


Article

The Power of Context: A Novel Hybrid Context-Aware Fake News Detection Approach

Jawaher Alghamdi ^{1,2,*}, Yuqing Lin ¹  and Suhuai Luo ¹

¹ School of Information and Physical Sciences, College of Engineering Science and Environment, The University of Newcastle, Newcastle 2308, Australia; yuqing.lin@newcastle.edu.au (Y.L.); suhuai.luo@newcastle.edu.au (S.L.)

² Department of Computer Science, King Khalid University, Abha 62521, Saudi Arabia

* Correspondence: jawaher.alghamdi@uon.edu.au

Abstract: The detection of fake news has emerged as a crucial area of research due to its potential impact on society. In this study, we propose a robust methodology for identifying fake news by leveraging diverse aspects of language representation and incorporating auxiliary information. Our approach is based on the utilisation of Bidirectional Encoder Representations from Transformers (BERT) to capture contextualised semantic knowledge. Additionally, we employ a multichannel Convolutional Neural Network (mCNN) integrated with stacked Bidirectional Gated Recurrent Units (sBiGRU) to jointly learn multi-aspect language representations. This enables our model to effectively identify valuable clues from news content while simultaneously incorporating content- and context-based cues, such as user posting behaviour, to enhance the detection of fake news. Through extensive experimentation on four widely used real-world datasets, our proposed framework demonstrates superior performance ($\uparrow 3.59\%$ (PolitiFact), $\uparrow 6.8\%$ (GossipCop), $\uparrow 2.96\%$ (FA-KES), and $\uparrow 12.51\%$ (LIAR), considering both content-based features and additional auxiliary information) compared to existing state-of-the-art approaches, establishing its effectiveness in the challenging task of fake news detection.

Keywords: fake news detection; deep learning; language representation; BERT model; contextualized semantic knowledge; auxiliary information integration



Citation: Alghamdi, J.; Lin, Y.; Luo, S. The Power of Context: A Novel Hybrid Context-Aware Fake News Detection Approach. *Information* **2024**, *15*, 122. <https://doi.org/10.3390/info15030122>

Academic Editors: Jennifer D'Souza and Chengzhi Zhang

Received: 24 January 2024

Revised: 14 February 2024

Accepted: 19 February 2024

Published: 21 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Fake news detection is a research area that has attracted much attention. People are increasingly utilising social networking platforms, such as Twitter, to share their opinions. Detecting fake news is crucial, yet it is challenging due to the intricate semantics of natural language and the high dimensionality of textual data, leading to data sparsity. Furthermore, malicious actors frequently modify their writing style to imitate credible content, making it difficult to identify fake news based solely on cues from the news content. It is, therefore, prudent to consider auxiliary information, such as user behaviour clues, to advance fake news detection.

Previous studies in the field of automatic fake news detection have predominantly focused on content-based features, including n-grams and part-of-speech tags [1–5]. However, this approach often fails to consider the significance of simultaneously learning various aspects of language to achieve effective detection. Moreover, some researchers [6] have attempted to differentiate fake content by incorporating both content and context characteristics, employing unidirectional encoding of news content with widely-used pre-trained word embedding models such as GloVe [7]. Nonetheless, a practical approach is needed to capture contextualised semantic patterns that classical statistical approaches or context-independent representation models cannot adequately model. To overcome the limitations of previous approaches and address the issue of considering only a single aspect

of language, our proposed framework for fake news detection takes a multi-aspect representation approach to model the input text. This comprehensive approach incorporates various language levels, including content, style, morality, and sentiment. By considering multiple aspects simultaneously, our framework provides a more holistic understanding of the textual data, enhancing the detection process. In addition, we propose the adoption of context-aware representation models, such as BERT [8], to encode the text input. By utilising BERT and its ability to capture contextual knowledge, our framework gains an additional layer of understanding, leading to improved performance in detecting fake news. To evaluate the effectiveness of our framework, we conduct experiments using two different types of text representations for news content: context-independent pre-trained embedding models like GloVe and context-aware pre-trained embedding models like BERT. This comparison allows us to analyse the impact of incorporating contextual information on the detection performance. Furthermore, we assess the performance of our framework across various scenarios, including short- and long-text news content, small- and large-scale datasets, and the inclusion of a rich set of crucial auxiliary features. This comprehensive evaluation enables us to thoroughly analyse our architecture's capabilities and potential for real-world applications.

This paper focuses on two problems: (i) how to detect fake news by using multi-aspect language representations and (ii) how to process multiple resolutions of news at once while simultaneously learning how to best integrate these interpretations and other contextual information through joint feature learning.

To respond to these challenges, this work presents a context-aware fake news detection framework (BERT_{base}-mCNN-sBiGRU) by jointly modelling context- and content-based clues through a coherent process that consists of (1) encoding news content using the pre-trained BERT model, (2) using multi-channel CNN [9] (mCNN) with three input channels to process different resolutions of the input text, and (3) introducing a stacked BiGRU [10] (sBiGRU) to encode the given auxiliary information, allowing the model to capture more contextual semantical information. Our key contributions are summarised as follows:

1. The development of a novel hybrid deep learning framework that can effectively learn from multiple sources to detect fake news. These sources include news content, social user behaviour information, and various language aspects.
2. The evaluation of the proposed framework uses two different embedding models, BERT and GloVe, to determine its efficacy.
3. The evaluation of the proposed framework through extensive experimentation on four real-world datasets to demonstrate its effectiveness in detecting fake news.
4. The discovery that incorporating user behavioural representation with content-based information can lead to more accurate outcomes compared to existing state-of-the-art baselines.

Overall, our study focused on developing a more effective way of detecting fake news by combining various sources of information and using a hybrid deep learning framework. The findings of the study suggest that this approach can yield superior outcomes compared to existing state-of-the-art methods.

The remainder of this paper is structured as follows: Section 2 summarises relevant literature on fake news detection. Section 3 outlines our research methodology and introduces the proposed model. In Section 4, we present comprehensive outcomes on the performance of the predictive models, including all other models developed in this study. Finally, Section 5 concludes the paper.

2. Related Work

Current research focuses on detecting fake news using both contextual and content-based approaches. The content-based features used in such research can be broadly classified into two groups [11]: general features and latent features. General textual features, commonly used in traditional machine learning models, employ statistical techniques like

bag-of-words (BoW) models to calculate the frequency statistics of lexicons and parts-of-speech (POS) tags [12], such as nouns and verbs, to evaluate the syntax [11].

Conversely, latent textual features capture implicit patterns or embeddings that can be generated at the word [13], sentence [14], or document level [14]. The outcome of this process is the generation of compact vector representations, which can be exploited for further analysis. The existence of irrelevant or noisy text in fake news datasets, particularly those extracted from social media platforms like Twitter, poses a challenge to automatic fake news detection. Failure to process this text can negatively affect the detection performance. Addressing this challenge requires encoding news content in a way that mitigates the problem. In order to achieve this goal, several neural network-based models have been suggested, each offering distinct and valuable features that aid in distinguishing between real and fake news [5]. As an example, Wang et al. [15] conducted a study that employed Bidirectional Long Short-Term Memory (BiLSTM) and Convolutional Neural Networks (CNNs) to encode the combination of textual and speaker metadata, aiming to identify fake news.

News content could have various sized lengths. In fact, despite the informative information longer text can provide, it may lead to the presence of noisy words or sentences. Contrary to static embeddings produced by traditional context-independent word embedding methods (word2vec and GloVe), more advanced pre-trained contextualised embedding models based on the attention mechanism such as BERT [8] can be exploited to provide embeddings of a word based on its context. However, using such an advanced pre-trained embedding model presents limitations to the length of a given input text where such a text must be truncated or padded to the maximum length imposed by the model.

In [16], the authors proposed a deep learning technique termed FakeBERT, which combines BERT with parallel blocks of a deep CNN featuring diverse kernel sizes and filters. This integration has demonstrated efficacy in mitigating ambiguity, a notable challenge in natural language comprehension. Nonetheless, the authors overlooked the potential benefits of integrating user behaviour cues, which could potentially improve classification accuracy. The comparative study conducted by [5] investigates the effectiveness of different machine learning and deep learning models, showcasing the capacity of BERT and its variations to improve detection performance across a range of datasets. Alghamdi et al. [17] proposed a computational framework for automatic fake news detection leveraging BERT with the LIAR dataset. The methodology employed BERT for encoding the input text, accompanied by a CNN for extracting local features. Furthermore, metadata information was encoded using a combination of CNN, BiLSTM, and a classification layer. The findings showcased enhanced performance in comparison to prior state-of-the-art techniques on the LIAR multiclass classification task. Additional data from social media platforms can provide valuable insights into the dissemination of fake news, despite potential challenges such as noise and inconsistency. Alongside the news content itself, the study incorporates auxiliary information, such as post-based features, extracted from source tweets within the Twitter context. The authors in [18] proposed an outlier knowledge management framework designed to identify fake news during emergency situations, incorporating principles from complex adaptive systems theory. Their hybrid model, which integrates CNN, BiLSTM networks, and attention mechanisms, demonstrates enhanced detection metrics while providing valuable insights into the characteristics of fake news. The authors in [19] introduced an arithmetic optimization algorithm (AOA)-based approach designed to improve classification accuracy through feature reduction. Utilizing AOA as a wrapper feature-selection technique, the study conducted extensive simulations, comparing the proposed method against established classifiers and alternative evolutionary approaches.

