrix, and value matrix. These three are transformed linearly to generate multiple independent query, key, and value subspaces, each having a header. The multi-head attention mechanism calculates multiple attention heads in parallel, thereby focusing on various aspects of the input sequence and capturing more fine-grained contextual information, thereby improving the model's performance. Its expression is displayed in Eq. (6).

$$MultiHead(Q,K,V) = Concat(head_1, head_2, ..., head_h)W^O \quad (6)$$

In Eq. (6), $head$ represents the calculation result of a head, as shown in Eq. (7).

$$head = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (7)$$

In Eq. (7), W represents the value of the weight matrix, but there is a network degradation in the model. Residual connection is introduced in the model. The fake news recognition model on the basis of an attention mechanism and multi-feature is shown in Fig. 6.

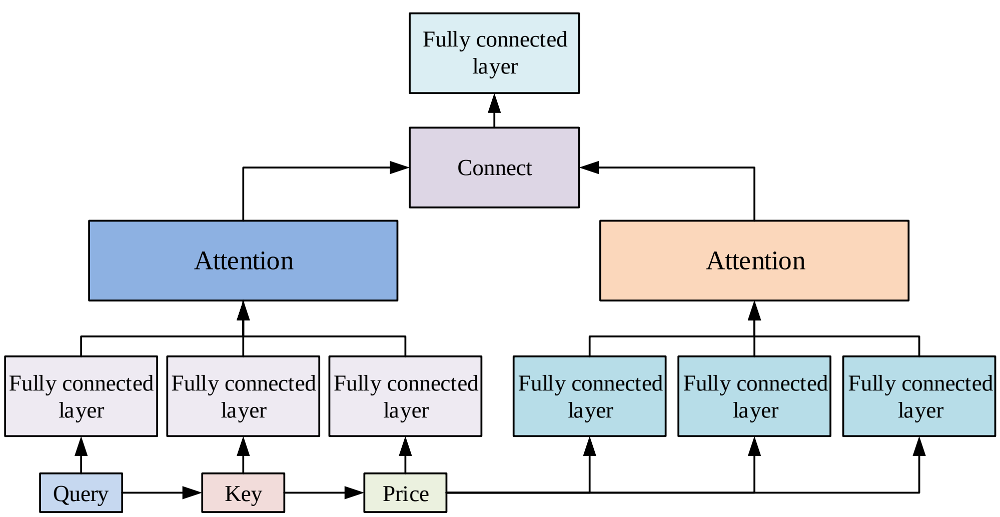


Fig. 5. Multi-head attention mechanism.

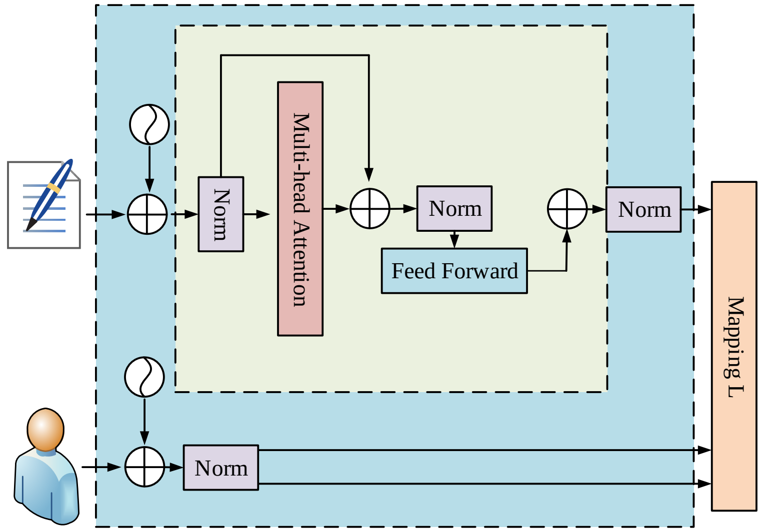


Fig. 6. Fake news recognition model based on an attention mechanism and multiple features.

