

Received 31 December 2023, accepted 20 January 2024, date of publication 23 January 2024, date of current version 1 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3357661

RESEARCH ARTICLE

Utilizing Ensemble Learning for Detecting Multi-Modal Fake News

MUHAMMAD LUQMAN¹, MUHAMMAD FAHEEM², (Member, IEEE),
WAHEED YOUSUF RAMAY³, MALIK KHIZAR SAEED⁴, AND MAJID BASHIR AHMAD⁵

¹Department of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China

²School of Technology and Innovations, University of Vaasa, 65200 Vaasa, Finland

³Department of Computer Science, Air University, Multan 60000, Pakistan

⁴Department of Computer Sciences, COMSATS University Islamabad, Vehari 61000, Pakistan

⁵School of Software and Microelectronics, Northwestern Polytechnical University, Xi'an 710072, China

Corresponding author: Muhammad Faheem (muhammad.faheem@uwasa.fi)

ABSTRACT The spread of fake news has become a critical problem in recent years due extensive use of social media platforms. False stories can go viral quickly, reaching millions of people before they can be mocked, i.e., a false story claiming that a celebrity has died when he/she is still alive. Therefore, detecting fake news is essential for maintaining the integrity of information and controlling misinformation, social and political polarization, media ethics, and security threats. From this perspective, we propose an ensemble learning-based detection of multi-modal fake news. First, it exploits a publicly available dataset *Fakeddit* consisting of over 1 million samples of fake news. Next, it leverages Natural Language Processing (NLP) techniques for preprocessing textual information of news. Then, it gauges the sentiment from the text of each news. After that, it generates embeddings for text and images of the corresponding news by leveraging Visual Bidirectional Encoder Representations from Transformers (V-BERT), respectively. Finally, it passes the embeddings to the deep learning ensemble model for training and testing. The 10-fold evaluation technique is used to check the performance of the proposed approach. The evaluation results are significant and outperform the state-of-the-art approaches with the performance improvement of 12.57%, 9.70%, 18.15%, 12.58%, 0.10, and 3.07 in accuracy, precision, recall, F1-score, Matthews Correlation Coefficient (MCC), and Odds Ratio (OR), respectively.

INDEX TERMS Ensemble learning, convolutional neural network, multi-modal fake news, classification, boosted CNN, bagged CNN.

I. INTRODUCTION

The concept of fake news is not new. Its roots existed long ago in our society. It refers to false information which can be disseminated to mislead or deceive the Public. For example, fake news about COVID-19 vaccines could discourage people from getting vaccinated, leading to increased rates of illness and death. In the past, every kind of distinct material was considered fake news, like satires, conspiracies, news manipulation, and click-bait. However, fake news is now becoming jargon [1] and has a huge impact on the critical

The associate editor coordinating the review of this manuscript and approving it for publication was Donato Impedovo¹.

events happening in our society, e.g., spreading fake news (false stories) on social media was very concerning in US presidential election 2016 [2].

Fake news can spread quickly through social media and other online platforms. It can have serious consequences, such as causing panic, influencing elections, and eroding public trust in legitimate news sources. Individuals need to distinguish real news and critically evaluate sources of information before sharing or responding to them. Additionally, news organizations and social media platforms are responsible for combating the spread of fake news by fact-checking and removing false content. The surveys show that about 70% of Americans use social media as a source

of news and circulating information [3]. The accessibility of news and information on the Internet is very low-cost and convenient. However, spreading fake news on these carriers is straightforward and effortless [4]. Fake news can lead to false assumptions that drastically affect our society. Consequently, it is critical to design an automated fake news detection system.

Many researchers are actively developing new and better methods for identifying and combating the spread of misinformation. Some of the key research areas and trends in this field include deep learning approaches, e.g., Convolutional Neural Network (CNN); linguistic features, e.g., sentiment analysis, topic modeling, and stylometric analysis; source-based approaches, e.g., analyzing the domain name, social media presence, or history of the news source, and ensemble approaches, e.g., combining linguistic, source-based, and deep learning models to create a more robust and accurate detection system. Although recent research has identified the issues of the said problem and proposed different solutions, e.g., pre-trained language models have shown their effectiveness in alleviating feature engineering efforts, such as Bidirectional Encoder Representations from Transformers (BERT) [5], OpenAI GPT [6], and Elmo [7], however; the problem requires significant performance improvement.

From this perspective, this paper proposes an ensemble learning-based detection of multi-modal fake news (ELD-FN). It first exploits a publicly available dataset Fakeddit, a novel multi-modal dataset consisting of over 1 million samples from multiple categories of fake news. Second, it leverages Natural Language Processing (NLP) techniques for preprocessing textual information of news. Third, it gauges the sentiment from the text of each news. Fourth, it generates embeddings for text and images of the corresponding news by leveraging V-BERT [8], respectively. Finally, it passes the embeddings to the deep learning ensemble model for training and testing. The 10-fold evaluation technique is used to check the performance of ELD-FN. The evaluation results are significant and outperform the state-of-the-art approaches with the performance improvement of 12.57%, 9.70%, 18.15%, 12.58%, 0.10, and 3.07 in accuracy, precision, recall, F1-score, Matthews Correlation Coefficient (MCC), and Odds Ratio (OR), respectively.

The main contributions made in this paper are as follows.

- The proposed approach integrates news sentiment as a crucial feature and employs ensemble learning to identify multi-modal fake news.
- It is evident from the evaluation results that ELD-FN is significant and outperforms the baseline approaches with the performance improvement of 12.57%, 9.70%, 18.15%, 12.58%, 0.10, and 3.07 in accuracy, precision, recall, F1-score, MCC, and OR, respectively.

The organization of the rest of the paper is as follows. Section III describes the details of ELD-FN. Section IV describes the evaluation methods for ELD-FN, obtained results, and their threats to validity. Section II discusses the

research background. Section V summarizes the paper and suggests future work.

II. RELATED WORK

Although extensive research on fake news detection has been performed [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], most research is conducted on textual data or uni-modal features. However, two most relevant researches [24], [25] proposed deep learning-based solutions for detecting fake news. The proposed approach (ELD-FN) differs from baseline approaches as it does not work for the multi-modal features but also considers the sentiments involved in the textual information of news.

Most of the state-of-the-art fake news classification approaches can be categorized as follows: 1) fake news classification approaches for single-modality and 2) fake news classification approaches for multi-modality.

A. FAKE NEWS CLASSIFICATION APPROACHES FOR SINGLE-MODALITY

The fake news classification approaches for single-modality can be further divided into two categories based on the text/image features.

1) SINGLE-MODALITY BASED CLASSIFICATION APPROACHES USING TEXTUAL FEATURES

Textual features can be divided into generic and latent categories. Usually, traditional machine learning algorithms utilize Generic textual features. These algorithms analyze text based on linguistic levels such as lexicon, syntax, discourse, and semantics. Previous research has compiled a detailed table summarizing these features [10]. However, Latent textual features consist of the embeddings extracted from textual data of news at the word, sentence, or document level. Latent vectors are constructed from the textual news data. Furthermore, these latent vectors are used as input for classifiers, i.e., SVM.

Recurrent neural networks (RNNs) are potent in modeling and analyzing sequential data. For example, Ma et al. used RNNs to capture relevant information over time by learning hidden layer representations [11]. Meanwhile, Chen et al. proposed a CNN-based approach for the classification [12]. Moreover, a novel technique Attention-Residual Network (ARC) is introduced to acquire long-range features. Ma et al. introduced a Generative Adversarial Network (GAN)-based model that employs a Generator network based on Gated Recurrent Units (GRU) to generate contentious instances. Furthermore, a Discriminator network based on RNNs is designed to identify essential features [13].