Numerous studies have leveraged contextual cues from social user posts, including temporal patterns observed in sequences of responses on platforms like Twitter, along with other features reflecting their engagements and interactions [5]. User-based features are also believed to be valuable clues in detecting fake news content. Because users prone to sharing fake news have distinctive traits from those who do not, researchers have become

interested in exploring user-based cues for identifying fake news [20]. For instance, in their study, Shu et al. [20] examined user profiles to distinguish fake content from real.

It has also been demonstrated that the language used by purveyors of fake news contains strong explicit or implicit indicators that can be harnessed to advance fake news detection. The former includes lexicon and linguistic cues, such as stylistic features like part-of-speech (POS) tags, while the latter refers to implicit clues such as sentimental and emotional cues. Table 1 shows different features used in previous related work. Few studies have integrated such multiple aspect clues to detect fake news. Incorporating various aspects of language alongside a diverse array of user behaviour cues could potentially enhance the accuracy of fake news detection. In this study, we introduce a context-aware hybrid framework that integrates multi-aspect language representations, behavioural information, and a diverse set of relevant text-based features. Our experiments illustrate that the fusion of these features yields significantly improved results compared to existing baselines.

Table 1. Previous studies on fake news and rumours detection using various features.

Study	Cues			Approach
	TF	UF	PF	
Vosoughi et al. [21]	✓		✓	Hidden Markov Models
Ma et al. [22]	✓			RNN
Chen et al. [23]	✓		✓	Anomaly detection, KNN
Wu et al. [24]	✓			LSTM-RNN
Gupta et al. [25]	✓		✓	Graph-based method
Gupta et al. [26]		✓	✓	Graph-based method, DT
Qazvinian et al. [27]	✓			L_1 -regularized log-linear model
Zhao et al. [28]	✓			DT ranking method
Chua et al. [29]	✓			Linear Regression (LR)
Ma et al. [30]	✓			Kernel-based method
Kwon et al. [31]	✓		✓	Random Forest
Kwon et al. [32]	✓		✓	SpikeM
Zubiaga et al. [33]	✓		✓	Conditional Random Fields
Qin et al. [34]	✓			SVM
Shu et al. [35]	✓		✓	Neural Network
Jin et al. [36]	✓		✓	LDA, Graph
Li et al. [37]	✓		✓	SVM
Li et al. [38]	✓		✓	LSTM
Shu et al. [6]	✓			Hierarchical Attention Network
Ours	✓	✓	✓	BERT-mCNN-sBiGRU

Note that TF: Textual Features, UF: User Features, PF: Propagation Features.

2.1. Preliminaries

2.1.1. Global Vectors for Word Representation (GloVe)

The Global Vectors for Word Representation method, known as GloVe, was developed by Pennington et al. [7] to enhance the process of learning word vectors. Building upon the word2vec approach, GloVe is more efficient in acquiring word embeddings. By combining global statistics from matrix factorization methods like LSA with context-based learning techniques such as word2vec, GloVe has gained widespread recognition as a superior method for generating word embeddings. Transfer learning, a machine learning technique, involves storing knowledge gained from a specific task and utilising it to solve related problems, thereby improving the learning process. This technique is particularly valuable when faced with limited training data and the need to evaluate models. In Natural Language Processing (NLP), transfer learning has made significant advancements by utilising pre-trained embedding models trained on large text corpora. This has resulted in remarkable breakthroughs in NLP. However, these methods have encountered challenges in distinguishing the context in which words are written. To address this issue, contextualised word embedding models have been introduced, such as BERT [8], which is specifically designed to capture contextual information effectively.

2.1.2. Bidirectional Encoder Representations from Transformers (BERT)

Contextualised word embedding models have become increasingly significant in recent years, surpassing the limitations of traditional context-free neural embedding models like word2vec and GloVe in capturing deep contextual relationships. These traditional models focus on short-range context within a specific co-occurrence window, which restricts their ability to grasp nuanced contextual information [5]. Consequently, their popularity has diminished in favor of transfer learning methods, with Google's BERT model leading the way. The BERT model, introduced by Devlin et al. [8], is an unsupervised language representation model that revolutionised the field. Unlike its predecessors, BERT incorporates a deeply bidirectional architecture that simultaneously considers both the forward and backward contexts in all layers, resulting in highly context-aware embeddings. The attention mechanism plays a pivotal role in improving the semantic representation of words within a given context by recognising their varying impacts. This mechanism is a crucial component of the transformer architecture, designed to assign different weights to different parts of the input text, thus distinguishing their contributions to the final output [5]. To achieve this, the attention mechanism transforms each word into matrix vectors, namely Q , K , and V , through separate linear transformations, and independently calculates the associations between words. This process allows BERT to capture intricate contextual dependencies and generate more nuanced and informative word embeddings.

The formula for scaled dot-product attention can be observed in Equation (1) [39].

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

The query, key, and value vectors are represented by Q , K , and V , respectively. In order to normalise the inputs to a value between 0 and 1, the attention mechanism employs the Softmax activation function. BERT utilises a multi-head attention mechanism, which is based on the transformer's encoder, as expressed in Equation (2) [39], where the subscript i represents each specific head and its corresponding weight matrices.

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O \quad (2)$$

where each $head_i$ is calculated as follows:

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (3)$$

Text classification tasks have demonstrated remarkable success through the implementation of the BERT model, but their success comes with a computational cost due to the millions of parameters required. Specifically, BERT_{base} has 110 million parameters, while BERT_{large} has 340 million parameters [8]. Nevertheless, BERT_{base} provides exceptional results and is simpler to train than BERT_{large}. In this study, the authors utilised BERT_{base} to generate context-aware text representations for the input text provided.

2.1.3. Long Short-Term Memory (LSTM)

The LSTM network, introduced by [40], represents a notable advancement within the realm of recurrent neural networks (RNNs). By incorporating three distinct types of gates—namely the input, forget, and output gates—LSTM overcomes certain limitations of traditional RNNs. These gates play a critical role in regulating the flow of information within the LSTM cells, thereby mitigating issues such as gradient vanishing and explosion. Consequently, LSTM has proven to be highly effective in handling long sentences, as evidenced by prior research [41]. LSTM components can be formulated mathematically as follows [42]:

$$f_t = \sigma(W_{xf} \cdot x_t + W_{hf} \cdot h_{t-1} + b_f), \quad (4)$$

$$i_t = \sigma(W_{xi} \cdot x_t + W_{hi} \cdot h_{t-1} + b_i), \quad (5)$$

$$\tilde{C}_t = \tanh(W_{xc} \cdot x_t + W_{hc} \cdot h_{t-1} + b_i), \quad (6)$$

$$C_t = f_t \odot C_{t-1} + i \odot \tilde{C}_t, \quad (7)$$

$$o_t = \sigma(W_{xo} \cdot x_t + W_{ho} \cdot h_{t-1} + b_o), \quad (8)$$

$$h_t = o_t \odot \tanh(C_t). \quad (9)$$

In the formulas above, σ represents the logistic sigmoid activation function. W , b , and C_t , represent the weight matrix, the bias, and the state of the memory unit at time t , respectively.

Nonetheless, one key drawback of the basic RNN architecture, including LSTM, is its limited capacity to consider future context. While these models can successfully capture dependencies based on previous context, their inability to account for subsequent context poses a notable limitation. Alternative architectures, such as Bidirectional LSTM (BiLSTM) and Bidirectional Gated Recurrent Unit (BiGRU), have been proposed as potential solutions to address this issue. These models comprise both forward and backward hidden layers, which are subsequently merged to facilitate the flow of temporal information in both directions [5]. Consequently, BiLSTM and BiGRU models offer superior learning performance by effectively considering both the preceding and subsequent contexts.

2.1.4. Gated Recurrent Unit (GRU)

As a type of RNN, the GRU has two gates: the update and reset gates. The update gate controls the amount of information that needs to be transferred to the current state, which combines the forget and input gates. Additionally, the reset gate determines when to discard the previous hidden state. The update and reset gates are computed similarly to LSTM as follows [10]:

$$r_t = \delta(W_r h_{t-1} + U_r x_t + b_r), \quad (10)$$

$$z_t = \delta(W_z h_{t-1} + U_z x_t + b_z), \quad (11)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t, \quad (12)$$

$$\tilde{h}_t = \tanh(W_{\tilde{h}_t} (h_{t-1} \odot r_t) + U_{\tilde{h}_t} x_t). \quad (13)$$

In the formulas above, $\delta(\cdot)$ signifies the logistic sigmoid function, W and U are gate weight matrices, and h_t and b are the hidden state and bias vectors, respectively.

