IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 13 April 2024, accepted 1 June 2024, date of publication 10 June 2024, date of current version 18 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3411926

RESEARCH ARTICLE

Elevating Fake News Detection Through Deep Neural Networks, Encoding Fused Multi-Modal Features

AHMED HASHIM JAWAD ALMARASHY[1], MOHAMMAD-REZA FEIZI-DERAKHSHI[1], AND PEDRAM SALEHPOUR[2], (Member, IEEE)

[1]Computerized Intelligence Systems Laboratory, Department of Computer Engineering, University of Tabriz, Tabriz 51666, Iran
[2]Department of Computer Engineering, University of Tabriz, Tabriz 51666, Iran

Corresponding author: Mohammad-Reza Feizi-Derakhshi (mfeizi@tabrizu.ac.ir)

ABSTRACT Textual content was initially the main focus of traditional methods for detecting fake news, and these methods have yielded appointed results. However, with the exponential growth of social media platforms, there has been a significant shift towards visual content. Consequently, traditional detection methods have become inadequate for completely detecting fake news. This paper proposes a model for detecting fake news using multi-modal features. The model involves feature extraction, feature fusion, dimension reduction, and classification as its main processes. To extract various textual features, a pre-trained BERT, gated recurrent unit (GRU), and convolutional neural network (CNN) are utilized. For extracting image features, ResNet-CBAM is used, followed by the fusion of multi-type features. The dimensionality of fused features is reduced using an auto-encoder, and the FLN classifier is then applied to the encoded features to detect instances of fake news. Experimental findings on two multi-modal datasets, Weibo and Fakeddit, demonstrate that the proposed model effectively detects fake news from multi-modal data, achieving 88% accuracy with Weibo and 98% accuracy with Fakeddit. This shows that the proposed model is preferable to previous works and more effective with the large dataset.

INDEX TERMS Social media platforms, fake news detection, multi-model features, deep learning, fusion, auto-encoder, dimensionality reduction.

I. INTRODUCTION

In today's digital age, the rapid spread of information through social media and online platforms has revolutionized how we consume news. However, this exceptional proliferation of information has simultaneously facilitated the propagation of disinformation, thereby developing the prevalent issue of bogus news. Fake news consists of intentionally contrived or misleading information that is presented as authentic news, intending to deceive and manipulate public sentiment or profit, and exploiting the followers of establishments that have a large fan base through the distribution of this misinformation. For example, Kylian Mbappe's fake contract, as shown in Figure 1. The proliferation of fake news poses significant challenges to society, affecting political discourse, and public perception, and even influencing critical decision-making processes. For example, there were numerous fabrications in the wake of the 2020 U.S. presidential election. Many voters who were preoccupied with election news erroneously believed that election fraud had occurred, with 40 percent of them preserving that Biden's election was illegal [1]. As the impact of fake news continues to grow, there is an urgent need to develop robust and reliable methods for its detection.

Fake news detection (FND) is an interdisciplinary domain that combines expertise in natural language processing (NLP), machine learning (ML), data analysis, and media literacy. The primary goal is to identify and distinguish between legitimate news articles and fabricated, misleading, or biased content.

ML algorithms play a pivotal role in fake news detection, leveraging vast amounts of labeled data to learn patterns and

The associate editor coordinating the review of this manuscript and approving it for publication was Jiankang Zhang.

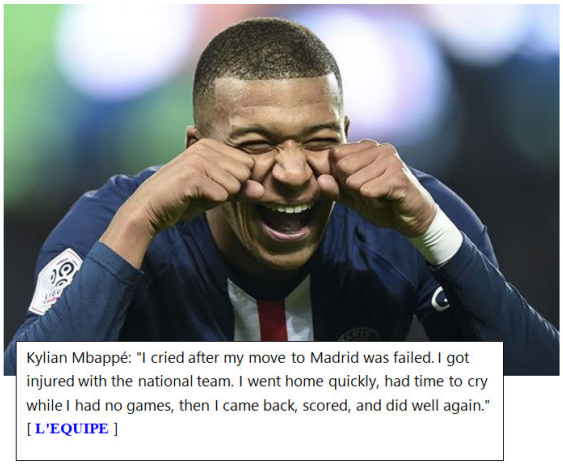

FIGURE 1. An example of a fake news post [29] with an image that was posted on an authenticated account [30] with a different caption.

features that distinguish between trustworthy and deceptive information [2]. These algorithms can analyze textual content, metadata, user behavior, and the source's credibility to make informed judgments about the veracity of news articles. Moreover, with the advancement of deep learning techniques, researchers and practitioners are continually developing sophisticated models capable of handling the dynamic and evolving nature of fake news. Deep neural networks can automatically extract complex patterns, enabling the detection of subtle linguistic cues and context-specific signals that might indicate the presence of fake news. Despite the ongoing efforts, FND remains a challenging task due to the adaptive tactics employed by misinformation propagators. The detection process needs to be agile, able to adjust and improve as new forms of misinformation emerge. In this pursuit, collaboration between academia, industry, and policymakers is vital to developing comprehensive strategies and systems to combat the fake news epidemic effectively. By implementing effective detection techniques and media literacy advertisements, we can enable individuals to evaluate information critically and foster a society that is better informed and more resilient to the threats posed by misinformation.

In this research paper, we delve into the techniques, methodologies, and challenges associated with FND. We explore the latest advancements in ML and NLP, discussing how these technologies can be harnessed to protect the integrity of information in the digital era. By understanding the landscape of fake news and the tools for its identification, we take a significant step toward fostering a more trustworthy and reliable information ecosystem. This paper is composed of multi-type feature extracting modules and a feature fusion module, then adopts an attention mechanism to reduce the dimensionality of fusion features. The feature extraction modules include text feature extractors and a visual feature extractor.