RNN-based models have proven very effective in classifying fake news detection datasets. However, the RNN-based models prioritize the recent input sequence, and the essential features may be located at the end of the sequence. Yu et al. proposed a CNN-based approach that resolves this issue. The proposed technique does not prioritize recent input sequences. This approach applies feature extraction based

on the relationship of the essential features [14]. Vaibhav and Hovy utilize a graphical approach for classifying news articles [15]. For this purpose, they used Graph Neural Networks, such as Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT), to create graph embeddings for fake news detection.

Wu et al. utilize multi-task learning techniques to classify and detect fake news. Moreover, the stance classification task optimizes shared layers concurrently, improving news representations [16]. Cheng et al. utilized LSTM model to classify the textual news data. They used a variational autoencoder to extract essential textual features at the tweet-level text.

Some researchers have assumed that complex and multi-dimensional news are not accessible initially. The accessibility of only text-based news depends on the popularity [17]. Qian et al. developed a text-based model that utilizes word/sentence level data from legitimate papers to produce user feedback for early detection [18]. This addressed the scarcity of user reviews as an auxiliary source of information. For example, Qian et al. proposed an approach for generating user feedback on the text. Such feedback was along with word/sentence level information from real articles for the classification process [18]. Giachanou et al. investigated the influence of emotional cues in the proposed model. They propose an LSTM model that integrates emotional signals extracted from claim texts to differentiate between true and false news [19].

2) SINGLE-MODALITY BASED CLASSIFICATION APPROACHES USING IMAGE FEATURES

As multimedia becomes more prevalent in social networks, news now contains text and visual information such as images and videos that convey rich meaning. However, textual feature-based approaches face challenges in effectively capturing visual information because of the heterogeneity between text and image data. Consequently, many researchers have proposed image-based approaches for detecting fake news.

Classical image-based models utilized basic fundamental numerical features of images [20], [26], such as image count, popularity [27], and type to identify fake news. For impaired images, complex forensics features were extracted. Furthermore, post and user-based features are integrated to identify fake news [28]. However, it was evident that basic numerical features are inadequate to describe complex visual information of the news images.

Deep learning models such as CNNs have proven effective in capturing visual features in news images. Many researches have shown that feature extraction from CNN models can be used in visual recognition tasks to generate generic image representation [29].

Building on the success of CNNs, recent studies have utilized pre-trained deep CNNs like VGG19 [30], [31] to obtain generic visual representations [32], [33]. Researchers

suggested multi-domain visual neural models to capture the inherent traits of fabricated news images more effectively. These multi-domain models merged frequency and pixel domain visual data to differentiate between genuine and fabricated news based on visual characteristics [34]. Poor quality is a common trait in fake news images. The poor quality feature and image semantics are visible in frequency and pixel domains. However, the quality feature is extracted by CNN model, and the semantics of the images are extracted by CNN-RNN model.

B. FAKE NEWS CLASSIFICATION APPROACHES FOR MULTI-MODALITY

Word-based and Image-based information are both important in detecting fake news. As social networks often contain both types of information, combining them can improve performance. This section discusses the different multi-modal approaches for fake news detection, categorized based on the different perspectives they adopt.

1) PROBLEMS IN MULTI-MODALITY

Several studies have explored using visual information to complement textual information in detecting fake news. These studies typically use text-based and image-based encoders to extract textual and visual features, respectively. Furthermore, these feature vectors construct an overall feature vector for each news. For example, Wang et al. proposed Event classification as an additional task to enhance the generalizing ability of the model for event-invariant multi-modal features [32]. Other researchers, such as Singhal et al., use a combination of text-based and image-based features. They utilize BERT and XLNet pre-trained models for encoding text-based and image-based data, respectively [35]. However, these approaches are proven to be limited in effectively detecting multi-modal fake news because of their ability to capture complex cross-modal correlations. More advanced multi-modal techniques are needed to improve the performance of fake news detection.

2) FLEXIBILITY IN MULTI-MODALITY

Some studies have recognized that irrelevant images are a common characteristic of multi-modal fake news and have focused on measuring the consistency between the text and visual components in detection. One approach by Zhou and Zafarani [36] used an image captioning model to generate sentences from images and then measured the similarity between those sentences and the original text. However, this approach was constrained by the discrepancies that existed between the training data of the image captioning model and the real news corpus. Another approach by Xue et al. projected the visual and textual features into a shared feature space and computed the similarities between resulting multi-modal features. However, they encountered difficulties capturing multi-modal inconsistencies because of the semantic gap between the two types of features [37].

Ghorbanpour et al. [38] proposed the Fake-News-Revealer (FNR) method, which uses a Vision-transformer [39] and BERT [5] to extract image and text features respectively. The model extracted textual and visual features separately and determined their similarities by loss.

3) IMPROVEMENT IN MULTI-MODALITY

Several researchers have proposed different approaches for fake news detection using multi-modal data. Jin et al. utilized an RNN model and applied an attention mechanism to combine information extracted from textual, visual, and social context data [40]. Zhang et al. [41] used a multi-channel CNN with an attention mechanism to combine multi-modal information, while Song et al. [42] proposed the co-attention transformer to model the bidirectional enhancement between images and text. Qian et al. developed a Hierarchical Multi-modal Contextual Attention Network (HMCAN), which was designed to collectively capture multi-modal context data and the hierarchical semantics of text [43]. Wu et al. introduced the Multi-modal Co-Attention Network (MCAN) that extracts spatial-domain and frequency-domain features from the image and text, and fuses visual and textual features using multiple co-attention layers [44]. Other researchers have also utilized Graph Convolutional Networks (GCN) and entity-centric cross-modal interaction to model the relationship between word-based and image-based objects. Finally, Zhang et al. and Laura et al. proposed a BERT-based multi-modal model to encode text-based and image-based information. The model effectively captures the interplay between text and images and employs contrastive learning to enhance multi-modal representations. [24], [45] integrated visual entities to enhance the comprehension of high-level semantics in news images and to model the inconsistencies and mutual enhancements of multi-modal entities [22].

In summary, when performing multi-modal fake news detection, there are three important inductive biases to consider when examining text-image correlations. Firstly, images provide additional information to the text, highlighting the need for multi-modal. Secondly, problems between text and images can serve as a potential signal for detecting fake news using multiple modalities. Finally, text-based and image-based data can improve performance by identifying essential features.

III. METHODOLOGY

A. OVERVIEW

The overview of ELD-FN is depicted in Fig. 1. The following are the main steps of ELD-FN.

- 1) First, the publicly available multi-modal dataset (Fakeddit) is collected from Google Drive.¹
- 2) Next, it leverages NLP techniques, e.g., tokenization, stop-word removal, lowercase conversion, and lemmatization, for preprocessing textual information of news.

- 3) Then, it computes the sentiment from the text of each news.
- 4) After that, it generates embeddings for text and images of the corresponding news by leveraging V-BERT, respectively.
- 5) Finally, it passes the embeddings to the deep learning ensemble model for training and testing.

B. PROBLEM DEFINITION

A news n from a set of multi-modal dataset of news \mathbb{N} can be represented as follows:

$$n = \langle t, i, s \rangle \quad (1)$$

where, t is the textual information of n , i is the image of n , and s is an assigned status to n whether n is fake or true.

The ELD-FN suggests the status of new news as either *true* or *false*, where *true* represents that the news is real and *false* represents that the news is *fake*. Consequently, the automatic classification of a new news n could be defined a mapping f :

$$f : n \rightarrow c \\ c \in \{true, false\}, \quad n \in \mathbb{N} \quad (2)$$

where, c is a suggested status from a news status set (*true*, *false*).

C. PREPROCESSING

The news may contain inappropriate and unnecessary text, e.g., English stop-words. Such information is considered an overhead for the machine learning classification algorithms because of processing time and memory utilization. Therefore, preprocessing of news text is essential for the performance of ELD-FN to make it fast and memory efficient. We perform the following preprocessing steps to clean the text of news.

1) TOKENIZATION

Text tokenization breaks down a piece of text into smaller units called tokens. Tokens are individual words, phrases, or other meaningful text elements, which can be analyzed and processed further.