As shown in Fig. 6, the model has a total of 4 layers, each of which has two modules: text mode and emotional text mode. The features of the model are extracted through a Transformer encoder, which is specifically designed to process input sequences and generate rich contextual representations. It is made up of several encoding layers stacked on top of each other, each layer consisting of two main parts: SAM and a feedforward neural network. Its expression is shown in Eq. (8):

$$F_m = Trans(T_L) \oplus Trans(T_A) \oplus Trans(T_V) \quad (8)$$

In Eq. (8), F_m represents the feature after multi-modal fusion. $Trans(\cdot)$ represents the encoder. The expression for sentiment classification through activation function and a fully connected layer is shown in Eq. (9).

$$\begin{cases} F_m^* = \mathrm{Re}\,LU(W_{f1}F_m + b_{f1}) \\ \hat{y}_m = W_{f2}F_m^* + b_{f2} \end{cases} \quad (9)$$

In Eq. (9), b represent bias. W signify weight. $\hat{y}_m$ represents multi-modal sentiment polarity.

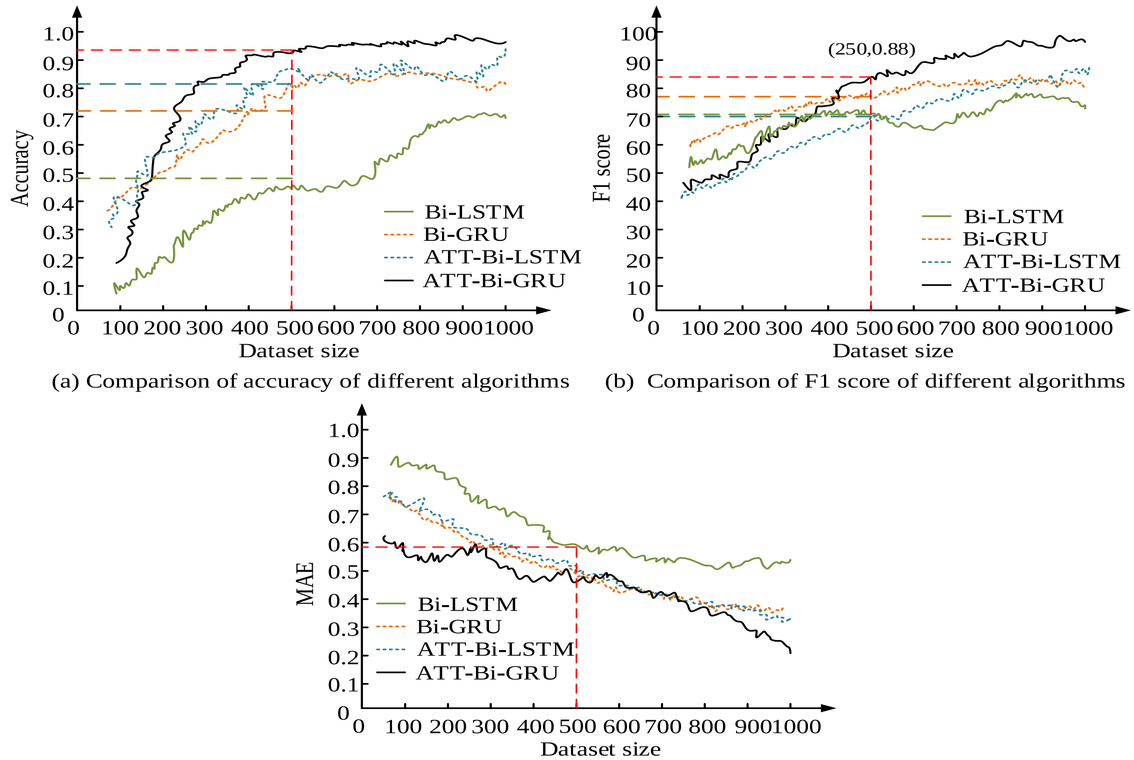
IV. Results

The first section uses publicly available datasets to analyze the performance. The second section uses simulation analysis to identify news with different types of emotions.