A pre-trained BERT module to extract contextual features, the deep neural network convolutional neural network (CNN) and a gated recurrent unit (GRU) to extract spatial and sequential features subsequently. ResNet-50 combined with the CBAM attention mechanism for extracting the visual feature. The concatenate module will fuse the text and image features to get fusion features. An auto-encoder reduces the dimensionality of fusion feature, then fed the reduction features into a fast learning network classifier to get detection results. The model integrates the features of multiple models more fully so that it can learn the deeper correlation between the models, which will be conducive to improving the performance of detection tasks. The main contributions of this paper are mentioned below:

- Employed a novel stacked model of state-of-the-art multi-modal deep neural networks for efficiently classifying between fake and real news.
- The deep learning ensemble model is proposed by leveraging the power of deep neural networks to extract multi-type features and then fuse these features.
- Adopted structured multi-modal datasets that contain pairs of texts and images to get more features in order to elevate the fake news detection.
- Dimensionality reduction of the fused feature via the auto-encoder to enhance the classifier results.

The following parts of this work are structured as follows: A literature review has been provided in Section II. We have presented and discussed the materials and methodology of the proposed model in Section III. In Section IV, the discussion and experimental results are presented. In Section V, the conclusion and future work direction are discussed.

II. LITERATURE REVIEW

Researchers are increasing their efforts to identify answers in response to the rise of misleading content on social media. A multitude of scholarly investigations have been conducted in this specific domain; we shall now highlight a carefully chosen subset of them. Prior research in this domain, similar to numerous other domains in NLP, mostly concentrated on probabilistic techniques. Nevertheless, subsequent to the introduction of deep learning (DL), numerous researchers embraced DL and made progress in this field. In the beginning, research efforts were focused on contrasting probabilistic approaches with DL methods. Following that, they advanced towards proposing more complex DL models. For example, article [3] provides a comprehensive study of the methods employed to identify misinformation on social media, encompassing classifications of fake news according to social and psychological theories, current algorithms analyzed through the lens of data mining, evaluation metrics, and representative datasets.

The authors of [4] examine cutting-edge and advanced methods for detecting fake news and elaborate on the dataset and NLP techniques utilized in prior investigations. A thorough examination of DL-based methodologies has

been presented in order to classify illustrative approaches into distinct categories. In [5], propose a novel higher-order user-to-user mutual-attention progression (HiMaP) method to capture the cues related to the authority or influence of the users by modelling direct and indirect (multi-hop) influence relationships among each pair of users present in the propagation sequence. The authors of [6] suggest a model to reduce the feature vectors' dimensionality before feeding them to the classifier. SAFE is a method described in [7] that analyzes news articles for multi-modal information to detect fake news. To get started, neural networks are utilized to autonomously extract textual and visual features for news representation. Further investigation is conducted into the relationship between the extracted features across modalities. The relationship between such visual and textual representations of news information are also jointly learned and utilized to predict fake news.

A multi-modal variational auto-encoder (MVAE) network is presented in [8] that combines a binary classifier with a bimodal variational auto-encoder to detect fake news. Three primary components comprise the model: an encoder, a decoder, and a module for detecting fake news. The variational auto-encoder can be used to learn probabilistic models for latent variables by finding the best bound on the observed data's marginal likelihood.

The authors of [9] present an innovative hybrid system for detecting fake news. This system merges the strengths of linguistic and knowledge-based approaches by utilizing two distinct sets of features: linguistic and a novel set of knowledge-based features. The system's performance on a simulated news dataset demonstrates that it is capable of achieving a commendable level of accuracy in identifying fake news.

A novel multi-modal topic memory network (MTMN) is described in [10]. It makes a good representation by using multi-modal fusion to take advantage of the connections between modes within each mode as well as the connections between modes between text terms and image regions. Reference [11] A multi-level multi-modal cross-attention network (MMCN) is presented in this article; it employs a network to facilitate cross-attention between various modalities. The MMCN has been purposefully developed to integrate the feature embedding of image regions and text words by concurrently taking into account duplicate data relationships and various modalities.

However, most of the studies mentioned do not take the variety of features in multi-modal datasets into account, resulting in limited results. In addition, most of the models are unsuccessful in obtaining adequate detection performance. To overcome these limitations, we employ multi-model feature extraction and dimensionality reduction for the fusion features, followed by the classification stage.

III. MATERIALS AND METHODS

The overall proposed model is shown in Figure 2.

This research paper employs a combined analysis of textual and visual data to assess the reliability of news. Based on this, we propose a multi-modal deep neural network with a fusion module to obtain deep connections among textual and visual features. This section covers the proposed model in detail.

A. FEATURE EXTRACTION

- Textual-Feature-Extractors

1) PREPROCESSING

Preprocessing of the text data is required, involving techniques such as stop word elimination, tokenization, and removal of punctuation. These processes can greatly assist in the selection of the most significant statements and enhance the performance of the model [12].

2) EMBEDDING

Utilizing the BERT pre-trained language model, this module is primarily tasked with transforming every word in the preprocessed sentence into a dense vector of the same dimension [13]. BERT accepts an entire sentence as input, in contrast to the static word embedding method which converts a word to a vector by consulting the word representation table directly. It uses the hidden state of the final hidden layer or the second-to-last hidden layer to represent each word in the sentence dynamic vector representation. Thus, various contexts will result in distinct vector representations for a given word, which provides a more precise expression of its semantics [14]. The procedures for the overall pre-training and subsequent fine-tuning of BERT are mentioned in [15]. Except for the output layers, exact configurations are utilized in both the pre-training and fine-tuning processes. Parameters derived from pre-existing models are employed to initialize models for different subsequent tasks. Each parameter is meticulously adjusted during the entire procedure. The special symbol [CLS] comes before each input example, and [SEP] serves as a separator token. For example, [SEP] is used to separate questions and answers, as demonstrated in Figure 3.