2.1.5. Convolutional Neural Network (CNN)

One-dimensional Convolutional Neural Networks (Conv1D) have emerged as a popular choice for prediction generation and have demonstrated their efficacy across a range of NLP tasks. Conv1D models employ a fixed window size filter, which slides over the input data during training. Each cell within the filter is initialised with a weight, and at each processing step, the input, typically consisting of word vectors, undergoes element-wise multiplication with the filter weights [5]. This process generates an output array referred to as the feature map or filter output array, encoding salient features extracted from the input data. Multi-channel CNNs have been particularly effective for text classification tasks [9]. CNNs are well-regarded for their ability to automatically extract relevant features, allowing them to capture local features with precision. A multi-channel CNN can capture diverse features across different regions of the input data by employing multiple channels, each with its own set of filters. This multi-channel approach enables the model to capture both low-level and high-level features, enhancing its ability to discern intricate patterns and improve classification performance.

3. Proposed Hybrid Model

3.1. Features

3.1.1. Content-Based Features

News content representations. News content provides valuable clues for distinguishing between fake and real news. We utilised BERT representations (embeddings) to encode each news text. The selection of this pre-trained BERT model is justified by its outstanding performance across a variety of NLP tasks. For the text representations, we initialised the networks with 100-dimensional pre-trained embeddings using the off-the-shelf context-free embedding method (i.e., GloVe). Preprocessing of the input text was performed using the NLTK package, including lowercasing, tokenization, stemming, and punctuation removal. Notably, these preprocessing steps were not applied to BERT's input text due to its built-in tokenizer and punctuation handling capability; for more details, refer to [43]. Our BERT-based models exhibited a significant drop in performance when punctuation was removed. According to [11], distinguishing fake news from the truth involves various factors, including quality, style, and quantity (e.g., word count and expressed sentiments). We, therefore, explored a range of content-based features as follows.

Stylistic features. We extract word-based features such as word length, the number of words beginning with capital letters and those starting with lowercase letters, character length, count of digits, exclamation marks, question marks, and periods.

Sentimental features. Fake news creators tend to write content that provokes readers' emotions to promote their creations' success and draw wide public attention. That is, fake news usually has a strong positive or negative sentiment of hate, anger or resentment [44]. Social science research shows that news stories that evoke high-arousal or triggering emotions (such as awe, anger, or anxiety) go viral more often on social media [45,46]. It is thus imperative to extract useful sentimental clues in our text. We extract such features (i.e., containing two categories: positive and negative) using the NRC lexicon [47]. It is claimed that negative emotions with stronger intensity expressed in fake news content are expected to provoke intense emotions in the public [48]. Consequently, hyperbolic-based characteristics, such as words that convey intense positive or negative sentiments like "terrifying", extracted from clickbait news headlines are also considered [49].

Morality features. Similar to [50], we extract useful moral-based features from the Moral Foundations Dictionary (<https://moralfoundations.org/other-materials/> (accessed on 15 July 2023)) where the categorisation scheme involves assigning words to certain categories, which include fairness, unfairness, care, loyalty, purity, authority, harm, betrayal, subversion, and degradation.

3.1.2. Context-Based Features

For the FakeNewsNet dataset, to obtain the contextual information (i.e., user posting behaviour) for each news article, we collect set of tweets related to each news article and summarise some features (e.g., the number of likes, number of retweets, number of verified tweets, etc.) over such set of tweets for each news article. For modelling, we use the news articles and such tweets. For the LIAR dataset, we consider useful cues from user profile information. We extract temporal features derived from the date attribute as additional useful cues for the FA-KES dataset.

3.2. Model Architecture

The proposed approach to detecting fake news involves a binary classification task in which we aim to develop a model that predicts the credibility of a given news article. Specifically, the model determines whether a given article is fake or real based on certain attributes. The BERT_{base}-mCNN-sBiGRU (See Figure 1) consists of two main modules: the mCNN module and the sBiGRU module. The former (right branch) consists of four layers: input, embedding, convolution (different filter channels), max-pooling, and flattening. First, BERT_{base} is used to generate 768-d vector representations for input text (statements/news), which are fed into what we call the mCNN module. The mCNN module is defined with

three input channels (We also experimented with increasing the number of input channels in our model; however, this led to a decrease in performance) for processing different n-grams of the input text. The model consists of several channels, each comprising three layers: convolution, ReLU activation, and kernel size set to 4, 6, or 8 g. These layers work together to extract different word n-gram features and capture more complex ones. Additionally, a max-pooling layer consolidates the output and facilitates the extraction of the most important features from each feature map. A flattening layer then reduces the three-dimensional output to a two-dimensional one in preparation for concatenation later. Finally, the extracted features from the three channels are concatenated into a single vector to prepare for concatenation with the output of the stacked BiGRU. Using multiple convolution kernels at the cost of the model’s complexity gives the model far more expressive power by allowing it to learn high-level contextual features.

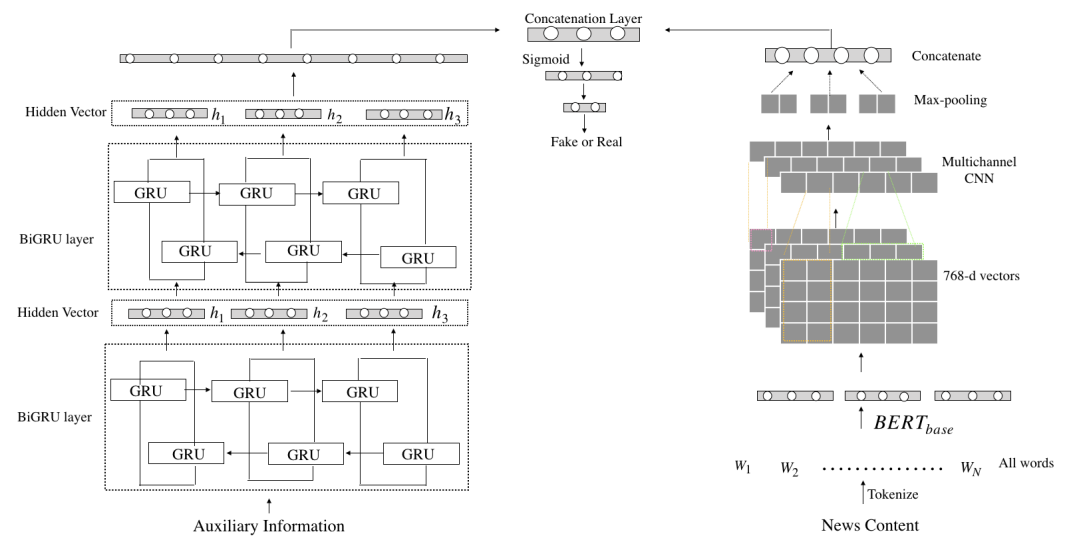


Figure 1. The high-level structure of the proposed approach comprises three components: (1) the news content module; (2) the multi-modalities module, where the former is used to model semantic contextualised representations from the news content while the latter runs gated layers to learn the multimodal information and (3) the classifier component to make prediction by fusing information of these two modules.

The latter (left branch) consists of a stacked BiGRU layer with 50 units to encode the auxiliary information by using multiple BiGRU layers run for the same number of steps in order to learn long-term bidirectional (forward and backward) dependencies. The resultant output is concatenated with the previously mentioned single merged vector. After passing the data through a dense layer comprising one unit with a Sigmoid activation function, the model is constructed and trained using the Adam optimizer and binary-cross entropy as the loss function. The training is conducted for seven epochs, with a batch size of 16, on a single GPU.

It is important to mention that these hyper-parameter values were chosen after multiple runs, as they consistently yielded the best results.

We have chosen to utilise a multi-channel CNN approach for encoding the input text due to its ability to handle multiple news resolutions simultaneously. This allows us to process different aspects of the news content and integrate them effectively by jointly learning features. By employing this approach, we aim to capture a comprehensive understanding of the input text. In addition to capturing textual information, we recognise the importance of behavioural patterns in detecting fake news. We have incorporated various behavioural features to address this, including user posting behaviour, sentiments, morality, etc. To model these patterns effectively, we have employed a stacked BiGRU architecture. In this architecture, multiple BiGRU layers are run for the same number of

time steps to encode behavioural patterns so as to capture the implicit and explicit temporal dependencies and context in such data. By leveraging the bidirectional nature of BiGRU, we can capture information from both past and future contexts, allowing us to model the dynamics of the behavioural clues effectively.