A. *Performance Analysis of False News Recognition Model Based on Multiple Features*

The central processor used in this experiment is Intel Core i5-8750H, the graphics processor is NVIDIA Geforce GTX2080Ti, the video memory is 8GB, and the memory is 16GB, taking Windows 10 as the operating system. To evaluate the effectiveness of the proposed ATT-Bi-GRU model, experiments were conducted using the publicly available LIAR dataset, developed by the Stanford University Computer Science Department. The LIAR dataset is widely recognized in the field of fake news detection and is specifically designed for assessing the veracity of political statements and short news articles. It contains 12,836 manually labeled short statements, each annotated with one of six truthfulness levels: true, mostly-true, half-true, barely-true, false, and pants-on-fire. Each data sample in the dataset includes not only the textual content of the news but also accompanying metadata such as speaker identity, political affiliation, venue, subject category, and context, which allows for multimodal feature fusion. For the purpose of binary classification in this study, we relabeled the samples into two categories: true (combining true, mostly-true, half-true) and false (combining barely-true, false, pants-on-fire). The dataset was randomly partitioned into 70 per cent training, 15 per cent validation, and 15 per cent testing subsets to ensure robust performance evaluation. The deigned model is named ATT-Bi-GRU, and the Bi-LSTM model, Bi-GRU and ATT-Bi-LSTM are introduced as comparison models for comparison. The accuracy, F1 score, and Mean Absolute Error (MAE) results are shown in Fig. 7.

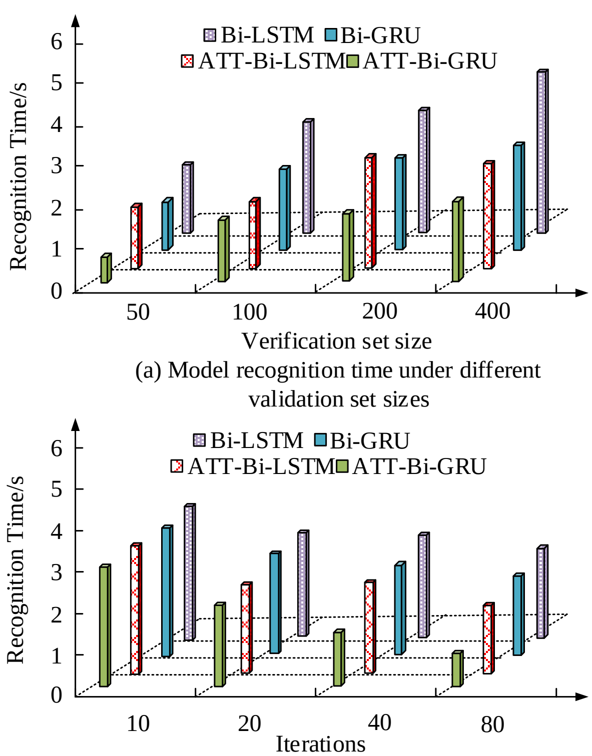
Fig. 7 illustrates the performance metrics of four distinct algorithmic models across various dataset sizes. In Fig. 7(a), the ATT-Bi-GRU model demonstrated peak recognition accuracy at a training set size of approximately 500 samples. Meanwhile, the accuracies of the other models consistently increased with larger dataset sizes. Specifically, when the training set reached 1000 samples, the recognition accuracies of the four models were recorded at 0.68, 0.79, 0.91, and 0.97, respectively. In Fig. 7(b), it is evident that the F1 scores for all four models increased with training set size, achieving scores of 76, 79, 93, and 97 at the 1000 sample size. Similarly, Fig. 7(c) reveals a trend of decreasing Mean Absolute Error values across the models with larger training datasets. At a dataset size of 1000, the MAE values recorded for the four models were 0.54, 0.47, 0.36, and 0.19, respectively. These findings underscore the effectiveness of the proposed ATT-Bi-GRU model, which outperformed the competing algorithms in terms of accuracy, F1 score, and MAE. Overall, the ATT-Bi-GRU model demonstrates superior performance metrics compared to its counterparts, highlighting its potential for effective fake news detection. The recognition efficiency is displayed in Fig. 8.



(a) Comparison of accuracy of different algorithms (b) Comparison of F1 score of different algorithms

(c) Comparison of F1-Score of different algorithms

Fig. 7. Comparison of Accuracy, F1 score, and MAE among different models.



(a) Model recognition time under different validation set sizes

(b) Model recognition time under different iteration times

Fig. 8. Analysis of model recognition efficiency.