2) SPECIAL CHARACTER REMOVAL

The text of news may contain special characters, e.g., semicolon (;). This step removes the special characters from the list of tokens.

3) STOP-WORD REMOVAL

English text contains meaningless words that are used to make sentences meaningful, called stop-words. This step removes stop-words from the working list.

4) SPELL CORRECTION AND LOWERCASE CONVERSION

This step identifies and corrects the spelling mistakes from the working list of tokens of news.

¹<https://fakeddit.netlify.app/>, accessed on 15-01-2023.

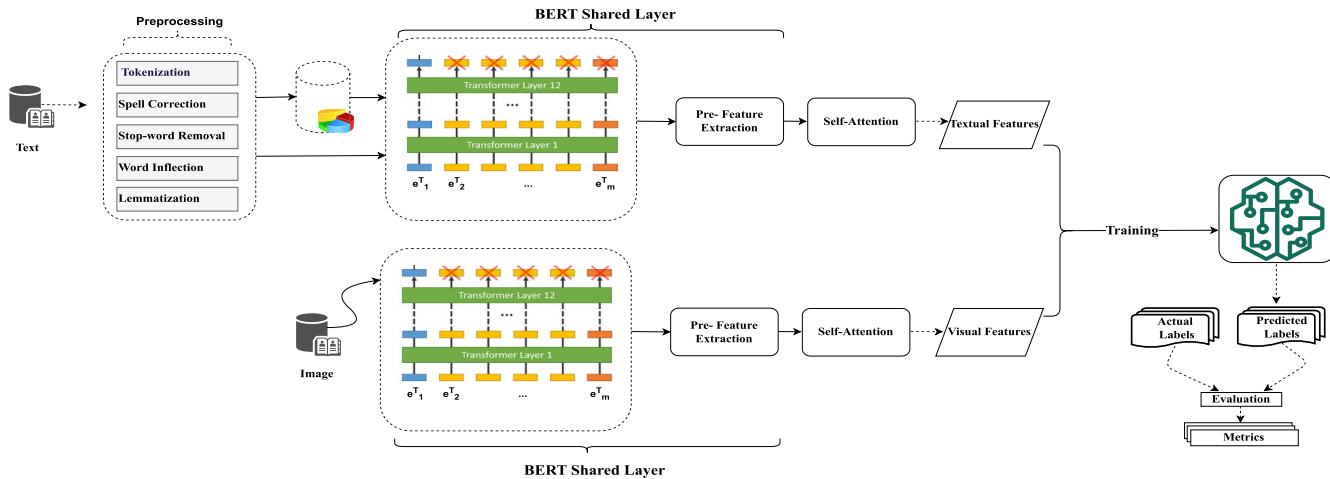


FIGURE 1. The overview of ELD-FN.

5) LEMMATIZATION

The lemmatization step converts higher-degree and comparative words into their lower-degree words, e.g., lemmatization converts the word *darker* into *dark*.

We exploit Python Natural Toolkit (NLTK)² for the preprocessing of news. The preprocessed news can be represented as follows:

$$n' = \langle t', i, s \rangle \tag{3}$$

$$t' = \langle t_1, t_2, \dots, t_n \rangle \tag{4}$$

where, $t' = t_1, t_2, \dots, t_n$ are the tokens from the text of n after preprocessing.

D. SENTIMENT ANALYSIS

Sentiment analysis is a NLP technique that involves identifying and extracting subjective information from text, i.e., opinions, attitudes, emotions, and sentiments towards a particular topic. It automatically classifies the polarity of a text as positive, negative, or neutral. We exploit TextBlob API³ for the computation of sentiment analysis. The news (mentioned in Eq. e3) after sentiment computation can be represented as follows:

$$n' = \langle v, t', i, s \rangle \tag{5}$$

where, v is the sentiment of n' .

E. FEATURE MODELING

This step passes the preprocessed text and images from the multi-modal dataset to V-BERT to generate the embeddings. V-BERT is an extension of the BERT model that combines the power of the BERT model with a visual grounding mechanism, allowing it to understand the relationship between the text and the visual information in an image. This is achieved

by combining a region-based visual feature extractor with the BERT model, where each image region is encoded into a vector using a CNN. These visual features are concatenated with the input text, and the resulting sequence is fed into the BERT model. During training, V-BERT is optimized to minimize a joint loss function. This allows Visual BERT to learn language and vision representations in a unified framework and capture the complex interactions between the two modalities. The layers/steps involved in ELD-FN for identifying fake/real news.

1) BERT SHARED LAYER

For the news text, the BERT shared layer is implemented using a pre-trained Seq2Seq model [8]. The fine-tuning learning process is required and indispensable to achieve better results. To improve its efficiency, separate BERT-shared layers are adopted for model-to-model textual features. The output of news text feature extractor O_{BERT}^T can be represented as follows:

$$O_{BERT}^T = BERT^T(X^T) \tag{6}$$

where, $BERT^T$ is the relevant BERT-shared layer modeling for news text and X^T is the input representation of textual data.

2) IMAGE EMBEDDING LAYER

For the news image, Faster-RCNN model [8] is applied to extract features from the image. The detected objects may provide visual contexts of the whole picture and be linked to specific terms through detailed region details. We also add a position embedding feature to images by encoding the object location. The output of the image feature extractor O_{BERT}^I can be represented as follow:

$$O_{BERT}^I = BERT^I(X^I) \tag{7}$$

²<https://www.nltk.org/>, accessed on 15-01-2013.

³<https://textblob.readthedocs.io/en/dev/>, accessed on 15-01-2023.

where, $BERT^I$ is the relevant BERT-shared layer modeling for images, and X^I is the input representation of images.

3) PRE-FEATURE EXTRACTION

The BERT-shared layer is strong enough for feature extraction. It includes a pre-feature extractor to enhance the ability of BERT to learn semantic characteristics. Pre-feature extractor consists of the Position-wise Convolution Transformation (PCT) and the Multi-Head Self-Attention (MSA) layer.

4) MULTI-MODAL FEATURE CONCATENATION

After extracting the latent features of text and image, these are concatenated together to obtain the desired multi-modal feature representations. The multi-modal concatenated features O^f can be represented as follows:

$$O_{BERT}^f = O_{BERT}^T + O_{BERT}^I \quad (8)$$

F. ENSEMBLE MODEL

Bagging and boosting [46] are two approaches to ensemble machine learning models. We applied both approaches with CNN and LSTM models. Four different architectures (bagged CNN, bagged LSTM, boosted CNN, boosted LSTM) of ensemble machine learning models have experimented using bagging and bootstrap aggregating to predict the fake/real news. Note that bagged CNN is the proposed ensemble model as it yields the other mentioned ensemble architectures. The predictions through different architectures are made using Algorithm 1.

IV. EVALUATION

This section constructs the research questions to evaluate ELD-FN, explains the exploited dataset, defines the metrics and evaluation process, and reports the findings and threats to validity.

A. RESEARCH QUESTIONS (RQs)

The following research questions are investigated to evaluate ELD-FN.

- RQ1: Does ELD-FN outperform the baseline approaches?
- RQ2: Does news sentiment influence the identification of fake news?
- RQ3: Does preprocessing influence the identification of fake news?
- RQ4: Does ELD-FN outperform other classifiers regarding identifying fake news?

The RQ1 compares the ELD-FN with the baseline approaches [24], [25] names as *FakeNED* and *MultiFND* in the rest of this paper. The reason to select these approaches as baseline approaches is that both are recently proposed approaches, closely related to our work and exploited the same dataset.

The RQ2 investigates the influence of news sentiment to detect fake news. It evaluates whether positive news will likely be considered true or vice versa.