3) CNN

We have presented a comprehensive analysis of the benefits of employing convolutional neural networks (CNNs) for extracting features in NLP tasks. Convolutional neural networks, known for their remarkable performance in computer vision tasks, have also proven to be highly effective in NLP tasks. A key benefit of convolutional neural networks in NLP is their capacity to effectively collect and analyze local patterns and relationships present in the text [16]. CNN can use one-dimensional convolutions (Conv1D) to get local features by treating words or characters as one-dimensional data. This is especially helpful for tasks that need to capture n-grams or short phrases, like sentimental analysis or finding important parts of text categorization. CNN can handle inputs of different lengths well, which makes it a good choice for

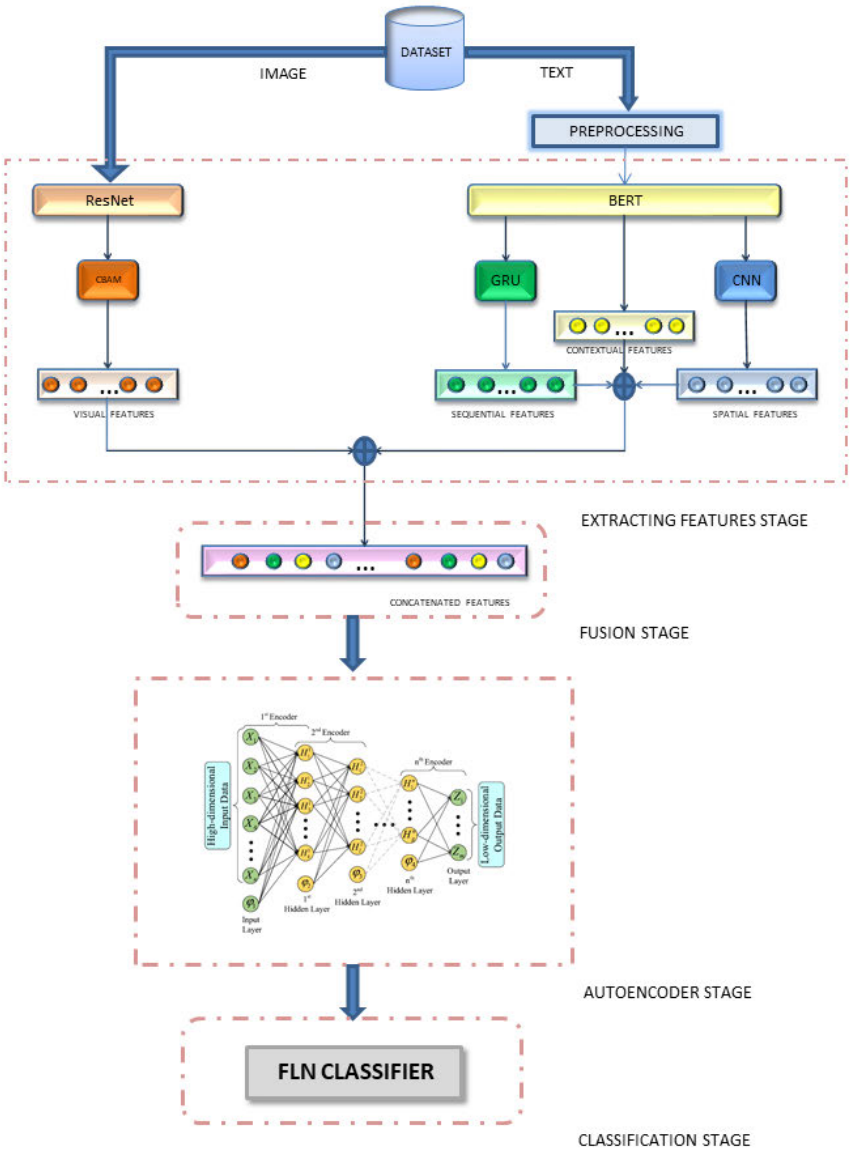

FIGURE 2. The proposed model stages.

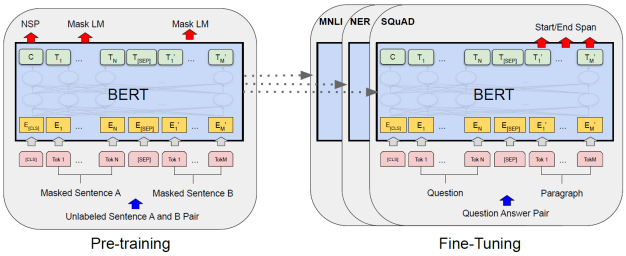

FIGURE 3. Bert procedures [15].

handling texts of different lengths. For many NLP jobs, these models can be used because they are flexible and do not require set input sizes. The number of filters and the kernel size must be set to train the CNN. Figure 4 shows a picture of the Conv1D process. CNN can get hierarchical statements of the text and local features at different levels by setting up many convolutional layers with various filter sizes and hyperparameters. This lets them get useful information from the data they put in.

4) GRU

The gated recurrent unit (GRU), which is a variation of LSTM, has a reduced number of model parameters and exhibits superior training efficiency when compared to some other LSTM models. It has successfully extracted the main characteristic from the text. GRU can model sequential dependencies, capture short-term memory, and address the vanishing gradient problem. The gating mechanism allows it

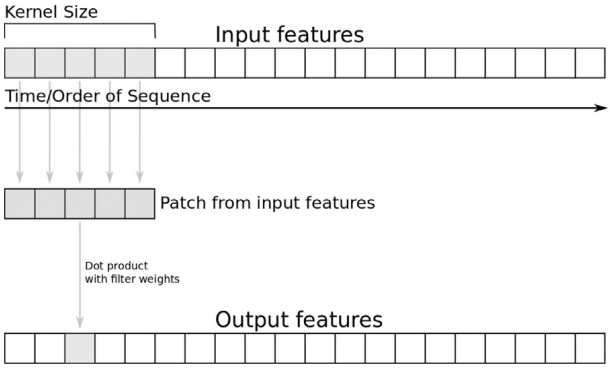

FIGURE 4. A Conv1D operation.