Fig. 8 depicts the model recognition time under various conditions. From Fig. 8(a), the times for every model were 5.1s, 3.8s, 2.9s, and 2.2s, respectively, when the data set was 400. In Fig. 8(b), the number of iterations significantly enhanced model performance. The greater the number of iterations, the greater the performance improvement observed. When the iterations were set to 80, the recognition times for the four models were 2.5 s, 1.9 s, 1.7 s, and 0.8 s, respectively. The ATT-Bi-GRU model exhibited the best performance, demonstrating superior recognition efficiency. The dataset is divided into two different sizes, and the performance is compared. Table I displays the results.

Table I presents a comprehensive performance analysis of various models across two datasets. For Dataset 1, the models exhibited varying performances; notably, the ATT-Bi-GRU model achieved the highest performance, attaining an accuracy of 87.3% and an F1 score of 88.3%. The accuracy of the Bi-LSTM model is the lowest among all models, at merely 78.8%. For Dataset 2, the ATT-Bi-GRU model continued to perform well, achieving an accuracy of 87.5% and an F1 score of 88.9%. In contrast, the accuracy of the Bi-GRU model remained consistent across both datasets, with accuracies of 83.5% and 83.3%, respectively. The experimental results indicate significant differences in Mean Absolute Error (MAE) across the various models, with the ATT-Bi-GRU model exhibiting an MAE of only 0.05 in Dataset 2, further demonstrating its accuracy advantage. Overall, the ATT-Bi-GRU model outperforms other models in terms of both accuracy and F1 score, indicating its superior performance in managing these datasets.

TABLE I COMPREHENSIVE PERFORMANCE ANALYSIS

/	Model	Accuracy-2	F1	MAE	Correlation	Accuracy-7
Dataset 1	Bi-LSTM	78.8	79.8	0.35	0.587	74.2
	Bi-GRU	83.5	84.4	0.27	0.676	81.1
	ATT-Bi-LSTM	84.4	85.4	0.15	0.697	85.6
	ATT-Bi-GRU	87.3	88.3	0.08	0.699	85.8
/	Model	Accuracy-2	F1	MAE	Correlation	Accuracy-7
Dataset 2	Bi-LSTM	83.7	85	0.22	0.671	95.3
	Bi-GRU	83.3	84.8	0.11	0.662	95.5
	ATT-Bi-LSTM	84.0	85.9	0.09	0.672	95.1
	ATT-Bi-GRU	87.5	88.9	0.05	0.699	97.3

B. Simulation Result Analysis

To further validate the model, a simulation analysis is conducted on each model. The recognition performance is analyzed. The results are shown in Fig. 9.

In Fig. 9, TN represents news that is judged as real by the model, but is actually real news. FP stands for news that has been deemed true, but in reality, it is fake news. FN indicates that it has been judged as false news, but in reality, it is true news. TP stands for true news, but it is actually true news. Fig. 9 displays the recognition results. As for ATT-Bi-GRU, only a few false positives occurred, while in the Bi-LSTM model, many false positives occurred. Among the four algorithm models, the proposed ATT-Bi-GRU model exhibits good recognition performance. Three types of news with different emotional types are tested. The results are displayed in Fig. 10.

In Fig. 10, A, B, C, and D represent the Bi-LSTM, Bi-GRU, ATT-Bi-LSTM model, and ATT-Bi-GRU. Fig. 10(a) displays the recognition accuracy of various emotional types of news by various models, while Fig. 10(b) shows the false positive rate of different emotional types of news. According to Fig. 10(a), among the three types of news emotions, active emotion news had the highest recognition rate, while neutral emotion news had the lowest recognition rate. The ATT-Bi-GRU model had recognition accuracy of 95.6%, 93.2%, and 94.1% for active emotion news, negative emotion news, and neutral emotion news, respectively. According to Fig. 10(b), among the three types of news emotion, negative sentiment news had the lowest false positive rate, while neutral sentiment news had the highest recognition rate. The ATT-Bi-GRU model had false positive rates of 8.2%, 7.3%, and 9.4% for active sentiment news, negative sentiment news, and neutral sentiment news. The designed ATT-Bi-GRU has excellent performance for news of different emotional types. 60 researchers selected in six groups to assess, as displayed in Table II.