Algorithm 1 Ensemble Model

```

1: procedure Ensemble Model
2:   Input:  $X_{iT+1}$ ,  $\alpha_{\frac{\sigma}{b}}$ ,
       $\int_h^{\frac{\sigma}{b}}(\sigma, W_h^{\frac{\sigma}{b}}, b_h^{\frac{\sigma}{b}}) \cdots \int_{N_{\frac{\sigma}{b}}}^{\frac{\sigma}{b}}(\sigma, W_{N_{\frac{\sigma}{b}}}^{\frac{\sigma}{b}}, b_{N_{\frac{\sigma}{b}}}^{\frac{\sigma}{b}})$ 
3:   Initialize:  $\hat{y}_{i+1}, h \leftarrow 2$ 
4:    $X \xrightarrow{f_1^i(\sigma, W_1^i, b_1^i)} y$ 
5:   while  $h \leq N_{\frac{\sigma}{b}}$  do
6:      $\hat{y}_{iT+1} \leftarrow \hat{y}_{iT+1} + \alpha_{\frac{\sigma}{b}} \int_h^{\frac{\sigma}{b}}(\sigma, W_b^{\frac{\sigma}{b}}, b_h^{\frac{\sigma}{b}}, X_{iT+1})$ 
7:      $h = h + 1$ 
8:   end while
9:   Output:  $\hat{y}_{iT+1}$ 
10: end procedure

```

where, X_{iT+1} is the feature set at time instances, $\alpha_{\frac{\sigma}{b}}$ are Weights of bagging or boosting, $\int_h^{\frac{\sigma}{b}}(\sigma, W_h^{\frac{\sigma}{b}}, b_h^{\frac{\sigma}{b}}) \cdots \int_{N_{\frac{\sigma}{b}}}^{\frac{\sigma}{b}}(\sigma, W_{N_{\frac{\sigma}{b}}}^{\frac{\sigma}{b}}, b_{N_{\frac{\sigma}{b}}}^{\frac{\sigma}{b}})$ are set of ensembled bagged or boosted models, \hat{y}_{iT+1} is the output of the ensembled model, X is the feature set, Y is the instance of the output, α is the activation function, and $W_b^{\frac{\sigma}{b}}$ are Weights of bagging or boosting models.

The RQ3 examines the impact of preprocessing the news text to detect fake news.

The RQ4 investigates the impact of different deep-learning classification algorithms on ELD-FN. We analyze the ELD-FN and other deep learning approaches to evaluate the performance of ELD-FN

B. DATASET

The description of the exploited dataset of fake news *Fakeddit* is presented in Table 1 which is public (available online⁴). Nakamura et al. [47] collected the data from a social news and discussion website *Reddit*. It consists of over 1 million pieces of news (1,063,106) from 22 subreddits. It is classified in three different ways: 2-way, 3-way, and 6-way. The dataset samples with 6-way classification are represented in Fig. 2. Out of the total samples, 59.12% (628,501) and 40.48% (527,049) are fake and real news, correspondingly. However, only 64.25% (682,966) samples are multi-modal. Note that we only use the multi-modal data samples with 2-way classification to evaluate the proposed approach. Moreover, Fig. 3 and Fig. 4 represent the wordcloud (most common words in the dataset) and frequency of the words, respectively.

C. PROCESS

This section explains the evaluation process of ELD-FN. After performing the preprocessing and feature modeling as mentioned in Section III, a 10-fold cross-validation technique is applied to train and test ELD-FN. The reason for considering 10-fold cross-validation is that it helps avoid data biasness and reduces the variance in performance estimation

⁴<https://github.com/entitize/fakeddit>, accessed on 15-01-2023.

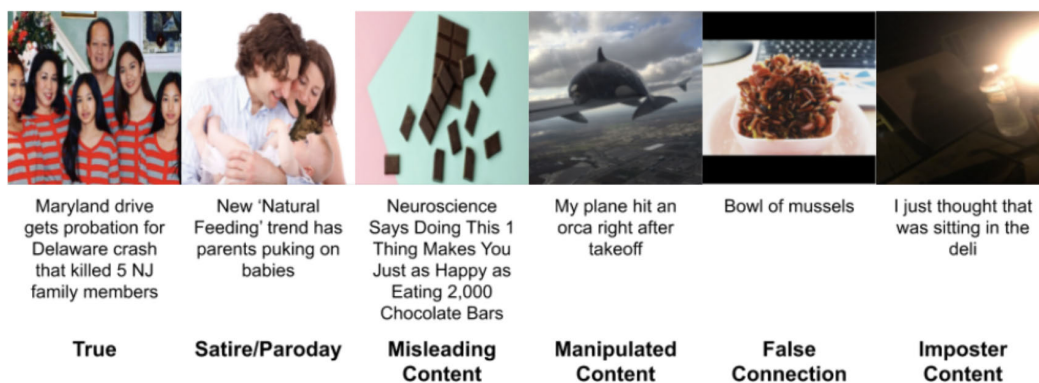


FIGURE 2. Dataset Example with 6-way Classification [47].

TABLE 1. Description of fakeddit dataset.

Total No. of News	1,063,106
Fake News	628,501
Real News	527,049
Multi-modal News	682,966

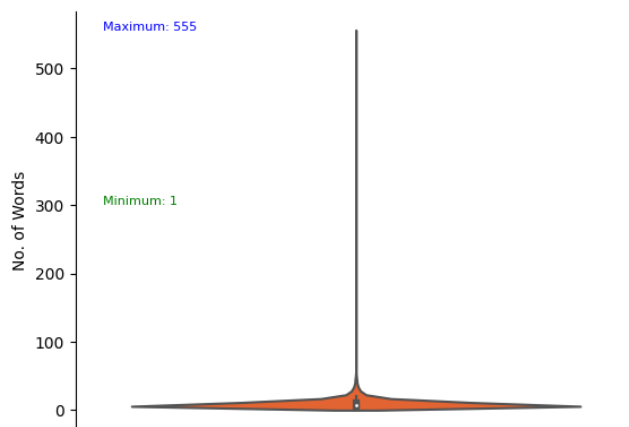


FIGURE 4. Minimum and maximum words.



FIGURE 3. Word cloud - most common words in details.

that might be observed with a single train-test split [48]. The dataset’s total multi-modal news N are broken down into ten (10) slices C_i , where $i = 1, 2, \dots, 10$. For each cross-validation, the slices of N are selected that are not from C_i as training samples (N_t) and news from C_i as testing samples (N_v).

A bit-by-bit evaluation process for i^{th} cross-validation is as follows: 1) All news N_t from N but C_i are extracted and combined; 2) an ensemble deep learning classifier is trained on N_t ; 3) a CNN classifier is trained on N_t ; 4) a LSTM classifier is trained on N_t ; 5) baseline classifiers are trained on N_t ; 6) we predict whether each news from the testing samples C_i is real or fake; and 7) the below-mentioned evaluation metrics are computed for each classifier.

D. METRICS

We train and test the deep learning classifiers to evaluate the performance of ELD-FN. We select the most accepted

metrics (accuracy, precision, recall, and f1-score) for this purpose. Furthermore, we compute the MCC and OR to check the effectiveness of the classifiers. The selected metrics can be presented as follows:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

$$precision = \frac{TP}{TP + FP} \tag{10}$$

$$recall = \frac{TP}{TP + FN} \tag{11}$$

$$f1 - score = \frac{2 * precision * recall}{precision + recall} \tag{12}$$

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{13}$$

$$OR = \frac{TP/FP}{FN/TN} \tag{14}$$

where, TP and TN are the numbers of correctly predicted news as real and fake, respectively. Similarly, FP and FN are the numbers of incorrectly predicted news as real and fake, respectively.

TABLE 2. Performance of ELD-FN and baseline approaches.