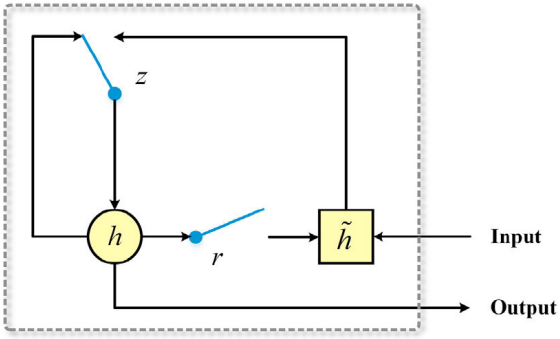

FIGURE 5. GRU cell.

to selectively update and utilize information from previous time steps, making them suitable for tasks involving sequential data, including NLP tasks, such as machine translation, named entity recognition, and classification. Figure 5 depicts the overall architecture of the GRU, as described in [17]. In GRU, there are two gates: an update gate (z) and a reset gate (r). These gates control the amount of information that is updated and forgotten, respectively. Furthermore, h and $\tilde{h}$ correspondingly convey current and updated information. The computation of z, r, h, and $\tilde{h}$ at time step s is conveniently given as follows:

$$\mathbf{z}_s = \sigma(W^{(z)}x_s + U^{(z)}h_{s-1}) \quad (1)$$

$$\mathbf{r}_s = \sigma(W^{(r)}x_s + U^{(r)}h_{s-1}) \quad (2)$$

$$\tilde{\mathbf{h}}_s = tanh(W^{(h)}x_s + U^{(h)}(h_{s-1} \odot rs)) \quad (3)$$

$$\mathbf{h}_s = (1 - z_s) \odot \tilde{h}_s + z_s \odot h_{s-1} \quad (4)$$

where σ is a nonlinear function, e.g., the sigmoid function, x_s denotes the input vector at the time step s, and $\odot$ represents an element-wise multiplication. $W^{(z)}$, $W^{(r)}$, $W^{(h)}$, $U^{(z)}$, $U^{(r)}$, and $U^{(h)}$ are all weights to be learned.

- Visual-Feature-Extractors

Previous methods commonly employed a VGG-based model to extract visual features for the analysis of multi-modal data, specifically in the context of visual content processing, such as images [18]. Nevertheless, compared to the features extracted by VGG, those from ResNet are more representative and discriminative. Thus, we utilize the ResNet model in our work to acquire the visual features of image data. Certain existing approaches solely rely on the attributes of the last layer of ResNet, which, under certain circumstances, disregard a significant amount of intricate visual data. To improve the representation of visual semantic information, we derive detailed area features from the image using the second-to-last layer of ResNet. The visual feature extractor utilizes the ResNet model [19], which has undergone pre-training on a vast array of picture datasets, to extract visual features.Residual network enhances the transformation capabilities of neural networks by introducing residual connections, also known as skip connections. These connections allow the network to bypass one or more layers and add the input of those layers directly to their output. This approach helps mitigate vanishing gradients, which is common in intense networks and can lead to decreasing model accuracy as the network depth increases. ResNet addresses the issue of reducing model accuracy with increased depth by using residual connections, which facilitate more effective training of deep networks by maintaining gradient flow and allowing layers to learn residual mappings more easily. The visual feature extractor in this paper utilizes the ResNet-50 model for extracting visual features. Combining the convolutional attention module (CBAM) [20] into the model in this study is meant to help it focus on the important parts of the image. The CBAM module will deduce the attention weights for a given intermediate feature graph by considering two separate dimensions (channel and space) sequentially. It will then multiply the attention weights with the input feature map to achieve adaptive feature optimization, as shown in Figure 6. A fully connected layer is added to the output of CBAM-ResNet-50 to keep the original model structure of ResNet-50 and use its pre-training parameters to stop overfitting. This makes sure that the text has the same size as the image's features and hidden state.

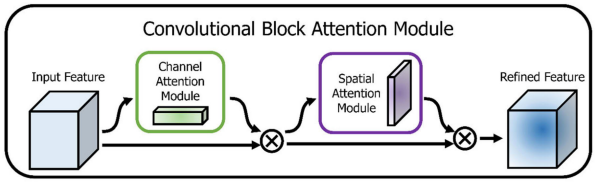

FIGURE 6. The CBAM module.

B. FUSION FEATURES

The series method results in feature fusion. The text feature $\mathbf{R}_T$, and visual feature $\mathbf{R}_V$ that are extracted are merged to obtain the fusion feature. The fusion process is as follows:

$$\mathbf{R} = concat(\mathbf{R}_T + \mathbf{R}_V) \quad (5)$$

where $\mathbf{R}$ is the input of the auto-encoder phase.

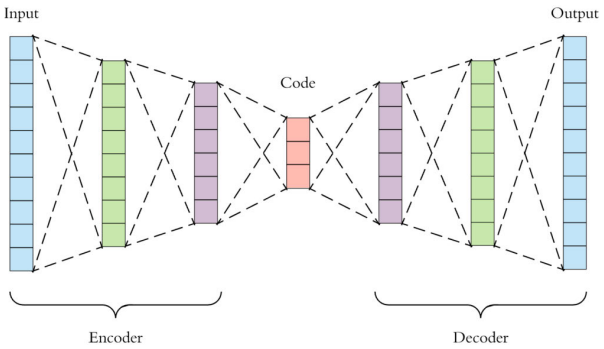

FIGURE 7. Autoencoder structure.