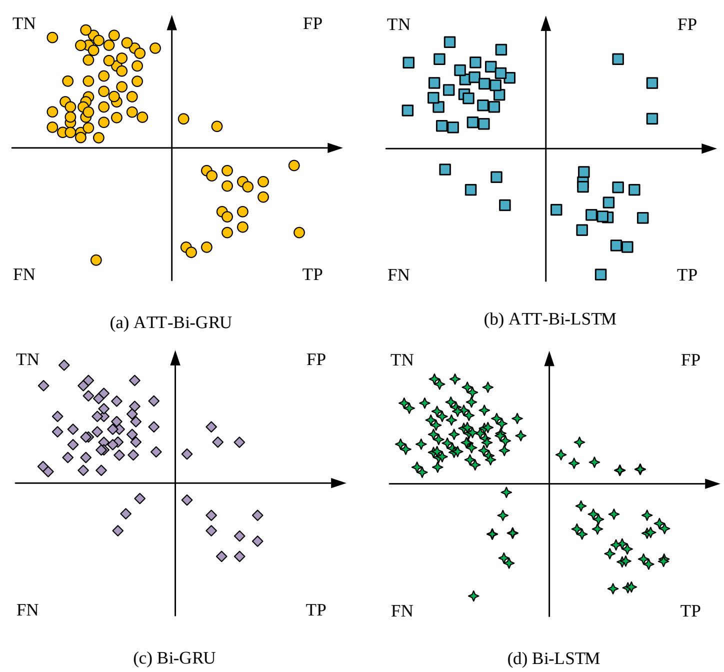
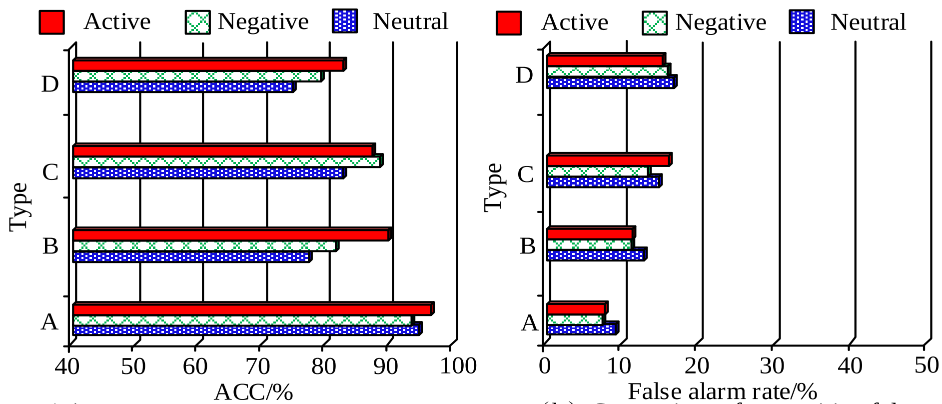


Fig. 9. Comparison of recognition performance of various algorithms.



（a）Comparison of recognition accuracy of different emotional types of news by various models

（b）Comparison of recognition false positive rate of different emotional types of news by various models

Fig. 10. Analysis of news recognition rates for different types of emotions by various models.

TABLE II MODEL RATINGS BY RESEARCHERS

Type	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
ATT-Bi-GRU	94.6	89.8	90.9	83.6	83.2	81.7
ATT-Bi-LSTM	80.5	84.7	80.1	79.1	81.5	78.7
Bi-GRU	79.1	82.1	77.3	77.3	79.3	76.6
Bi-LSTM	75.7	78.3	72.8	70.8	76.7	75.1

According to Table II, the six groups of researchers rated the ATT-Bi-GRU model as 94.6, 89.8, 90.9, 83.6, 83.2, and 81.7, respectively. The ratings for the Bi-LSTM model were 75.7, 78.3, 72.8, 70.8, 76.7, and 75.1. The ATT-Bi-GRU has received widespread praise from users. To evaluate the influence of key parameters on model performance, several ablation experiments were conducted. The parameters tested include the number of attention heads in the self-attention mechanism, the dimensionality of the GRU hidden layer, and the total number of training iterations.