4. Experiments

In this section, we describe the experiments carried out to evaluate the effectiveness of the proposed architecture. The experiments were conducted using the Python programming language, and the models were built using the TensorFlow library. We evaluate the performance of the models using five commonly used evaluation criteria for text classification tasks: accuracy, precision, recall, and F1 measure. The datasets are divided into training (80%) and test (20%) sets to evaluate the model's performance.

4.1. Datasets

The effectiveness of the proposed framework is evaluated using three publicly available datasets, detailed as follows.

4.1.1. FakeNewsNet Dataset

In line with [51], we recognise the importance of analysing user engagement with news articles on social media platforms to improve the detection of fake news. Therefore, we incorporate relevant information from user-news interactions to enhance the accuracy of our detection framework. To facilitate this, we utilised the FakeNewsNet (<https://github.com/KaiDMML/FakeNewsNet> (accessed on 1 January 2023)) dataset, which was collected from two fact-checking platforms—PolitiFact and GossipCop. This dataset consists of labelled news articles and includes social context information obtained from Twitter, such as user engagements/activities. This social user posting behaviour encompasses various features, including follower count, favourite count, retweets count, and verified tweets count. In our approach, we utilised this dataset as input for the BERT-mCNN module to encode the news articles. Through the utilisation of this module, we captured the contextualised semantic information within the news articles, facilitating a comprehensive representation of the textual content. Simultaneously, we used the user posting behaviour clues as input to the sBiGRU module. This module effectively encoded the temporal dynamics and patterns of user interactions by leveraging the sBiGRU module. The statistics of the dataset are shown in Table 2.

Table 2. The statistical information of the FakeNewsNet dataset.

Dataset	PolitiFact	GossipCop
No. Candidate News	694	18676
No. True News	356	14129
No. Fake News	338	4547

4.1.2. FA-KES Dataset

The FA-KES (<https://zenodo.org/record/2607278#.X3oK8WgzaUk> (accessed on 2 February 2023)) dataset comprises 804 news articles concerning the Syrian war. Each article includes the full body of text, headline, news sources, date, location, and two class labels denoted as '0' and '1' for fake news and real news, respectively. We utilise the BERT-mCNN module, as detailed in Section 3.2, to encode both the headline and news text. Additionally, numerical features such as date features (month, day, year, and weekday) are incorporated. These numerical features are vital for providing temporal information associated with the news and are encoded using the sBiGRU module. The statistics of the dataset are shown in Table 3.

Table 3. The statistical information of the FA-KES dataset.

No. Candidate News	804
No. True News	426
No. Fake News	378

4.1.3. LIAR Dataset

The LIAR dataset (<https://www.cs.ucsb.edu/william/data/liardataset.zip> (accessed on 19 February 2023)) consists of approximately 12,800 pieces of information. This dataset comprises two main components: the user profile (e.g., subject, speaker’s name, credit history, etc.) and short political statements. The statements were reported between 2007 and 2016 and were categorised by the editors of Politifact.com using six fine-grained categories. For our study, we considered a binary classification problem where we categorised statements labelled as true, mostly true, and half true as “true” while the remaining were considered “false”. In our approach, we have incorporated textual auxiliary features, such as subject, speaker’s name, etc., by appending them to the end of each corresponding statement. This allows us to integrate these additional textual attributes into the encoding process of the model. On the other hand, numerical features such as credit history features are used as direct inputs to the sBiGRU module; see Section 3.2 for more details. The statistics of the dataset are tabulated in Table 4.

Table 4. The statistical information of the LIAR dataset.

No. Candidate News	12791
No. True News	7134
No. Fake News	5657

We hypothesise that utilising content-based features and contextual cues can potentially improve detection performance. Based on this hypothesis, we extract content-based features described in Section 3.1.1 and experiment with different datasets containing different user behavioural information. We conduct three experiments: (1) To evaluate the effectiveness of the proposed model, we compare it against baseline methods. (2) We experiment using contextual features such as user posting behaviour, user profile and temporal information with text content (e.g., news article) solely as an input to our models; we experiment with two pre-trained embedding models, namely, BERT_{base} and GloVe. (3) We experimented using merely content-based features (i.e., features extracted from text input such as sentiment, morality, and other linguistic-based clues), with original text content (e.g., news article) as an input to our models, with the above-mentioned pre-trained word representation models.

4.2. Comparison of Fake News Detection Methods

4.2.1. Baseline Methods

In this study, we conduct a comparative analysis between the proposed method and state-of-the-art algorithms across multiple datasets. Our primary objective is to assess the predictive efficacy of our model and the utility of the features introduced for detecting fake news. The comparative evaluation encompasses a range of methods proposed for fake news detection and general text classification models:

1. **mCNN [9]:** a model consisting of multiple convolution filters to capture different granularity from text data (e.g., a news article). We use BERT as an encoding model; thus, we call this baseline BERT_{base}-mCNN.
2. **SAF [52]:** a model which uses the FakeNewsNet dataset that integrates social user activities-related features with linguistic-based features. The results reported here are adopted from [52].

3. **BiLSTM-BERT [53]**: a natural language inference approach is used to determine the truthfulness of news using the PolitiFact dataset. This method utilises BiLSTM and BERT embeddings. The results reported here are adopted from [53].
4. **LNN-KG [1]**: a neural network model was applied to the PolitiFact dataset, which used different representations for both the textual patterns and embeddings of concepts found in the input text. The results reported here are adopted from [1].
5. **Multinomial Naive Bayes [2]**: a hybrid model leveraging the FA-KES dataset integrates features derived from both textual content and metadata associated with news articles to discern fake news. The results reported here are adopted from [2].
6. **Hybrid CNN-RNN [3]**: a deep learning model that utilises both CNN and RNN in a hybrid approach for the classification of fake news using the FA-KES dataset. The results reported here are adopted from [3].
7. **Naive Bayes (NB) [4]**: a probabilistic model used to detect fake from real news on the LIAR dataset. The results reported here are adopted from [4].
8. **Support Vector Machine (SVM) [2]**: a popular classifier to detect fake from real news on the LIAR dataset. The results reported here are adopted from [2].
9. **BERT_{base}-BiGRU(Att)**: a stacked BiGRU architecture is employed, incorporating multiple BiGRU layers operating for an identical number of time steps, followed by an attention layer. This configuration enables the model to capture bidirectional dependencies, encompassing both forward and backward contexts, thereby enhancing effectiveness compared to unidirectional GRU models.
10. **BERT_{base}-text-sBiLSTM**: this model focuses solely on textual content, disregarding user posting features and other auxiliary factors. By solely analysing the text, the model may overlook the nuances where certain portions of the text are true but are used to bolster false claims. Incorporating BiLSTM provides an advantage as it enables a thorough examination of the input text, encompassing both preceding and subsequent events, thereby enhancing the model's ability to discern the veracity of the content [4].
11. **BERT_{base}-CNN-sBiGRU**: this framework employs a combination of BERT-CNN for encoding text representations and stacked BiGRU layers for modelling additional auxiliary features. Subsequently, the outputs from both models are merged, followed by the application of a Sigmoid layer.

4.2.2. Results and Discussion

Based on the results presented in Tables 5–8, the proposed model consistently outperforms the state-of-the-art results by notable percentages: \uparrow 3.59% (PolitiFact), \uparrow 6.8% (GossipCop), \uparrow 2.96% (FA-KES), and \uparrow 12.51% (LIAR), considering both content-based features and additional auxiliary information. For the FakeNewsNet dataset, the proposed model exhibits superior performance compared to the models presented in [52]. The referenced model in [52] considered various feature sets learned jointly with news article content, providing significant contextual knowledge to the model. The improved performance of the sBiGRU(Att) model can be attributed to its ability to select the most salient parts of a sequence, facilitated by the attention mechanism. In contrast, the text-sBiLSTM model shows suboptimal performance across all datasets, especially on the LIAR dataset. Consistent with [15], we acknowledge that BiLSTM is more prone to overfitting on the LIAR dataset, leading to underperformance. Additionally, the mCNN model performs poorly on the same dataset, indicating vulnerability to overfitting. Figure 2a illustrates the training and validation loss and accuracy on the LIAR dataset using BERT_{base}-text-sBiLSTM, while Figure 2b depicts those of BERT_{base}-mCNN. Similar insights can be observed in the FA-KES dataset. Surprisingly, the CNN-sBiGRU model achieves the lowest F1 score, particularly on the LIAR dataset, suggesting its incapacity to capture useful patterns for detecting fake content. However, the same network ranks as the second-best performing model on FA-KES. In contrast, the proposed model consistently yields the best results across datasets, demonstrating its capability to capture dataset intricacies effectively. Models integrating

metadata, content-based cues, and news content representation consistently outperform those relying solely on context-based cues (metadata) or content-based clues. We can see that $BERT_{base}\text{-mCNN-sBiGRU}_{all} > BERT_{base}\text{-mCNN-sBiGRU}_{metadata} > BERT_{base}\text{-mCNN-sBiGRU}_{content}$. The results can be seen in the tables below. Numbers in bold refer to the highest scores achieved among all the implemented algorithms. For brevity, we exclude the outcomes achieved by GloVe when utilising the fusion of content- and context-based features. It should be noted that A%, P%, R%, and F1% denote accuracy, precision, recall, and F1 score, respectively.