C. DIMENSIONALITY REDUCTION

In this research endeavor, we delve into the application of auto-encoders for the reduction of dimensionality in fusion features [21], [22]. Auto-encoders, a class of neural networks, excel at learning compact representations of input data by employing an encoder-decoder architecture. The encoding process captures the essential information from the fusion features, resulting in a lower-dimensional representation, while the subsequent decoding step reconstructs the original input as shown in Figure 7. The dimensionality reduction is achieved through mathematical operations at each layer of the auto-encoder. Specifically, the encoding function h(x) transforms the input x into a compressed representation h using weights (W) and biases (b).

Encoder mapping:

$$\mathbf{h} = f(\mathbf{X}.\mathbf{W} + \mathbf{b}) \tag{6}$$

here, f is typically a nonlinear activation function.

The decoding process, aiming to reconstruct the input, whereÂ (H)Â is the encoded vector from hidden layer, ($\bar{\mathrm{W}}$) is the weights associated to hidden layer neurons with biases ($\bar{\mathrm{b}}$).

Decoder mapping:

$$\bar{\mathbf{x}} = g(\bar{\mathbf{W}}.\mathbf{H} + \bar{\mathbf{b}}) \tag{7}$$

The ultimate goal is to minimize the reconstruction error, encouraging the auto-encoder to capture the most salient features in the reduced space. This study shows how auto-encoders can be used to reduce the number of dimensions in fusion features. It also goes into detail about the mathematical methods used for this purpose, which is useful for the fields of feature learning and representation optimization.

D. CLASSIFICATION STAGE

In our proposed model, we utilize the Fast Learning Network (FLN) classifier [23] to classify the encoded features. The FLN exhibits a distinctive aspect in its procedure of initializing weights and constructing weighted connections. A random method is used to set the weights and biases of the hidden layer. Merely start with random weights, the network can carefully explore different parts of the weight space and find more appropriate responses. The FLN aims to optimize the learning efficiency and precision of the network by building connections between the input nodes, the hidden layer, and the output nodes as shown in Figure 8. Least squares methods are used to find the best line or gradient for a set of data points. This connectivity design facilitates a more direct flow of information from the input to the output, which has the potential to enable the network to catch significant aspects and generate precise predictions. Hence, the FLN successfully addressed most of the limitations associated with conventional learning methods while simultaneously demonstrating an exceptionally high learning velocity [12].

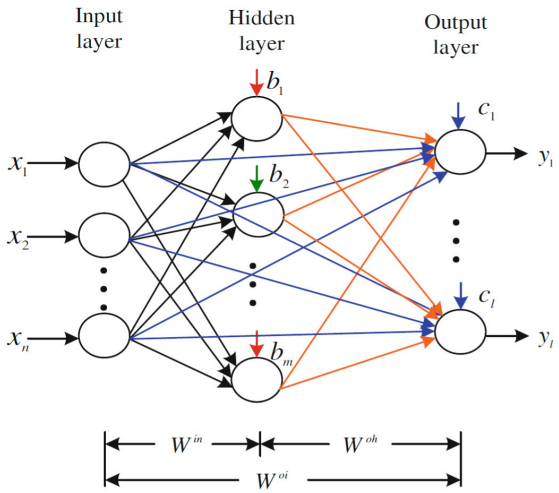

FIGURE 8. The FLN structure.

TABLE 1. Confusion matrix.

Actual	Predicted:Fake news	Predicted:Real news
Fake news	TP	FN
Real news	FP	TN

IV. EVALUATION CRITERIA

To assess the efficacy of the model suggested in this research, it was evaluated based on four key metrics: accuracy, precision, recall, and F_1 score [24], [25].

The confusion matrix of classification results, where (TP) and (FP) are the numbers of correctly and incorrectly classified instances of the positive class, and (TN) and (FN) are the numbers of correctly and incorrectly classified instances of the negative class, respectively.

According to Table 1, we compute the following metrics and mathematically expressed them in Equations 8, 9, 10, and 11.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

TABLE 2. The statics of the Weibo dataset.

News	Labeled
Real posts	4779
Fake posts	4749
Attached post images	9528

Accuracy (Acc) refers to the ratio of accurately predicted outcomes to the total number of outcomes.

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

Recall (Re): the ratio of accurately predicted fake news results to the total number of fake news results.

$$Precision = \frac{TN}{TN + FP} \tag{10}$$

Precision (Pr) refers to the ratio of correctly recognized instances of fake news to the total number of instances that have been detected as fake news.

$$F_1 = 2\frac{RePr}{Re + Pr} \tag{11}$$

F_1 score: the harmonic mean of recall and accuracy.

V. EXPERIMENTS

A. DATASET

We adopted two multi-modal datasets:

1) WEIBO

The dataset that was provided in [26]. It was collected from May 2012 to January 2016, Xinhua, China's reputable news agency, collected the dataset from the Weibo's official dispelling system and true tweets from verified tweets. The balanced dataset statistics are shown in Table 2. Xinhua News Agency verifies the dataset's posts to determine whether they are fake or legitimate news.