TABLE III MODEL PERFORMANCE EVALUATION

Model Variant	Attention Heads	GRU Hidden Units	Accuracy (%)	F1 Score (%)	MAE	Training Time (s)
Bi-GRU (baseline)	-	128	83.5	84.4	0.27	4.3
Bi-GRU + Attention (4 heads)	4	128	91.2	92.6	0.21	3.9
Bi-GRU + Attention (8 heads)	8	128	94.6	95.5	0.15	4.5
Bi-GRU + Attention (8 heads) + 256D	8	256	97	97	0.19	4.8
Bi-GRU + Attention (8 heads) + 512D	8	512	95.9	96	0.23	5.6
Bi-GRU + Attention (12 heads) + 256D	12	256	96.7	96.8	0.2	5.3

According to Table III, when the number of attention heads increased from four to eight, the recognition accuracy improved from 94.6 per cent to 97 per cent, indicating that a higher number of attention heads enables better parallel extraction of diverse semantic features. However, beyond eight heads, the performance gains plateaued while computational cost increased significantly. Similarly, increasing the GRU hidden layer size from 128 to 256 enhanced the model's ability to retain contextual dependencies, improving the F1 score from 93 to 97. Yet, when expanded to 512, the model began to overfit, as seen in rising validation loss. The learning rate was also found to be crucial. A learning rate of 0.001 yielded the fastest convergence and best accuracy, while higher rates led to instability, and lower rates caused underfitting. In terms of iteration count, most performance improvements occurred within the first 80 iterations. Beyond that point, additional training brought marginal gains but increased computation time. Therefore, the optimal parameter configuration was determined based on a balance between accuracy, generalization, and computational efficiency.

V. Discussion

Accurate recognition of false news has become increasingly critical in the digital information ecosystem, as it directly affects public trust, media credibility, and the stability of societal cognition. While conventional fake news detection models based on single-modal text analysis have achieved some progress, they often suffer from limitations such as insufficient semantic understanding, poor emotional discrimination, and a lack of generalization across content types. These shortcomings hinder their performance in detecting misinformation embedded in emotionally charged or subtly manipulated narratives. In particular, models such as Bi-LSTM or single-directional GRU often struggle with long-distance dependencies and fail to selectively focus on the most informative parts of the text. To address these challenges, this study proposes a multimodal fake news recognition framework based on a bidirectional gated recurrent unit and a self-attention mechanism. By introducing the ATT-Bi-GRU model, the architecture enables forward and backward semantic modeling of news content while dynamically assigning attention weights to critical features such as emotional polarity, metadata, and textual salience. This deep fusion of multiple features effectively overcomes the bottlenecks of previous approaches. For example, compared with Bi-LSTM and Bi-GRU, the proposed model improved recognition accuracy by up to 28% and reduced mean absolute error by more than 60%. This is similar to the findings of Gorai J et al [25]. Particularly under high-iteration settings, the model achieved a recognition time of 0.8 seconds—demonstrating both high efficiency and accuracy. The substantial improvements are primarily attributed to the attention mechanism's ability to identify core emotional signals and context markers, coupled with Bi-GRU's capability to model long-range dependencies in both directions. Another major advantage of the ATT-Bi-GRU model is its strong adaptability and generalization across different types of emotional news. In scenarios involving active, negative, or neutral sentiment news, the model consistently maintained high recognition accuracy above 93%, and significantly lowered false positive rates—especially for emotionally ambiguous texts. This is similar to the findings of Wu D et al. [26]. This is due to the model's integration of sentiment-enhanced embeddings and context-aware attention filtering, which allows it to dynamically adapt to varying linguistic tones and deceptive strategies. Moreover, the model demonstrates robustness across both small and large training datasets, making it suitable for practical deployment in real-time misinformation detection systems.

In summary, by leveraging a hybrid structure of Bi-GRU and multi-head attention, the proposed model effectively resolves the long-standing issues of low robustness, poor emotional sensitivity, and inefficient feature extraction in fake news detection tasks. Its superior performance underlines its potential as a scalable, adaptive, and highly accurate solution for combating the spread of misinformation in dynamic and emotionally complex media environments.