Approach	Accuracy	Precision	Recall	F1-score	MCC	OR
ELD-FN	88.83%	93.54%	90.29%	91.89%	0.49	17.02
FakeNED	89.25%	91.12%	87.54%	89.29%	0.45	15.78
MultiFND	78.91%	85.27%	76.42%	80.60%	0.39	13.95

E. RESULTS

1) RQ1: COMPARISON OF ELD-FN AGAINST BASELINE APPROACHES

Table 2 and Fig. 5 present the evaluation metrics for three different approaches (ELD-FN, FakeNED, MultiFND) based on their accuracy, precision, recall, F1-score, MCC, and OR. The results advised that the average values of these metrics for ELD-FN, FakeNED, and MultiFND are (88.83%, 93.54%, 90.29%, 91.89%, 0.49, and 17.02), (89.25%, 91.12%, 87.54%, 89.29%, 0.45, and 15.78), and (78.91%, 85.27%, 76.42%, 80.60%, 0.39, and 13.95), respectively.

The f1-score distribution of cross-validation for ELD-FN, FakeNED, and MultiFND are presented in Fig. 6. A beanplot is a visualization that displays a continuous variable's distribution across different groups. The beanplot compares the f1-score distributions by plotting one bean for each approach. Across a bean, the width of the bean represents the density of the data, with wider beans indicating higher density.

The following observations are made from Table 2, Fig. 5, and Fig. 6.

- ELD-FN has the accuracy (88.83%) and highest precision (93.54%), indicating that it has the highest percentage of correctly classified instances and true positive instances.
- ELD-FN has the highest recall (90.29%) and F1-score (91.89%), indicating that it has the highest ability to correctly identify positive instances and achieve a balance between precision and recall.
- ELD-FN also has the highest MCC (0.49) and OR (17.02), indicating a better correlation between predicted and actual classifications and higher odds of event occurrence than FakeNED and MultiFND. The average results of MCC (0.49 > 0.45 > 0.39) > 0 and OR (17.02 > 15.78 > 13.95) > 1 are true for ELD-FN and confirm its effectiveness.
- The minimum f1-score of ELD-FN is higher than the maximum f1-scores of FakeNED and MultiFND (shown in Fig. 6).

To validate the significant difference in the means of performance (f1-score) for all iterations of ELD-FN, FakeNED, and MultiNED, we perform a single-factor Analysis of Variance (ANOVA). ANOVA is a statistical method used to test whether there is a significant difference in the means of three or more independent groups or samples. It is conducted on Excell with its default settings and presented in Fig. 7. It suggests that $F > F_{crit}$ and $p\text{-value} < (\alpha = 0.05)$ are

true for f1-score, and the factor (using different approaches) significantly differs in f1-score.

Moreover, we utilize two re-sampling methods, over-sampling and under-sampling to tackle the class imbalance within the dataset. Over-sampling involves generating additional samples for the minority class through RandomOverSampler, while under-sampling entails removing surplus records from the majority class in imbalanced datasets using RandomUnderSampler. The findings reveal that employing under-sampling results in accuracy, precision, recall, and F1-score values of 86.12%, 92.54%, 88.76%, and 90.61%, respectively. However, it's important to note that under-sampling diminishes the number of majority class samples, leading to a loss of information. Consequently, the performance of both majority and minority classes in the fine-tuned BERT model declines when under-sampling is applied. Likewise, utilizing the over-sampling technique yields accuracy, precision, recall, and F1-score values of 90.26%, 94.37%, 91.88%, and 93.11%, respectively. This enhancement is attributed to BERT being exposed to a larger dataset, enabling it to learn meaningful patterns more effectively.

The preceding analysis concluded that ELD-FN outperforms the baseline approaches in detecting fake news.

2) RQ2: INFLUENCE OF SENTIMENT ON ELD-FN

The evaluation results of ELD-FN with and without sentiment analysis are presented in Table 3 and Fig. 8. The evaluation results of ELD-FN for different settings of sentiment (enable/disable) based on their accuracy, precision, recall, F1-score, MCC, and OR are (88.83%, 93.54%, 90.29%, 91.89%, 0.49, and 17.02) and (88.12%, 90.38%, 89.98%, 90.17%, 0.49, and 17.02), respectively.

From Table 3 and Fig. 8, it is observed that Disabling sentiment (i.e., textual features only) brings out the significant difference in precision from 93.54% to 90.38% and f1-score from 91.89% to 90.17%. However, MCC and OR remain the same.

Table 5 represents the relationship between sentiment and news. It presents that 65.84% of negative news are real, whereas only 34.16% of the positive news are real. However, 73.71% of negative news are fake, whereas only 26.29% of the positive news are fake. It means the possibility of spreading fake news is $180.37\% = (73.71\% - 26.29\%) / 26.29\%$, if the news is negative. For example, if a fake news article portrays a political figure negatively, it can contribute to a negative sentiment towards that figure among the public and will propagate quickly.

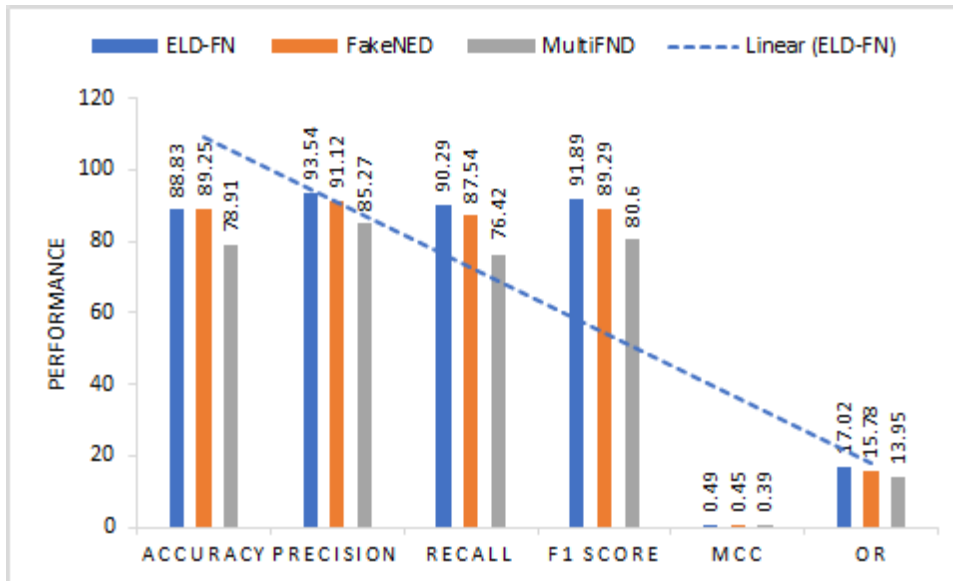


FIGURE 5. Performance of ELD-FN and baseline approaches.

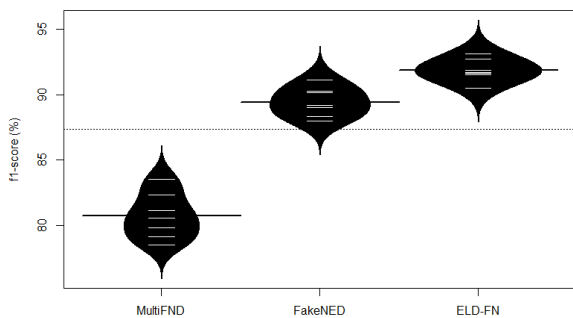


FIGURE 6. Distribution of f-measure.

The preceding analysis concluded that sentiment and features are critical for detecting fake news and disabling either would significantly reduce the performance of ELD-FN.

3) RQ3: INFLUENCE OF PREPROCESSING ON ELD-FN

The evaluation results of ELD-FN with and without preprocessing are presented in Table 4 and Fig. 9. The evaluation results of ELD-FN for different settings of preprocessing (enable/disable) based on their accuracy, precision, recall, F1-score, MCC, and OR are (88.83%, 93.54%, 90.29%, 91.89%, 0.49, and 17.02) and (88.49%, 92.95%, 90.11%, 90.50%, 0.49, and 17.02), respectively.