2) FAKEDDIT

The novel multi-modal dataset [27]. The dataset comprises more than 1 million examples of fake news across several categories, collected through multiple modes of data collection. Following a series of reviews, the samples are categorized into 2-way, 3-way, and 6-way classification groups using distant supervision and then labeled accordingly, as shown in Figure 9. These hybrid models combine text and image data and conduct thorough experiments to explore various categorization variations. Our research highlights the significance of the innovative concept of multimodality and fine-grained classification, which is specific to Fakeddit, to evaluate our architecture for fake news detection. The problem of imbalanced class distribution in real-world datasets severely impairs the performance of classification algorithms. The learning task becomes more complicated and challenging when there is also a class overlap problem with imbalanced

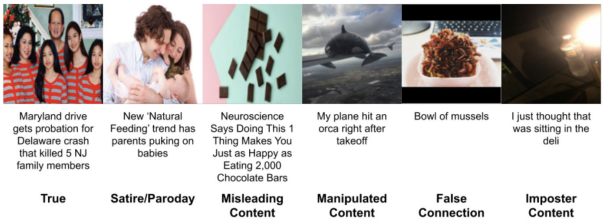

FIGURE 9. Fakeddit's categorizes.

data. The suggested method in [27] tries to find the best subset of the majority samples to deal with both the imbalanced and the class-overlap issues at the same time while avoiding getting rid of too many majority samples, especially in areas where classes overlap.

Accordingly, we selected 54,000 submissions from 878,000 entries randomly, with 30000 as the training dataset, 15000 as real news, and 15000 as fake news, while the test set had 10800 submissions, of which 5400 were classified as real news and 5400 as fake news. We adopted random selection to ensure fairness, minimize bias, maintain data integrity by avoiding the loss of valuable data points and excluding redundant outliers, and foster transparency and credibility in the methodology of the study.

B. EXPERIMENTAL SETTING

The framework was built on Python 3.9. The train-to-test split ratio is 8:2. For textual content in multi-modal posts, we use the pre-trained BERT module with CNN and GRU simultaneously for the textual branch. For simplicity, we fix the weights of BERT during the training phase. The ResNet-50 model, combined with the CBAM attention mechanism, was used for the visual feature extraction module. The visual features were also shrunk from 2048 dimensions to 64 dimensions. To begin training on the dataset, we set the learning rate to 0.001 for 200 epochs and the batch size to 64, which are settings commonly used in studies.

C. ABLATION EXPERIMENT

To determine how crucial each component was to the experiment, the model's simplified version had some parts removed. These parts were the visual feature extractor and the auto-encoder module.

Model w/o image: removing the visual features and using only textual features as input to the model. At this point, the classification of the encoded content is only based on textual features.

Ours w/o auto-encoder: The auto-encoder is removed, and the text and visual features are simply combined as the classification basis. Table 3 and Table 4 list the results of the ablation experiment, and two conclusions can be drawn: First, each component of the model plays an important role in the fake news detection task because the classification accuracy will decrease to some extent if any part of the model

TABLE 3. The ablation comparison results of Weibo.

Ablation model	Acc	Re	Pr	F_1
Model w/o image	87%	86%	87%	86%
Model w/o Auto-encoder	86%	84%	88%	86%
Overall model	88%	88%	87%	87%

TABLE 4. The ablation comparison results of Fakeddit.

Ablation model	Acc	Re	Pr	F_1
Model w/o image	95%	92%	93%	92%
Model w/o Auto-encoder	92%	88%	90%	89%
Overall model	98%	95%	97%	96%

TABLE 5. Comparison of the approaches results on Weibo.

Methods	Acc	Re	Pr	F_1
att-RNN	78%	86%	68%	76%
MAVE	82%	85%	76%	80%
EANN	82%	84%	81%	82%
MMCN	87%	88%	87%	87%
Proposed method	88%	88%	87%	87%

is removed. Second, the model is most accurate when the auto-encoder is added. The experimental results show that the proposed model is superior to the single-mode model in fake news detection.

D. BASELINES

In order to verify the validity of the proposed model, it was compared with other multi-modal fake news detection models on the same data set. The main comparison models are as follows:

MMCN [11]: The multi-level multi-modal cross-attention network that utilizes a network for cross-attention between different modalities.

att-RNN [26]: Uses attention mechanisms to combine textual, visual, and social context features. In this model, the LSTM network is used to jointly represent text and social environment information, and then the attention mechanism is used to integrate visual features.

MAVE [8]: By training the multi-mode variational auto-encoder and reconstructing two modes from the learned shared representation to find the correlation between the modes, a better multi-mode shared representation can be obtained for fake news detection.

EANN [31]: CNN and VGG19 were used to extract multi-modal features, and then the multi-modal features obtained by the concatenation of the two were input into the fake news detector and event discriminator to identify the labels of each post.

Bi-GCN [32]: Bi-directional graph convolutional network is one of the state-of-the-art fake news detection methods, utilizing both the content of the post and the propagation path.

SeRN [33]: Stance Extraction and Reasoning Network is proposed to extract the stances implied in post-reply pairs implicitly and integrate the stance representations for fake news detection.

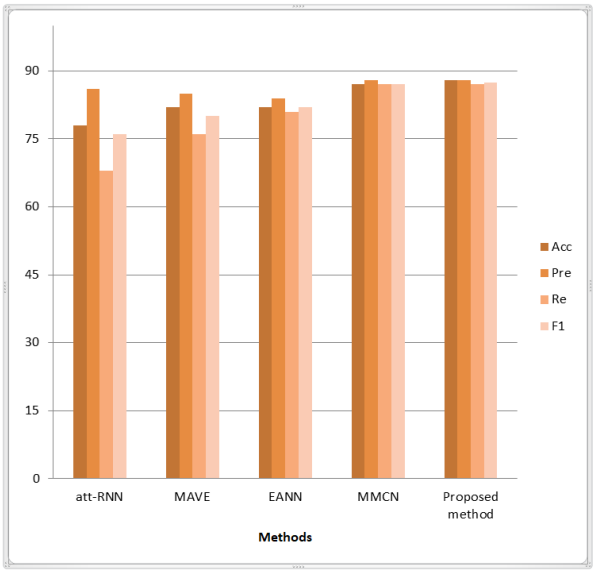

FIGURE 10. Performance of the different methods on Weibo .