Figure 2. LIAR training and validation accuracy and loss graphs using (a) $BERT_{base}\text{-text-sBiLSTM}$ and (b) $BERT_{base}\text{-mCNN}$ models.

Based on the experimental results, we have the following observations.

1. The interplay between the distinct properties of metadata and news content uncovers more patterns for machine learning models to identify, ultimately leading to enhanced detection performance. This seems to confirm the hypothesis that using both content- and context-based clues can improve the detection performance. More details on assessing impacts of selecting useful features can be found in Section 4.5.
2. The consideration of solely news content features for detecting fake news yielded sub-par results, underscoring the importance of behavioural information in distinguishing between fake and genuine news articles.
3. Using context-aware embedding models like $BERT_{base}$ has demonstrated exceptional performance compared to off-the-shelf context-independent pre-trained models like GloVe. This highlights the superiority of semantically contextual representations over context-independent embedding methods, albeit with the caveat of their higher computational complexity.

Table 5. A comparison (%) of detection effectiveness on the PolitiFact dataset. The best performance scores are bolded.

Model	A (%)	P (%)	R (%)	F1 (%)
SAF	0.691	0.638	0.789	0.706
BiLSTM-BERT	0.885	NA	NA	NA
LNN-KG	0.880	0.9011	0.880	0.8892
$BERT_{base}\text{-BiGRU(Att)}$	0.9137	0.9722	0.8750	0.9211
$BERT_{base}\text{-text-sBiLSTM}$	0.8705	0.8760	0.7750	0.8732
$BERT_{base}\text{-mCNN}$	0.8849	0.9571	0.8375	0.8933
$BERT_{base}\text{-CNN-sBiGRU}$	0.8921	0.9012	0.9125	0.9068
$BERT_{base}\text{-mCNN-sBiGRU (ours)}$	0.9209	0.9600	0.9000	0.9290

Table 6. A comparison (%) of detection effectiveness on the GossipCop dataset. The best performance scores are bolded.

Model	A (%)	P (%)	R (%)	F1 (%)
SAF	0.796	0.820	0.753	0.785
BERT _{base} -BiGRU(Att)	0.8450	0.8892	0.9101	0.8996
BERT _{base} -text-sBiLSTM	0.8389	0.8668	0.9319	0.8982
BERT _{base} -mCNN	0.8086	0.8906	0.8540	0.8719
BERT _{base} -CNN-sBiGRU	0.8555	0.8942	0.9193	0.9065
BERT _{base} -mCNN-sBiGRU (ours)	0.8640	0.8731	0.9614	0.9151

Table 7. A comparison (%) of detection effectiveness on the FA-KES dataset. The best performance scores are bolded.

Model	A (%)	P (%)	R (%)	F1 (%)
Multinomial Naive Bayes	0.5809	0.63	0.58	0.50
Hybrid CNN-RNN	0.60	0.59	0.60	0.59
BERT _{base} -BiGRU(Att)	0.5679	0.5890	0.8958	0.7107
BERT _{base} -text-sBiLSTM	0.5062	0.5769	0.6250	0.6000
BERT _{base} -mCNN	0.5280	0.5823	0.5169	0.5476
BERT _{base} -CNN-sBiGRU	0.5404	0.5547	0.8539	0.6726
BERT _{base} -mCNN-sBiGRU (ours)	0.6296	0.6552	0.7917	0.7170

Table 8. A comparison (%) of detection effectiveness on the LIAR dataset. The best performance scores are bolded.

Model	A (%)	P (%)	R (%)	F1 (%)
Naive Bayes	0.60	0.59	0.60	0.59
SVM	0.62	NA	NA	NA
BERT _{base} -BiGRU(Att)	0.7411	0.7487	0.8137	0.7799
BERT _{base} -text-sBiLSTM	0.6196	0.6391	0.7465	0.6886
BERT _{base} -mCNN	0.6030	0.6087	0.8277	0.7015
BERT _{base} -CNN-sBiGRU	0.7277	0.7524	0.7703	0.7612
BERT _{base} -mCNN-sBiGRU (ours)	0.7451	0.7576	0.8053	0.7807

4.3. Prediction Performance of BERT vs. GloVe Using Solely Context-Based Features

In this section, we assess the effectiveness of the proposed model and different baseline models (i.e., CNN-BiGRU and sBiGRU(Att)) on different datasets using solely basic contextual features such as user posting behaviour in FakeNewsNet (Tables 9 and 10), spatiotemporal features in the FA-KES dataset (Table 11) and the user profile in the LIAR dataset (Table 12). The investigation found that using mCNN coupled with stacked BiGRU (the proposed approach) achieved the best accuracy. The results are tabulated in the tables below, highlighting the best scores in bold. We have achieved better accuracy using pre-trained BERT_{base} embeddings than the off-the-shelf pre-trained GloVe. The analysis found that the attention mechanism boosts the detection performance, where the attention-based classifier (sBiGRU(Att)) yields comparable results to the proposed approach. This could be attributed to the fact that placing a stacked BiGRU on the input text representations results in more semantic representations that were harnessed by extracting both past and future contexts. More than that, such semantic representations might be improved using the attention layer on the output of the stacked BiGRU layers, leading to more accurate results. This model (sBiGRU(Att)) is found to be the best-performing model for GossipCop and LIAR datasets and the second-best among all the models. For the baseline CNN-sBiGRU, we found it to be the second-best model for FA-KES. Moreover, it has been shown that the proposed framework achieved the best accuracy compared to all baselines.

Exploiting user behaviour would give the model extra contextual knowledge, resulting in good detection performance. It should be noted that we conducted further experiments

to test the contextual features in FakeNewsNet, and as a result, considering features such as followers count, favourites count, retweets count, and verified tweets count only achieved 92% F1 score on PolitiFact, showed better results than using all features. The opposite is true for GossipCop data, where using all users posting behavioural information yields the best results with a 91% F1 score. Readers are referred to Section 4.5 for more details.

It is important to acknowledge that the imbalanced class distribution within the FakeNewsNet dataset could potentially impact the performance of the model. The inherent skewness in the number of fake and real news instances may lead to a model biased towards the majority class. We anticipate that by equalising the representation of fake and real news instances in such a dataset, the model will be better equipped to learn from both classes, thereby mitigating potential bias and improving its ability to generalise to a wider range of scenarios. Future iterations of our research will explore and implement class balancing techniques to ensure a fair and robust evaluation of our fake news detection model.

Moreover, the incorporation of user behavioural information indeed introduced heightened complexity to the model architecture. The utilisation of the stacked BiGRU layer, chosen for its effectiveness in capturing temporal dependencies in user behaviour, contributed to increased computational overhead and extended training times. As part of the future work, we aim to explore innovative approaches to further streamline the model, investigating techniques such as model quantization and pruning to reduce computational demands without compromising the model's ability to capture nuanced user behaviour.

Table 9. Performance comparison (%) of contextual features using (a) $BERT_{base}$ and (b) *GloVe* on the PolitiFact dataset. The best performance scores are bolded.

a		b	
	A (%)		A (%)
sBiGRU(Att)	0.8561	sBiGRU(Att)	0.8273
CNN-sBiGRU	0.8777	CNN-sBiGRU	0.8489
mCNN-sBiGRU	0.9137	mCNN-sBiGRU	0.8201

Table 10. Performance comparison (%) of contextual features using (a) $BERT_{base}$ and (b) *GloVe* on the GossipCop dataset. The best performance scores are bolded.

a		b	
	A (%)		A (%)
sBiGRU(Att)	0.8578	sBiGRU(Att)	0.8476
CNN-sBiGRU	0.8349	CNN-sBiGRU	0.8121
mCNN-sBiGRU	0.8560	mCNN-sBiGRU	0.8420

Table 11. Performance comparison (%) of contextual features using (a) $BERT_{base}$ and (b) *GloVe* on the FA-KES dataset. The best performance scores are bolded.

a		b	
	A (%)		A (%)
sBiGRU(Att)	0.4968	sBiGRU(Att)	0.4596
CNN-sBiGRU	0.5217	CNN-sBiGRU	0.4347
mCNN-sBiGRU	0.5590	mCNN-sBiGRU	0.5465

Table 12. Performance comparison (%) of contextual features using (a) $BERT_{base}$ and (b) *GloVe* on the LIAR dataset. The best performance scores are bolded.

a		b	
	A (%)		A (%)
sBiGRU(Att)	0.7332	sBiGRU(Att)	0.7277
CNN-sBiGRU	0.7205	CNN-sBiGRU	0.7127
mCNN-sBiGRU	0.7285	mCNN-sBiGRU	0.7158

4.4. Prediction Performance of BERT vs. GloVe Using Solely Various Content-Based Features

The tables (Tables 13–16) below show how deep learning models perform when relying merely on content-based features. The proposed model performs the best among all baselines (across all datasets), with stacked BiGRU(Att) being the first-best model on the LIAR dataset. It is noticeable that the proposed framework utilising merely user behavioural patterns outperforms these models that rely on just content-based features on all datasets. The stacked BiGRU(Att) model continues its promising results, while the CNN-sBiGRU network performs poorly using solely content-based features on LIAR and FA-KES datasets.