From Table 4 and Fig. 9, it is observed that disabling preprocessing brings out the significant difference in accuracy from 88.83% to 88.12%, precision from 93.54% to 92.95%, recall from 90.29 to 90.11, and f1-score from 91.89% to 90.50%. However, MCC and OR remain the same.

The preceding analysis concluded that text preprocessing and features are critical for detecting fake news and

SUMMARY

Groups	Count	Sum	Average	Variance
MultiFND	10	805.95	80.595	2.486583
FakeNED	10	892.92	89.292	1.373973
ELD-FN	10	918.85	91.885	0.577206

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	699.4185	2	349.7093	236.4092	7.75E-18	3.354131
Within Groups	39.93986	27	1.479254			
Total	739.3584	29				

where, SS = sum of squares, DF = degree of freedom, MS = mean square, F = F-test, and p-value determines the significance

FIGURE 7. ANOVA analysis on performance comparison.

disabling either would significantly reduce the performance of ELD-FN.

4) RQ4: COMPARISON OF ELD-FN AGAINST OTHER CLASSIFIERS

We select off-the-shelf deep learning classifiers (CNN and LSTM), the most widely adopted and well-known. Note that the preprocessed text, their sentiment, and feature embeddings are given as input to the selected classifiers for comparative analysis. We set hyper-parameters' values as dropout = 0.2, recurrent_dropout = 0.2, loss function = *binary-crossentropy*, and activation = *sigmoid* for ELD-FN and both baseline approaches.

Table 6 and Fig. 10 present the evaluation metrics for ELD-FN, CNN, and LSTM based on their accuracy, precision, recall, F1-score, MCC, and OR. The results advised that the average values of these metrics for ELD-FN, FakeNED, and MultiFND are (88.83%, 93.54%, 90.29%, 91.89%, 0.49, and 17.02), (86.73%, 92.56%, 85.81%, 89.06%, 0.48, and 16.97),

TABLE 3. Influence of sentiment on ELD-FN.

Sentiment	Accuracy	Precision	Recall	F1-score	MCC	OR
Enable	88.83%	93.54%	90.29%	91.89%	0.49	17.02
Disable	88.12%	90.38%	89.98%	90.17%	0.49	17.02

TABLE 4. Influence of preprocessing on ELD-FN.

Preprocessing	Accuracy	Precision	Recall	F-Measure	MCC	OR
Enable	88.83%	93.54%	90.29%	91.89%	0.49	17.02
Disable	88.49%	92.95%	90.11%	91.50%	0.49	17.02

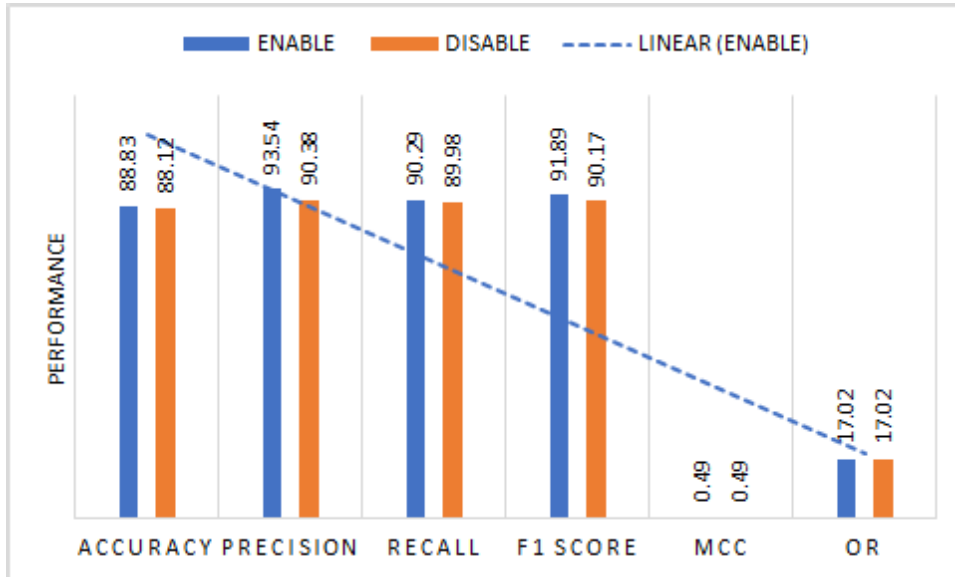


FIGURE 8. Influence of sentiment on ELD-FN.

and (86.51%, 90.22%, 86.19%, 88.21%, 0.48, and 16.92), respectively.

The following observations are made from Table 5 and Fig. 10.

- ELD-FN outperforms CNN and LSTM. The performance enhancement of ELD-FN upon CNN in accuracy, precision, recall, f1-score, MCC, and OR is 2.42%, 1.06%, 5.22%, 3.18%, 0.01, and 0.05, respectively. However, the performance enhancement of ELD-FN upon LSTM in accuracy, precision, recall, f1-score, MCC, and OR is 2.68%, 3.68%, 4.76%, 4.17%, 0.01, and 0.10, respectively.
- ELD-FN performs better than LSTM because LSTM requires short text and performs sequential processing, which is unnecessary in our case. In contrast, CNN is proven efficient for long text and works better to extract local invariant features.

The preceding analysis concluded that ELD-FN outperforms other classifiers in detecting fake news.

F. THREATS TO VALIDITY

The probability of incorrect labeling of news is the first threat to construct validity. This research assumes that the

TABLE 5. Relation between sentiment and news.

News	Total	Positive	Negative
Real	312,061 (45.69%)	34.16%	65.84%
Fake	370,935 (54.31%)	26.29%	73.71%

assigned labels by Nakamura et al. [47] are correct. However, incorrect labeling of data may cause the productivity of ELD-FN.

The choice of assessment metrics of ELD-FN is another threat to construct validity. The chosen metrics for detecting news are the most accepted in the literature for the classification task.

The choice of the sentiment analysis repository is the first threat to internal validity. The chosen repository III-E has been public and has good results in computing sentiment. Exploiting other repositories may cause the productivity of ELD-FN.

ELD-FN, FakeNED, and MultiFND coding is the second threat to internal validity. The coding and the produced results of ELD-FN, FakeNED, and MultiFND are verified to mitigate the threat. However, unknown errors may cause the productivity of ELD-FN.

The hyper-parameters setting of ELD-FN is the third threat to internal validity. The hyper-parameters setting for ELD-FN

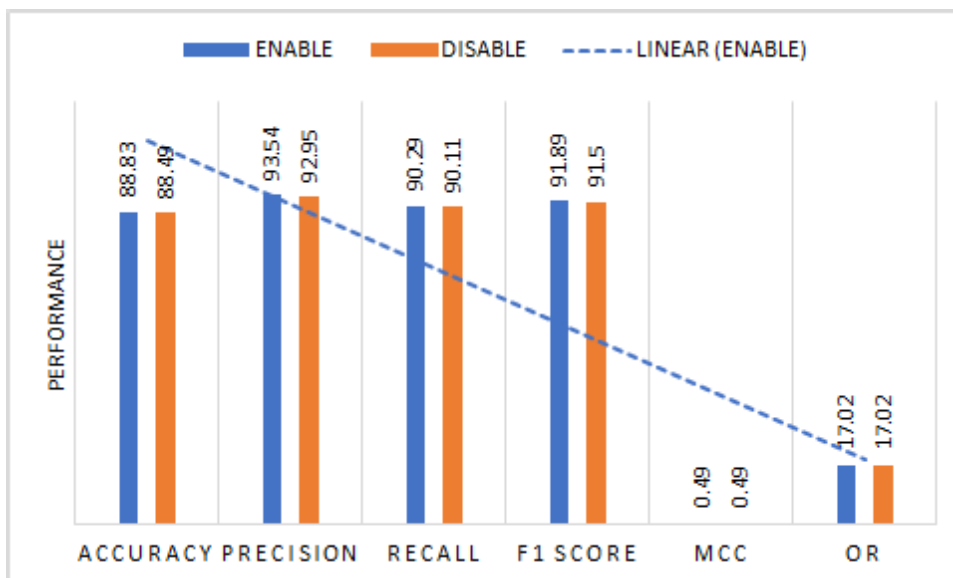


FIGURE 9. Influence of preprocessing on ELD-FN.