TABLE 6. Comparison of the approaches results on Fakeddit.

Methods	Acc	Re	Pr	F_1
Bi-GCN	95%	93%	94%	93%
SeRN	96%	95%	96%	95%
Proposed method	98%	95%	97%	96%

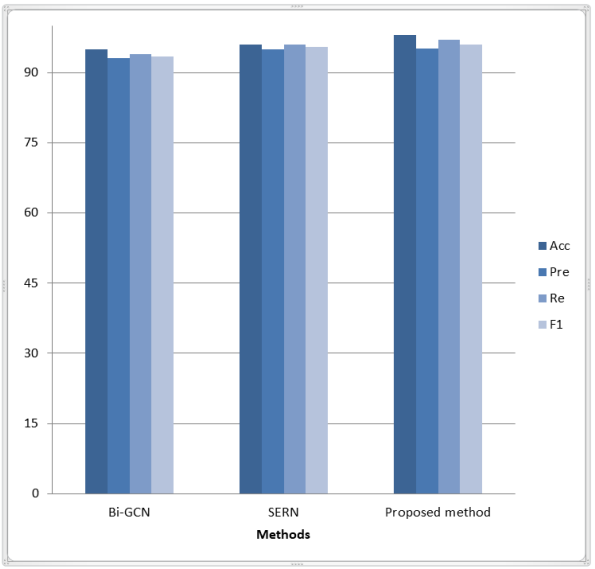

FIGURE 11. Performance of the different methods on Fakeddit.

VI. RESULTS AND DISCUSSION

Table 5 and Figure 10 show that the outcomes were close to those of the other baseline approaches in terms of accuracy on Weibo. While Table 6 and Figure 11 show our approach outperforms all the baselines on Fakeddit in terms of accuracy and F1 score, this demonstrates that deep neural networks are capable of learning better whenever the database increases,

enhancing the accuracy of the outcomes. In addition to the role of reducing the dimensionality of fusion features via the auto-encoder, managing the amount of these features in turn contributes to enhancing detection accuracy. Furthermore, the success of our approach underscores the importance of continually refining and optimizing feature extraction and dimensionality reduction methods. By efficiently managing the quantity and quality of features, we can further enhance the accuracy and robustness of deception detection models, ensuring their effectiveness across diverse datasets and real-world applications.

VII. CONCLUSION AND FUTURE WORK

This research paper proposes a fake news detection model based on multimodal deep learning to solve the problem of fake news detection in complex scenes where text and image coexist. The overall model consists of four parts, namely a feature extraction module, a feature fusion module, a dimensionality reduction stage, and a classification stage. The feature extraction module includes a text feature extractor and a visual feature extractor. The features of text and image are fused and then encoded via an auto-encoder, followed by classifying the encoded features by FLN. A large number of experiments with data collected on Weibo and Fakeddit demonstrate the effectiveness of the model proposed in this paper. Moving forward, there are several areas to explore and enhance in our fake news detection model. First off, we need to broaden our dataset, including a wider range of sources and social media platforms. This will make the model more adaptable to different scenarios. We also need to make the model more dynamic, allowing it to quickly adapt to new fake news tactics as they emerge. This means incorporating continuous learning mechanisms to keep the model updated in real-time. Optimizing for real-time processing is essential for integrating the model into live social media feeds and news streams. Considering the global nature of online content, it's crucial to extend the model's capabilities to detect fake news in multiple languages. This will make it more useful and applicable in diverse linguistic settings. Additionally, we should look into fine-tuning options, making the model easily customizable for specific domains or user preferences. This could involve creating user-friendly interfaces or providing settings for users to tailor the model to their needs. Lastly, user feedback is invaluable. We should create mechanisms for users to report potential fake news and integrate feedback loops to refine the model continuously. This collaborative effort between the model and its users ensures ongoing improvement.

REFERENCES

[1] G. Pennycook and D. G. Rand, "Examining false beliefs about voter fraud in the wake of the 2020 presidential election," *Harvard Kennedy School Misinformation Rev.*, vol. 2, no. 1, pp. 1–19, Jan. 2021, doi: 10.37016/mr-2020-51.

[2] H. Ahmed, I. Traore, and S. Saad, "Detection of online fake news using n-gram analysis and machine learning techniques," in *Proc. 1st Int. Conf.*, 2017, pp. 127–138.

[3] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explor. Newslett.*, vol. 19, no. 1, pp. 22–36, 2017.

[4] M. F. Mridha, A. J. Keya, M. A. Hamid, M. M. Monowar, and M. S. Rahman, "A comprehensive review on fake news detection with deep learning," *IEEE Access*, vol. 9, pp. 156151–156170, 2021.

[5] R. Mishra, "Fake news detection using higher-order user to user mutual-attention progression in propagation paths," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 2775–2783.

[6] M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. S. Choi, and B.-W. On, "Fake news stance detection using deep learning architecture (CNN-LSTM)," *IEEE Access*, vol. 8, pp. 156695–156706, 2020.

[7] X. Zhou, J. Wu, and R. Zafarani, "Similarity-aware multi-modal fake news detection," in *Proc. Pacific–Asia Conf. Knowl. Discovery Data Mining*, 2020, pp. 354–367.

[8] D. Khattar, J. S. Goud, M. Gupta, and V. Varma, "MVAE: Multimodal variational autoencoder for fake news detection," in *Proc. World Wide Web Conf.*, May 2019, pp. 2915–2921.