Table 13. Performance comparison (%) of content-based features using (a) $BERT_{base}$ and (b) *GloVe* on the PolitiFact dataset. The best performance scores are bolded.

a		b	
	A (%)		A (%)
sBiGRU(Att)	0.8993	sBiGRU(Att)	0.7554
CNN-sBiGRU	0.8705	CNN-sBiGRU	0.8633
mCNN-sBiGRU	0.9065	mCNN-sBiGRU	0.8849

Table 14. Performance comparison (%) of content-based features using (a) $BERT_{base}$ and (b) *GloVe* on the GossipCop dataset. The best performance scores are bolded.

a		b	
	A (%)		A (%)
sBiGRU(Att)	0.8445	sBiGRU(Att)	0.8340
CNN-sBiGRU	0.8480	CNN-sBiGRU	0.8252
mCNN-sBiGRU	0.8509	mCNN-sBiGRU	0.8263

Table 15. Performance comparison (%) of content-based features using (a) $BERT_{base}$ and (b) *GloVe* on the FA-KES dataset. The best performance scores are bolded.

a		b	
	A (%)		A (%)
sBiGRU(Att)	0.4658	sBiGRU(Att)	0.4783
CNN-sBiGRU	0.5093	CNN-sBiGRU	0.4907
mCNN-sBiGRU	0.5404	mCNN-sBiGRU	0.4410

Table 16. Performance comparison (%) of content-based features using (a) $BERT_{base}$ and (b) *GloVe* on the LIAR dataset. The best performance scores are bolded.

a		b	
	A (%)		A (%)
sBiGRU(Att)	0.6148	sBiGRU(Att)	0.5643
CNN-sBiGRU	0.5643	CNN-sBiGRU	0.5746
mCNN-sBiGRU	0.6046	mCNN-sBiGRU	0.5817

Table 17 gives insights into the computational efficiency of the proposed model for different datasets.

Table 17. Training and inference time of the proposed model in seconds.

Dataset	Training time	Inference Time
PolitiFact	79.81	9.64
GossipCop	1356.94	81.98
FA-KES	153.54	10.65
LIAR	788.57	20.52

4.5. Assessing Impacts of Selecting Useful Features

In this section, we aim to thoroughly evaluate the efficacy of the proposed model. Additionally, we plan to conduct a case study to examine the model's performance with various features. To achieve this, we feed the model with these features separately and compare their performance. According to a study by [54], false political news on Twitter tends to be retweeted by a large number of users and spreads very quickly. The authors in [55] also suggest that the dissemination of fake news on social media starts with the behaviour of the user who posts the news. As such, (1) we analyse the FakeNewsNet dataset by investigating the role played by each user posting behavioural features; and (2) we also analyse the performance of each language feature individually using all datasets. The results of these experiments are shown in the tables below.

For the first experiment, we evaluate the performance of user posting behaviour using the proposed model (BERT_{base}-mCNN-sBiGRU). We test different sets of feature combinations. Table 18 shows the performance of the proposed model using news article content and user posting behaviour on Twitter for the PolitiFact dataset. Similarly, Table 19 shows that of the GossipCop dataset. Using merely followers count, favourite count, retweets count and verified tweets features achieves a significantly better performance than using other features on the PolitiFact dataset. On the other hand, considering all features shows better results for the GossipCop dataset. For the second experiment (See Table 20), first, we run the model using solely sentimental features, which shows that the model did not perform well when considering merely such features across all datasets. Interestingly, the model performs even worse when considering the combination of sentimental and morality-based features.

Similar observations can be applied when using only morality-based features. In contrast, the model performs well when considering only the combination of linguistic and sentiment cues on PolitiFact, while considering merely sentiment clues yields better results on the GossipCop dataset. The combination of linguistic and morality-based features yields good performance on the FA-KES dataset; thus, identifying language indicators from user-generated information is critical for detecting fake news. It is noticeable that considering all features together further increases the performance of fake news detection. We deduce that content- and context-based features provide complementary information towards improving fake news detection.

To this end, the research methodology is strategically designed to stay adaptive and robust to the dynamic nature of fake news and the evolving landscape of online content. Recognising the limitations of static approaches, we incorporate advanced contextual embeddings, specifically leveraging pre-trained models like BERT, to capture the ever-changing semantics of news articles. The proposed framework places a strong emphasis on multi-aspect language representations, acknowledging the diverse forms fake news can take. Furthermore, the inclusion of user behavioural information adds a layer of adaptability, as user engagement patterns continually shift over time. The hybrid nature of the proposed framework, encompassing pre-trained embeddings, multi-channel CNN, and stacked BiGRU, allows the model to dynamically learn from multiple sources, ensuring its efficacy in the face of evolving content structures. Through extensive experimentation on diverse real-world datasets, we evaluate the performance of the proposed framework

across various scenarios, thereby affirming its adaptability and effectiveness in addressing the challenges posed by the dynamic nature of misinformation in online content.

Table 18. Evaluating (%) the effectiveness of selecting useful features using $BERT_{base}$ -mCNN-sBiGRU on the PolitiFact dataset. Note that **A**, **P**, **R**, and **F1**, refer to accuracy, precision, recall, and F1 score, respectively. The best performance scores are bolded.

Features	Metrics			
	A (%)	P (%)	R (%)	F1 (%)
+title	0.8345	0.8353	0.8875	0.8606
+followers	0.7698	0.7553	0.8875	0.8161
+friends	0.7482	0.7473	0.8500	0.7953
+favorites	0.8273	0.8333	0.8750	0.8537
+retweets	0.8058	0.7849	0.9125	0.8439
+statuses	0.8489	0.8831	0.8500	0.86625
+verified_count	0.7842	0.8571	0.7500	0.8000
+title+followers	0.8561	0.8750	0.8750	0.8750
+title+friends	0.8129	0.7879	0.8125	0.8000
+title+favorites	0.7914	0.7215	0.8906	0.7972
+title+retweets	0.8201	0.7826	0.8438	0.8120
+title+statuses	0.8417	0.7692	0.9375	0.8451
+title+verified_count	0.8633	0.8689	0.8281	0.8480
+followers+friends	0.7986	0.8600	0.6719	0.7544
+followers+favorites	0.7770	0.7089	0.8750	0.7832
+followers+retweets	0.7626	0.8605	0.5781	0.6916
+followers+statuses	0.8129	0.8065	0.7812	0.7937
+followers+verified_count	0.8273	0.7632	0.9062	0.8286
+friends+favorites	0.6978	0.6146	0.9219	0.7375
+friends+retweets	0.7122	0.6224	0.9531	0.7531
+friends+statuses	0.7842	0.7361	0.8281	0.7794
+friends+verified_count	0.8273	0.8030	0.8281	0.8154
+favorites+retweets	0.7985	0.7432	0.8594	0.7971
+favorites+statuses	0.7986	0.7308	0.8906	0.8028
+favorites+verified_count	0.8489	0.9388	0.7188	0.8142
+retweets+statuses	0.8345	0.8727	0.7500	0.8067
+retweets+verified_count	0.8561	0.8143	0.8906	0.8507
+statuses+verified_count	0.8057	0.7079	0.9844	0.8235
+title+followers+friends	0.8417	0.7917	0.8906	0.8382
+title+followers+favorites	0.8201	0.7532	0.9062	0.8227
+title+followers+retweets	0.8489	0.8413	0.8281	0.8346
+title+followers+statuses	0.8489	0.8644	0.7969	0.8293
+title+followers+verified_count	0.8633	0.8082	0.9219	0.8613
+title+followers+friends+favorites	0.8345	0.8361	0.7969	0.8160
+title+followers+friends+retweets	0.8058	0.7229	0.9375	0.8163
+title+followers+friends+verified	0.8417	0.8387	0.8125	0.8254
+followers+friends+favorites	0.7914	0.7869	0.7500	0.7680
+favorites+friends+verified	0.8849	0.8636	0.8906	0.8769
+followers+favorites+retweets+verified	0.9137	0.9595	0.8875	0.9221

Table 19. Evaluating (%) the effectiveness of selecting useful features using $BERT_{base}$ -mCNN-sBiGRU on GossipCop dataset. Note that **A**, **P**, **R**, and **F1**, refer to accuracy, precision, recall, and F1 score, respectively. The best performance scores are bolded.