VI. Conclusion

A multi-feature based fake news recognition model was proposed to address the serious impact of fake news on society. The model used Bi-GRU and SAM to judge the fake news in news texts. The recognition accuracy was 0.68, 0.79, 0.91, and 0.97, respectively, for Bi-LSTM, Bi-GRU, ATT-Bi-LSTM, and ATT-Bi-GRU. The F1 scores were 76, 79, 93, and 97, respectively, while the MAE values were 0.54, 0.47, 0.36, and 0.19. With the data set quantity of 400, required time for every models was 5.1 seconds, 3.8 seconds, 2.9 seconds, and 2.2 seconds. When the count of the iteration was 80, the time was 2.5 seconds, 1.9 seconds, 1.7 seconds, and 0.8 seconds, respectively. Selecting three different types of news with different emotions for detection, the ATT-Bi-GRU model achieved recognition accuracy of 95.6%, 93.2%, and 94.1% for active, negative, and neutral emotional news. The ATT-Bi-GRU can effectively improve the recognition ability of fake news, and also presents good recognition performance for news of different emotional types, providing a new and effective method for detecting fake news. Although the proposed ATT-Bi-GRU model demonstrates strong performance in fake news detection, several avenues remain for further exploration. First, the model's effectiveness on long-form textual content, such as full-length news articles or reports, remains unverified. Future work will explore hierarchical encoding strategies or long-sequence transformer encoders, such as Longformer or BigBird, to enhance scalability in document-level classification. Second, while this study integrates sentiment and metadata features, future research can extend the model toward multi-modal fusion by incorporating visual and auditory modalities, such as image-caption pairs or speaker tone in video news. This would be especially relevant for fake news detection in multimedia and social platforms. Third, the current model is trained and evaluated on English datasets. To support broader applicability, future research will consider cross-lingual and cross-cultural fake news detection by incorporating multilingual pre-trained models and annotated datasets in other languages, such as Chinese, Arabic, and Spanish. Lastly, deployment aspects such as real-time detection in streaming environments, model compression for edge devices, and adaptive learning under evolving misinformation patterns will be considered to bring the proposed method closer to practical application in news media monitoring systems.

Funding

The research is supported by Shaanxi Provincial Department of Education 2020 general special scientific research, research on the dissemination path of the “New Shaanxi Merchants Spirit” from the perspective of technological empowerment, 20JK0280.

References

[1] Mao Y, Wang T, Duan M, Zhang M. A new mooring failure detection approach based on hybrid LSTM-SVM model for semi-submersible platform.Ocean engineering, 2023, 275(5):1-23.DOI:10.1016/j.oceaneng.2023.114161.

[2] Rongjiang Cai. Can Environmental Information Disclosure Reduce Stock Price Synchronization of Listed Companies? — Empirical Analysis Based on China’S Heavy Pollution Industry. Advanced Management Science, 2022, 11(2).

[3] Rong Tian, Simin Chen, Enze Pan. Research on the Influence of Product Information on Consumers’ Willingness To Shop Online. Advances In Industrial Engineering and Management, 2023, 12(1).

[4] Tuli S, Gill S S, Garraghan P, Buyya R, Casale G, Nicholas R. START: Straggler Prediction and Mitigation for Cloud Computing Environments

Using Encoder LSTM Networks.IEEE transactions on services computing, 2023, 16(1): 615-627.DOI:0.1109/TSC.2021.3129897.

[5] Wang X Y. Research on Emotional Music App Design Under the Needs of Office Workers. Information Management and Computer Science, 2022, 5(1): 01-02.

[6] Xinyi Wang. Research on Emotional Music App Design Under the Needs of Office Workers. Information Management and Computer Science, 2022, 5(1).

[7] Zhao Y, Guo M, Sun X, Chen X, Zhao F. Attention-based sensor fusion for emotion recognition from human motion by combining convolutional neural network and weighted kernel support vector machine and using inertial measurement unit signals. IET signal processing, 2023, 17(4): 241-259.DOI:10.1049/sil2.12201.

[8] Ahmed M R, Islam S, Islam A K M M, S.An ensemble 1D-CNN-LSTM-GRU model with data augmentation for speech emotion recognition.Expert Systems with Applications, 2023, 218:119633-.DOI:10.1016/j.eswa.2023, 218(5):154-167.DOI:10.1016/j.eswa.2023.119633.

[9] Chen L, Wang K, Li M, et al.K -Means Clustering-Based Kernel Canonical Correlation Analysis for Multimodal Emotion Recognition in Human–Robot Interaction.IEEE Transactions on Industrial Electronics, 2023, 70(1):1016-1024.DOI:10.1109/TIE.2022.3150097.

[10] Pushpalatha M N, Meherishi H, Vaishnav A, Pillal R A, Gupta A. Facial emotion recognition and encoding application for the visually impaired.Neural computing & applications, 2023, 35(1), 749-755.DOI:10.1007/s00521-022-07807-z.