[9] N. Seddari, A. Derhab, M. Belaoued, W. Halboob, J. Al-Muhtadi, and A. Bouras, "A hybrid linguistic and knowledge-based analysis approach for fake news detection on social media," *IEEE Access*, vol. 10, pp. 62097–62109, 2022.

[10] L. Ying, H. Yu, J. Wang, Y. Ji, and S. Qian, "Fake news detection via multi-modal topic memory network," *IEEE Access*, vol. 9, pp. 132818–132829, 2021.

[11] L. Ying, H. Yu, J. Wang, Y. Ji, and S. Qian, "Multi-level multi-modal cross-attention network for fake news detection," *IEEE Access*, vol. 9, pp. 132363–132373, 2021.

[12] A. Hashim Jawad Almarashy, M.-R. Feizi-Derakhshi, and P. Salehpour, "Enhancing fake news detection by multi-feature classification," *IEEE Access*, vol. 11, pp. 139601–139613, 2023.

[13] E. Zafarani-Moattar, M. R. Kangavari, and A. M. Rahmani, "Topic detection on COVID-19 tweets: A comparative study on clustering and transfer learning models," *Tabriz J. Elect. Eng.*, vol. 52, pp. 281–291, Oct. 2022.

[14] Z. Xinxi, "Single task fine-tune BERT for text classification," in *Proc. 2nd Int. Conf. Comput. Vis., Image, Deep Learn.*, Oct. 2021, pp. 1–12.

[15] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.

[16] H. Alaa Al-Kabbi, M.-R. Feizi-Derakhshi, and S. Pashazadeh, "Multi-type feature extraction and early fusion framework for SMS spam detection," *IEEE Access*, vol. 11, pp. 123756–123765, 2023.

[17] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," 2014, *arXiv:1406.1078*.

[18] I. Khalaf Salman Al-Tameemi, M.-R. Feizi-Derakhshi, S. Pashazadeh, and M. Asadpour, "Interpretable multimodal sentiment classification using deep multi-view attentive network of image and text data," *IEEE Access*, vol. 11, pp. 91060–91081, 2023, doi: 10.1109/ACCESS.2023.3307716.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[20] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.

[21] P. Jindal and D. Kumar, "A review on dimensionality reduction techniques," *Int. J. Comput. Appl.*, vol. 173, no. 2, pp. 42–46, Sep. 2017.

[22] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–11.

[23] G. Li, P. Niu, X. Duan, and X. Zhang, "Fast learning network: A novel artificial neural network with a fast learning speed," *Neural Comput. Appl.*, vol. 24, nos. 7–8, pp. 1683–1695, Jun. 2014.

[24] Z. Jahanbakhsh-Nagadeh, M.-R. Feizi-Derakhshi, and A. Sharifi, "A deep content-based model for Persian rumor verification," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 21, no. 1, pp. 1–29, Jan. 2022.

[25] D. Wang, W. Zhang, W. Wu, and X. Guo, "Soft-label for multi-domain fake news detection," *IEEE Access*, vol. 11, pp. 98596–98606, 2023.

[26] Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo, "Multimodal fusion with recurrent neural networks for rumor detection on microblogs," in *Proc. 25th ACM Int. Conf. Multimedia*, Oct. 2017, pp. 1–17.

[27] K. Nakamura, S. Levy, and W. Y. Wang. (2019). *R/fakeddit: A New Multimodal Benchmark Dataset for Fine-Grained Fake News Detection*. [Online]. Available: https://fakeddit.netlify.app/
[28] P. Soltanzadeh, M. R. Feizi-Derakhshi, and M. Hashemzadeh, "Addressing the class-imbalance and class-overlap problems by a metaheuristic-based under-sampling approach," *Pattern Recognition*, vol. 143, Nov. 2023, Art. no. 109721.
[29] (2021). *Image from Google Images*. [Online]. Available: https://images.app.goo.gl/KRRLJ3XvwbXMwfFZ6
[30] K. Mbappé. (2019). Tweet by Kylian Mbappé. Twitter. [Online]. Available: https://twitter.com/KMbappe/status/1188603901953691649
[31] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, L. Su, and J. Gao, "EANN: Event adversarial neural networks for multi-modal fake news detection," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2018, pp. 1–16.
[32] T. Bian, X. Xiao, T. Xu, P. Zhao, W. Huang, Y. Rong, and J. Huang, "Rumor detection on social media with bi-directional graph convolutional networks," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, 2020, pp. 549–556.
[33] J. Xie, S. Liu, R. Liu, Y. Zhang, and Y. Zhu, "SERN: Stance extraction and reasoning network for fake news detection," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 2520–2524.

AHMED HASHIM JAWAD ALMARASHY received the B.S. degree in control and systems engineering from the University of Technology, Baghdad, Iraq, in 2003, and the M.Tech. degree in computer science engineering from Acharya Nagarjuna University, India, in 2016. He is currently pursuing the Ph.D. degree with the Department of Computer Engineering, University of Tabriz, Iran. His research interests include natural language processing, deep learning, image processing, and smart cities.

MOHAMMAD-REZA FEIZI-DERAKHSHI received the B.S. degree in software engineering from the University of Isfahan, Iran, and the M.Sc. and Ph.D. degrees in artificial intelligence from Iran University of Science and Technology, Tehran, Iran. He is currently a Professor with the Faculty of Computer engineering, University of Tabriz, Iran. His research interests include natural language processing, optimization algorithms, deep learning, social network analysis, and intelligent databases.

PEDRAM SALEHPOUR (Member, IEEE) received the B.S. and M.Sc. degrees in computer science and the Ph.D. degree in electrical engineering from the University of Tabriz, in 2007, 2009, and 2015, respectively. He is currently an Assistant Professor with the Faculty of Electrical and Computer engineering, University of Tabriz. His research interests include distributed computing, machine learning, image processing, and deep learning.