Features	Metrics			
	A (%)	P (%)	R (%)	F1 (%)
+title	0.8440	0.8425	0.9782	0.9053
+followers	0.8124	0.8482	0.9182	0.8818
+friends	0.8472	0.8591	0.9565	0.9052

Table 19. Cont.

Features	Metrics			
	A (%)	P (%)	R (%)	F1 (%)
+favorites	0.8314	0.8594	0.9312	0.8939
+retweets	0.8279	0.8330	0.9684	0.8956
+statuses	0.8381	0.8321	0.9867	0.9028
+verified_count	0.8504	0.8605	0.9593	0.9072
+title+followers	0.8191	0.8616	0.9087	0.8845
+title+friends	0.8431	0.8528	0.9600	0.9032
+title+favorites	0.8220	0.8900	0.8747	0.8823
+title+retweets	0.8522	0.8740	0.9421	0.9068
+title+statuses	0.8456	0.8606	0.9516	0.9038
+title+verified_count	0.8375	0.8920	0.8954	0.8937
+title+followers+friends+retweets	0.8466	0.8530	0.9653	0.9056
+title+followers+friends+statuses	0.8308	0.8431	0.9561	0.8961
+title+followers+friends+verified	0.8469	0.8508	0.9691	0.9061
+followers+friends+favorites	0.8448	0.8475	0.9712	0.9051
+followers+friends	0.8410	0.8405	0.9768	0.9036
+followers+favorites	0.8386	0.8706	0.9259	0.8974
+followers+retweets	0.8453	0.8690	0.9386	0.9025
+followers+statuses	0.8426	0.8512	0.9617	0.9031
+followers+verified_count	0.8461	0.8668	0.9431	0.9033
+friends+retweets	0.8319	0.8664	0.9217	0.8932
+friends+statuses	0.8311	0.8314	0.9765	0.8981
+friends+verified_count	0.8429	0.8733	0.9287	0.9002
+favorites+retweets	0.8386	0.8611	0.9400	0.8988
+favorites+verified_count	0.8541	0.8810	0.9351	0.9072
+retweets+statuses	0.8469	0.8548	0.9628	0.9056
+retweets+verified_count	0.8405	0.8666	0.9347	0.8994
+title+followers+friends	0.8324	0.8313	0.9789	0.8991
+title+followers+retweets	0.8512	0.8531	0.9723	0.9088
+title+followers+verified_count	0.8539	0.8764	0.9410	0.9076
+title+followers+friends+favorites	0.8453	0.8683	0.9396	0.9026
+favorites+friends+verified	0.8423	0.8615	0.9452	0.9014
+retweets+friends+verified	0.8472	0.8679	0.9431	0.9040
+retweets+statuses+verified	0.8480	0.8726	0.9375	0.9039
+followers+favorites+friends+verified	0.8330	0.8433	0.9593	0.8975
+follow+fav+fri+retw+stat+verif	0.8560	0.8556	0.9758	0.9118
+followers+favorites+retweets+verified	0.8549	0.8798	0.9379	0.9079

Table 20. Evaluating (%) the effectiveness of selecting useful features. Note that **P**, **G**, **F**, and **L**, respectively, refer to PolitiFact, GossipCop, FA-KES, and LIAR datasets using BERT_{base}-mCNN-sBiGRU. The best performance scores are bolded.

Model	Features	Datasets			
		P A (%)	G A (%)	F A (%)	L A (%)
Our model	+ sentiment	0.8561	0.8565	0.4720	0.7245
	+ morality	0.8345	0.8474	0.5031	0.7332
	+ sentiment and morality	0.8058	0.8399	0.5403	0.7088
	+ linguistic	0.8417	0.8520	0.4968	0.7301
	+ linguistic and sentiment	0.8633	0.8386	0.5155	0.7380
	+ linguistic and morality	0.8273	0.8526	0.5714	0.7167
	+ all	0.9209	0.8640	0.6296	0.7451

4.6. Limitations

While the proposed approach demonstrates promising results in fake news detection, several limitations need to be acknowledged. Firstly, the effectiveness of the proposed model may vary across different datasets, as it heavily relies on the quality and diversity of the training data. Additionally, the computational resources required for training and testing the model may pose practical constraints, particularly for researchers with limited access to high-performance computing facilities. Lastly, the generalisation of the findings to other languages or domains beyond the scope of the study remains uncertain and requires further investigation. Addressing these limitations would be essential for enhancing the robustness and applicability of the proposed approach in real-world scenarios.

Note

The data can be downloaded from the links provided in Section 4.1.

5. Conclusion

We have proposed a novel deep learning framework that integrates additional auxiliary feature representations together with news content representation. The interplay between user posting behaviour, temporal or user profile cues, and model training helps uncover informative features that distinguish falsified from genuine information. Enriching such features with extra valuable content-based knowledge significantly improves the detection performance. Experiments on four real-world datasets demonstrate the effectiveness of our proposed framework. Our experiments demonstrate that not only does BERT_{base}-mCNN-sBiGRU perform well, but individual components of it also outperform comparative methods. BERT offers a good advantage to improve performance but at the expense of merely considering sequences within a certain length. That is, BERT imposes a 512-token length limit where longer sentences are simply truncated, resulting in the loss of some important information. One way to overcome this is an effective summarising method, which can be applied so as to shrink the sequence to a length that is somewhat equivalent to or less than the token length limit imposed by BERT. Moreover, using knowledge-based networks to identify the veracity of news by checking the facts would give extraordinary and correct fake news classification results. Furthermore, future endeavours should consider adapting and refining these indicators to suit the specific linguistic and cultural nuances of diverse languages. This nuanced approach aims to enhance the framework's robustness and applicability in capturing each language's unique characteristics, thereby contributing to the development of more effective language-specific indicators for global fake news detection. Furthermore, although conducting a statistical analysis is a commendable idea, it presents significant challenges due to issues such as data imbalance and algorithmic randomness. Overcoming these challenges requires a robust statistical model capable of providing insights across all approaches. However, given the complexity and scope of this task, it exceeds the current paper's capacity, and therefore, we intend to explore it as part of our future work. These are some directions we will consider for our future research.