TABLE 6. Comparison of ELD-FN against other classifiers.

Classifier	Accuracy	Precision	Recall	F-Measure	MCC	OR
ELD-FN	88.83%	93.54%	90.29%	91.89%	0.49	17.02
CNN	86.73%	92.56%	85.81%	89.06%	0.48	16.97
LSTM	86.51%	90.22%	86.19%	88.21%	0.48	16.92

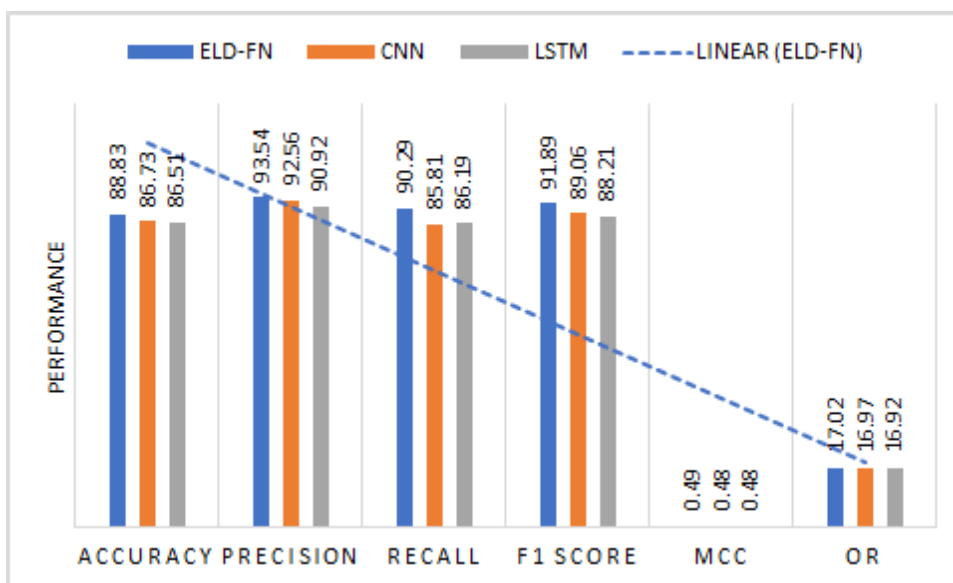


FIGURE 10. Comparison of ELD-FN against other classifiers.

is mentioned in Section IV-E4. The change in settings may cause the productivity of ELD-FN.

V. CONCLUSION AND FUTURE WORK

Automatic fake news detection is crucial to avoid spreading false information that can have serious consequences, ranging from reputational damage to social and political unrest. In some cases, fake news can even incite violence and

lead to harm or loss of life. Therefore, the ability to automatically identify and flag false information can help mitigate the threats of fake news. From this perspective, this paper proposes an ensemble deep learning-based detection of fake news. The proposed approach leverages NLP techniques for preprocessing textual information of news, computes the sentiment from the text of each news, generates embeddings for text and images of the corresponding news

by leveraging V-BERT, and passes the embeddings to the deep learning ensemble model for training and testing. The evaluation results significantly outperform the state-of-the-art approaches in identifying fake news.

In future, we would like to investigate the need to adapt detection algorithms to new types of media. Fake news is not limited to text-based content, and algorithms must be able to detect false information in images, videos, and audio as well. Moreover, we are interested in improving the interpretability of detection algorithms. Current methods often rely on opaque deep learning models, making it difficult to understand how decisions are being made. Future work could focus on developing more transparent models or tools that help users understand how algorithms arrive at their conclusions.

REFERENCES

- [1] S. De Sarkar, F. Yang, and A. Mukherjee, "Attending sentences to detect satirical fake news," in *Proc. 27th Int. Conf. Comput. Linguistics*, 2018, pp. 3371–3380.
- [2] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *J. Econ. Perspect.*, vol. 31, no. 2, pp. 211–236, May 2017.
- [3] A. Moon. (2017). *Two-Thirds of American Adults Get News From Social Media: Survey*. [Online]. Available: <https://uk.reuters.com/article/usa-internet-socialmedia/two-thirds-of-american-adults-get-news-from-social-media-survey-idUKKCN1BJ2A8>
- [4] 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.
- [5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.
- [6] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," Tech. Rep., 2018.
- [7] J. Sarzynska-Wawer, A. Wawer, A. Pawlak, J. Szymanowska, I. Stefaniak, M. Jarkiewicz, and L. Okruszek, "Detecting formal thought disorder by deep contextualized word representations," *Psychiatry Res.*, vol. 304, Oct. 2021, Art. no. 114135.
- [8] L. Harold Li, M. Yatskar, D. Yin, C.-J. Hsieh, and K.-W. Chang, "VisualBERT: A simple and performant baseline for vision and language," 2019, *arXiv:1908.03557*.
- [9] S. Afroz, M. Brennan, and R. Greenstadt, "Detecting hoaxes, frauds, and deception in writing style online," in *Proc. IEEE Symp. Secur. Privacy*, May 2012, pp. 461–475.
- [10] X. Zhou, J. Wu, and R. Zafarani, "SAFE: Similarity-aware multi-modal fake news detection," in *Proc. Advances in Knowledge Discovery and Data Mining*, Cham, Switzerland: Springer, 2020, pp. 354–367.
- [11] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, and M. Cha, "Detecting rumors from microblogs with recurrent neural networks," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, 2016, pp. 3818–3824.
- [12] Y. Chen, J. Sui, L. Hu, and W. Gong, "Attention-residual network with CNN for rumor detection," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, Nov. 2019, pp. 1121–1130.
- [13] J. Ma, W. Gao, and K.-F. Wong, "Detect rumors on Twitter by promoting information campaigns with generative adversarial learning," in *Proc. World Wide Web Conf.*, May 2019, pp. 3049–3055.
- [14] F. Yu, Q. Liu, S. Wu, L. Wang, and T. Tan, "A convolutional approach for misinformation identification," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 3901–3907.
- [15] V. Vaibhav, R. M. Annasamy, and E. Hovy, "Do sentence interactions matter? Leveraging sentence level representations for fake news classification," 2019, *arXiv:1910.12203*.
- [16] L. Wu, Y. Rao, H. Jin, A. Nazir, and L. Sun, "Different absorption from the same sharing: Sifted multi-task learning for fake news detection," 2019, *arXiv:1909.01720*.
- [17] M. Cheng, S. Nazarian, and P. Bogdan, "VRoC: Variational autoencoder-aided multi-task rumor classifier based on text," in *Proc. Web Conf.*, 2020, pp. 2892–2898.
- [18] F. Qian, C. Gong, K. Sharma, and Y. Liu, "Neural user response generator: Fake news detection with collective user intelligence," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 3834–3840.
- [19] A. Giachanou, P. Rosso, and F. Crestani, "Leveraging emotional signals for credibility detection," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2019, pp. 877–880.
- [20] K. Wu, S. Yang, and K. Q. Zhu, "False rumors detection on Sina Weibo by propagation structures," in *Proc. IEEE 31st Int. Conf. Data Eng.*, Apr. 2015, pp. 651–662.
- [21] P. Li, X. Sun, H. Yu, Y. Tian, F. Yao, and G. Xu, "Entity-oriented multi-modal alignment and fusion network for fake news detection," *IEEE Trans. Multimedia*, vol. 24, pp. 3455–3468, 2022.
- [22] P. Qi, J. Cao, X. Li, H. Liu, Q. Sheng, X. Mi, Q. He, Y. Lv, C. Guo, and Y. Yu, "Improving fake news detection by using an entity-enhanced framework to fuse diverse multimodal clues," in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 1212–1220.
- [23] C. Song, N. Ning, Y. Zhang, and B. Wu, "A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks," *Inf. Process. Manage.*, vol. 58, no. 1, Jan. 2021, Art. no. 102437.
- [24] L. D. Sciuca, M. Mameli, E. Balloni, L. Rossi, E. Frontoni, P. Zingaretti, and M. Paolanti, "FakeNED: A deep learning based-system for fake news detection from social media," in *Proc. Int. Conf. Image Anal. Process.*, 2022, pp. 303–313.
- [25] I. Segura-Bedmar and S. Alonso-Bartolome, "Multimodal fake news detection," *Information*, vol. 13, no. 6, p. 284, Jun. 2022. [Online]. Available: <https://www.mdpi.com/2078-2489/13/6/284>
- [26] F. Yang, Y. Liu, X. Yu, and M. Yang, "Automatic detection of rumor on Sina Weibo," in *Proc. ACM SIGKDD Workshop Mining Data Semantics*, Aug. 2012, pp. 1–7.
- [27] Z. Jin, J. Cao, Y. Zhang, J. Zhou, and Q. Tian, "Novel visual and statistical image features for microblogs news verification," *IEEE Trans. Multimedia*, vol. 19, no. 3, pp. 598–608, Mar. 2017.
- [28] C. Boididou, S. Papadopoulos, D.-T. Dang-Nguyen, G. Boato, and Y. Kompatsiaris, "The certh-unitn participation@ verifying multimedia use 2015," *MediaEval*, vol. 1, p. 2, May 2015.
- [29] B. Emek Soylu, M. S. Guzel, G. E. Bostanci, F. Ekinci, T. Asuroglu, and K. Acici, "Deep-learning-based approaches for semantic segmentation of natural scene images: A review," *Electronics*, vol. 12, no. 12, p. 2730, Jun. 2023. [Online]. Available: <https://www.mdpi.com/2079-9292/12/12/2730>
- [30] Q. S. Hamad, H. Samma, and S. A. Suandi, "Feature selection of pre-trained shallow CNN using the QLESCA optimizer: COVID-19 detection as a case study," *Appl. Intell.*, vol. 53, no. 15, pp. 18630–18652, Feb. 2023, doi: 10.1007/s10489-022-04446-8.
- [31] S. R. Shah, S. Qadri, H. Bibi, S. M. W. Shah, M. I. Sharif, and F. Marinello, "Comparing inception v3, VGG 16, VGG 19, CNN, and ResNet 50: A case study on early detection of a Rice disease," *Agronomy*, vol. 13, no. 6, p. 1633, Jun. 2023. [Online]. Available: <https://www.mdpi.com/2073-4395/13/6/1633>
- [32] 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*, 2018, pp. 849–857.
- [33] 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.
- [34] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [35] S. Singhal, R. R. Shah, T. Chakraborty, P. Kumaraguru, and S. Satoh, "SpotFake: A multi-modal framework for fake news detection," in *Proc. IEEE 5th Int. Conf. Multimedia Big Data (BigMM)*, Sep. 2019, pp. 39–47.
- [36] X. Zhou and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," *ACM Comput. Surv.*, vol. 53, no. 5, pp. 1–40, Sep. 2021.
- [37] J. Xue, Y. Wang, Y. Tian, Y. Li, L. Shi, and L. Wei, "Detecting fake news by exploring the consistency of multimodal data," *Inf. Process. Manage.*, vol. 58, no. 5, Sep. 2021, Art. no. 102610.
- [38] F. Ghorbanpour, M. Ramezani, M. A. Fazli, and H. R. Rabiee, "FNR: A similarity and transformer-based approach to detect multi-modal fake news in social media," *Social Netw. Anal. Mining*, vol. 13, no. 1, pp. 1–15, Mar. 2023.