[11] Zhang Y, Wang J, Liu Y, Rong L, Zhang Q, Song D, Tiwari P, Qin J. A Multitask learning model for multimodal sarcasm, sentiment and emotion recognition in conversations.Information Fusion, 2023, 93(15):282-301.DOI: 10.1016/j.inffus.2023.01.005.

[12] Wen J, Jiang D, Tu G, Liu C, Cambria E. Dynamic interactive multiview memory network for emotion recognition in conversation.Information Fusion, 2023, 91(13):123-133.DOI:10.1016/j.inffus.2022.10.009.

[13] Yang M, Liu W, Zhang W, Wang M, Fang X. Bearing Vibration Signal Fault Diagnosis Based on LSTM-Cascade CatBoost.Journal of Internet Technology, 2022, 23(5): 1155-1161.DOI:10.53106/160792642022092305024.

[14] Ahamed S A, Ravi C.Novel deep learning model for bitcoin price prediction by multiplicative LSTM with attention mechanism and technical indicators.International Journal of Engineering Systems Modelling and Simulation, 2022,13(2):164-177.DOI:10.1504/IJESMS.2022.123341.

[15] Cai Y, Xu G, Yang D. On-line multi-gas component measurement in the mud logging process based on Raman spectroscopy combined with a CNN-LSTM-AM hybrid model.Analytica chimica acta, 2023, 1259(21):154-162.DOI: 10.1016/j.aca.2023.341200.

[16] Ma G, Jin C, Wang H, Li P, Kang H. Study on dynamic tension estimation for the underwater soft yoke mooring system with LSTM-AM neural network.Ocean engineering, 2023, 267(7):1-15.

[17] Singh Y P, Lobiyal D K. Automatic prediction of epileptic seizure using hybrid deep ResNet-LSTM model.AI communications, 2023, 36(1): 57-72.DOI: 10.1016/j.oceaneng.2022.113287.

[18] Darekar R V, Chavan M, Sharanyaa S, Ranjan N M. A hybrid meta-heuristic ensemble based classification technique speech emotion recognition.Advances in engineering software, 2023, 180(21):154-168.DOI:10.3233/aic-220177.

[19] Dirik M. Optimized Anfis Model with Hybrid Metaheuristic Algorithms for Facial Emotion Recognition.International Journal of Fuzzy Systems, 2023, 25(2): 485-496.DOI:10.1007/s40815-022-01402-z.

[20] Rooj S, Routray A, Mandal M K. Feature based analysis of thermal images for emotion recognition.Engineering Applications of Artificial Intelligence, 2023, 120(24):1058-1065.DOI:10.1016/j.engappai.2022.105809.

[21] Zheng Q, Wang R, Tian X, Yu Z, Wang H, Elhanashi A, Saponara S. A real-time transformer discharge pattern recognition method based on CNN-LSTM driven by few-shot learning.Electric Power Systems Research, 2023, 219(6):1-12.DOI:10.1016/j.epsr.2023.109241.

[22] Wei J, Wu X, Yang T, Jiao R. Ultra-short-term forecasting of wind power based on multi-task learning and LSTM.International Journal of Electrical Power & Energy Systems, 2023, 149(5): 1249-1261.DOI: 10.1016/j.ijepes.2023.109073.

[23] De Lope J, Grana M. An ongoing review of speech emotion recognition.Neurocomputing, 2023.DOI:10.1016/j.neucom.2023, 528(8):1-11.DOI:10.1016/j.neucom.2023.01.002.

[24] Pal S, Roy A, Shivakumara P, Pal U. Adapting a Swin Transformer for License Plate Number and Text Detection in Drone Images. Artificial Intelligence and Applications, 2023, 1(3), 145-154.DOI:10.47852/bonviewAIA3202549.

[25] Gorai J, Shaw D K.Semantic difference-based feature extraction technique for fake news detection.Journal of Supercomputing, 2024, 80(15):22631-22653.DOI:10.1007/s11227-024-06307-2.

[26] Wu D, Tan Z, Zhao H.Domain- and category-style clustering for general fake news detection via contrastive learning.Information Processing and Management, 2024, 61(4):103725-103728.DOI:10.1016/j.ipm.2024.103725.