[39] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[40] 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. 795–816.

[41] H. Zhang, Q. Fang, S. Qian, and C. Xu, "Multi-modal knowledge-aware event memory network for social media rumor detection," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 1942–1951.

[42] C. Song, C. Yang, H. Chen, C. Tu, Z. Liu, and M. Sun, "CED: Credible early detection of social media rumors," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 8, pp. 3035–3047, Aug. 2021.

[43] S. Qian, J. Wang, J. Hu, Q. Fang, and C. Xu, "Hierarchical multi-modal contextual attention network for fake news detection," in *Proc. 44th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2021, pp. 153–162.

[44] Y. Wu, P. Zhan, Y. Zhang, L. Wang, and Z. Xu, "Multimodal fusion with co-attention networks for fake news detection," in *Proc. IJCNLP*, 2021, pp. 2560–2569.

[45] W. Zhang, L. Gui, and Y. He, "Supervised contrastive learning for multimodal unreliable news detection in COVID-19 pandemic," in *Proc. 30th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2021, pp. 3637–3641.

[46] T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "Comparing boosting and bagging techniques with noisy and imbalanced data," *IEEE Trans. Syst., Man, Cybern., A, Syst. Hum.*, vol. 41, no. 3, pp. 552–568, May 2011.

[47] K. Nakamura, S. Levy, and W. Y. Wang, "r/Fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection," in *Proc. Int. Conf. Lang. Resour. Eval.*, 2020, pp. 1–9.

[48] M. Tausif, S. Dilshad, Q. Umer, M. W. Iqbal, Z. Latif, C. Lee, and R. N. Bashir, "Ensemble learning-based estimation of reference evapotranspiration (ETO)," *Internet Things*, vol. 24, Feb. 2023, Art. no. 100973. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2542660523002962>



MUHAMMAD FAHEEM (Member, IEEE) received the B.Sc. degree in computer engineering from the Department of Computer Engineering, University College of Engineering and Technology, Bahauddin Zakariya University, Multan, Pakistan, in 2010, the M.S. degree in computer science from the Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia (UTM), Johor Bahru, Malaysia, in 2012, and the Ph.D. degree in computer science from the Faculty of Engineering, UTM, in 2021. From 2012 to 2014, he was a Lecturer with the COMSATS Institute of Information and Technology, Pakistan. From 2014 to 2022, he was also an Assistant Professor with the Department of Computer Engineering, Abdullah Gul University, Turkey. He is currently a Researcher with the School of Computing (Innovations and Technology), University of Vaasa, Vaasa, Finland. He has authored several papers in refereed journals and conferences. His research interests include cybersecurity, blockchain, artificial intelligence, smart grids, and smart cities. He served as a reviewer for numerous journals in IEEE, Elsevier, Springer, Wiley, Hindawi, and MDPI.



WAHEED YOUSUF RAMAY received the Ph.D. degree from the University of Science and Technology Beijing (USTB) China. He is currently an Assistant Professor with Air University. His academic and clinical focus is the use of algorithms (deep learning, machine learning, and big data analysis), advanced text analysis techniques, and sentiment analysis.



MALIK KHIZAR SAEED received the B.S. degree in information technology from the University of Gujrat, Gujrat, Pakistan, in 2013, and the M.S. degree in computer science from COMSATS University Islamabad, Vehari Campus, Pakistan. He is currently working as a Visiting Lecturer at COMSATS University Islamabad. His research interests include machine learning, deep learning, and artificial intelligence. He is also interested in classification-related tasks using different ML approaches.



MUHAMMAD LUQMAN received the bachelor's degree in computer science from the University of Gujrat, Pakistan, in 2017, and the master's degree in computer science from Northwestern Polytechnical University, China. He is currently a young Scholar in the field of computer science. His research interests include wide spectrum, primarily focusing on cutting-edge fields, such as artificial intelligence, deep learning, and data mining.



MAJID BASHIR AHMAD received the master's degree in computer science from COMSATS University Islamabad, Pakistan, in 2014, and the M.S. degree in computer science from The University of Lahore, Pakistan, in 2019. He is currently a Research Scholar in the field of computer science. His research interests include artificial intelligence, machine learning, and data mining.