Author Contributions: Conceptualisation, algorithm developing, experiments and formal analysis, J.A.; paper writing, J.A.; writing, review and editing, analysis and interpretation of the experimental data and supervision, Y.L. and S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The FakeNewsNet data is available from <https://github.com/KaiDMML/FakeNewsNet> (accessed on 1 January 2023); LIAR data is available from <https://www.cs.ucsb.edu/william/data/liardataset.zip> (accessed on 19 February 2023) and FA-KES is available from <https://zenodo.org/record/2607278#.X3oK8WgzaUk> (accessed on 2 February 2023)).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Koloski, B.; Stepišnik-Perdih, T.; Robnik-Šikonja, M.; Pollak, S.; Škrj, B. Knowledge Graph informed Fake News Classification via Heterogeneous Representation Ensembles. *arXiv* **2021**, arXiv:2110.10457.
- Elhadad, M.K.; Li, K.F.; Gebali, F. A Novel Approach for Selecting Hybrid Features from Online News Textual Metadata for Fake News Detection. In Proceedings of the Advances on P2P, Parallel, Grid, Cloud and Internet Computing, Antwerp, Belgium, 7–9 November 2019; Barolli, L.; Hellinckx, P.; Natwichai, J., Eds., Springer: Cham, Switzerland, 2020; pp. 914–925.
- Nasir, J.A.; Khan, O.S.; Varlamis, I. Fake news detection: A hybrid CNN-RNN based deep learning approach. *Int. J. Inf. Manag. Data Insights* **2021**, *1*, 100007. [\[CrossRef\]](#)
- Khan, J.Y.; Khondaker, M.T.I.; Afroz, S.; Uddin, G.; Iqbal, A. A benchmark study of machine learning models for online fake news detection. *Mach. Learn. Appl.* **2021**, *4*, 100032. [\[CrossRef\]](#)
- Alghamdi, J.; Lin, Y.; Luo, S. A Comparative Study of Machine Learning and Deep Learning Techniques for Fake News Detection. *Information* **2022**, *13*, 576. [\[CrossRef\]](#)
- Shu, K.; Cui, L.; Wang, S.; Lee, D.; Liu, H. DEFEND: Explainable Fake News Detection. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 395–405.
- Pennington, J.; Socher, R.; Manning, C.D. Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 25–29 October 2014; pp. 1532–1543.
- Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
- Kim, Y. Convolutional Neural Networks for Sentence Classification. *arXiv* **2014**. [\[CrossRef\]](#)
- Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv* **2014**, arXiv:1406.1078.
- Zhou, X.; Zafarani, R. A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Comput. Surv. (CSUR)* **2020**, *53*, 1–40. [\[CrossRef\]](#)
- Zhou, X.; Jain, A.; Phoha, V.V.; Zafarani, R. Fake News Early Detection: An Interdisciplinary Study. *arXiv* **2019**, arXiv:1904.11679.
- Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. *arXiv* **2013**, arXiv:1301.3781.
- Le, Q.; Mikolov, T. Distributed representations of sentences and documents. In Proceedings of the International Conference on Machine Learning, Beijing, China, 21–26 June 2014; pp. 1188–1196.
- Wang, W.Y. “Liar, Liar Pants on Fire”: A New Benchmark Dataset for Fake News Detection. *arXiv* **2017**, arXiv:1705.00648.
- Kaliyar, R.K.; Goswami, A.; Narang, P. FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimed. Tools Appl.* **2021**, *80*, 11765–11788. [\[CrossRef\]](#)
- Alghamdi, J.; Lin, Y.; Luo, S. Modeling Fake News Detection Using BERT-CNN-BiLSTM Architecture. In Proceedings of the 2022 IEEE 5th International Conference on Multimedia Information Processing and Retrieval (MIPR), Online, 2–4 August 2022; pp. 354–357. [\[CrossRef\]](#)
- Xia, H.; Wang, Y.; Zhang, J.Z.; Zheng, L.J.; Kamal, M.M.; Arya, V. COVID-19 fake news detection: A hybrid CNN-BiLSTM-AM model. *Technol. Forecast. Soc. Chang.* **2023**, *195*, 122746. [\[CrossRef\]](#)
- Zivkovic, M.; Stoean, C.; Petrovic, A.; Bacanin, N.; Strumberger, I.; Zivkovic, T. A Novel Method for COVID-19 Pandemic Information Fake News Detection Based on the Arithmetic Optimization Algorithm. In Proceedings of the 2021 23rd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), Timisoara, Romania, 7–10 December 2021; pp. 259–266. [\[CrossRef\]](#)
- Shu, K.; Zhou, X.; Wang, S.; Zafarani, R.; Liu, H. The role of user profiles for fake news detection. In Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Vancouver, BC, Canada, 27–30 August 2019; pp. 436–439.
- Vosoughi, S. Automatic Detection and Verification of Rumors on Twitter. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2015.
- Ma, J.; Gao, W.; Mitra, P.; Kwon, S.; Jansen, B.J.; Wong, K.F.; Cha, M. Detecting Rumors from Microblogs with Recurrent Neural Networks. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, New York, NY, USA, 9–15 July 2016; pp. 3818–3824.
- Chen, W.; Yeo, C.K.; Lau, C.T.; Lee, B.S. Behavior deviation: An anomaly detection view of rumor preemption. In Proceedings of the 2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 13–15 October 2016; pp. 1–7.
- Wu, L.; Liu, H. Tracing Fake-News Footprints: Characterizing Social Media Messages by How They Propagate. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, New York, NY, USA, 5–9 February 2018; pp. 637–645.
- Gupta, M.; Zhao, P.; Han, J. Evaluating event credibility on twitter. In Proceedings of the 2012 SIAM International Conference on Data Mining, Anaheim, CA, USA, 26–28 April 2012; pp. 153–164.

26. Gupta, A.; Lamba, H.; Kumaraguru, P.; Joshi, A. Faking Sandy: Characterizing and Identifying Fake Images on Twitter during Hurricane Sandy. In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 13–17 May 2013; pp. 729–736.
27. Qazvinian, V.; Rosengren, E.; Radev, D.R.; Mei, Q. Rumor has it: Identifying Misinformation in Microblogs. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Edinburgh, UK, 27–31 July 2011; pp. 1589–1599.
28. Zhao, Z.; Resnick, P.; Mei, Q. Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts. In Proceedings of the WWW '15: 24th International World Wide Web Conference, Florence, Italy, 18–22 May 2015; pp. 1395–1405. [[CrossRef](#)]
29. Chua, A.Y.; Banerjee, S. Linguistic predictors of rumor veracity on the internet. In Proceedings of the International MultiConference of Engineers and Computer Scientists, Hong Kong, China, 16–18 March 2016; Volume 1, pp. 387–391.
30. Ma, J.; Gao, W.; Wong, K.F. *Detect Rumors in Microblog Posts using Propagation Structure via Kernel Learning*; Association for Computational Linguistics: Vancouver, BC, Canada, 2017.
31. Kwon, S.; Cha, M.; Jung, K.; Chen, W.; Wang, Y. Prominent features of rumor propagation in online social media. In Proceedings of the 2013 IEEE 13th International Conference on Data Mining, Dallas, TX, USA, 7–10 December 2013; pp. 1103–1108.
32. Kwon, S.; Cha, M.; Jung, K. Rumor Detection over Varying Time Windows. *PLoS ONE* **2017**, *12*, e0168344. [[CrossRef](#)]
33. Zubiaga, A.; Liakata, M.; Procter, R. Exploiting context for rumour detection in social media. In Proceedings of the International Conference on Social Informatics, Oxford, UK, 13–15 September 2017; Springer: Cham, Switzerland, 2017, pp. 109–123.
34. Qin, Y.; Wurzer, D.; Lavrenko, V.; Tang, C. Spotting rumors via novelty detection. *arXiv* **2016**, arXiv:1611.06322.
35. Shu, K.; Wang, S.; Liu, H. Exploiting tri-relationship for fake news detection. *arXiv* **2017**, arXiv:1712.07709.
36. Jin, Z.; Cao, J.; Zhang, Y.; Luo, J. News verification by exploiting conflicting social viewpoints in microblogs. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 14–17 February 2016; Volume 30.
37. Li, Q.; Liu, X.; Fang, R.; Nourbakhsh, A.; Shah, S. User behaviors in newsworthy rumors: A case study of twitter. In Proceedings of the International AAAI Conference on Web and Social Media, Cologne, Germany, 17–20 May 2016; Volume 10.
38. Li, Q.; Zhang, Q.; Si, L. eventAI at SemEval-2019 task 7: Rumor detection on social media by exploiting content, user credibility and propagation information. In Proceedings of the 13th International Workshop on Semantic Evaluation, Minneapolis, MN, USA, 6–7 June 2019; pp. 855–859.
39. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All You Need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Red Hook, NY, USA, 4–9 December 2017; pp. 6000–6010.
40. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
41. Tang, D.; Qin, B.; Liu, T. Document Modeling with Gated Recurrent Neural Network for Sentiment Classification. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, 17–21 September 2015; pp. 1422–1432.
42. Graves, A. Supervised sequence labelling. In *Supervised Sequence Labelling with Recurrent Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 5–13.
43. Ek, A.; Bernardy, J.P.; Chatzikyriakidis, S. How does Punctuation Affect Neural Models in Natural Language Inference. In Proceedings of the Probability and Meaning Conference (PaM 2020), Gothenburg, Sweden, 14–15 October 2020; pp. 109–116.
44. Singh, V.; Dasgupta, R.; Sonagra, D.; Raman, K.; Ghosh, I. Automated fake news detection using linguistic analysis and machine learning. In Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation (SBP-BRiMS), Washington, DC, USA, 5–8 July 2017; pp. 1–3.
45. Stieglitz, S.; Dang-Xuan, L. Emotions and information diffusion in social media—Sentiment of microblogs and sharing behavior. *J. Manag. Inf. Syst.* **2013**, *29*, 217–248. [[CrossRef](#)]
46. Ferrara, E.; Yang, Z. Quantifying the effect of sentiment on information diffusion in social media. *PeerJ Comput. Sci.* **2015**, *1*, e26. [[CrossRef](#)]
47. Mohammad, S.; Turney, P. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, Los Angeles, CA, USA, 5 June 2010; pp. 26–34.
48. Guo, C.; Cao, J.; Zhang, X.; Shu, K.; Yu, M. Exploiting emotions for fake news detection on social media. *arXiv* **2019**, arXiv:1903.01728.
49. Chakraborty, A.; Paranjape, B.; Kakarla, S.; Ganguly, N. Stop clickbait: Detecting and preventing clickbaits in online news media. In Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Davis, CA, USA, 18–21 August 2016; pp. 9–16.
50. Ghanem, B.; Ponzetto, S.P.; Rosso, P.; Rangel, F. Fakeflow: Fake news detection by modeling the flow of affective information. *arXiv* **2021**, arXiv:2101.09810.
51. Shu, K.; Wang, S.; Liu, H. Beyond news contents: The role of social context for fake news detection. In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, Melbourne, Australia, 11–15 February 2019; pp. 312–320.
52. Shu, K.; Mahudeswaran, D.; Wang, S.; Lee, D.; Liu, H. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data* **2020**, *8*, 171–188. [[CrossRef](#)] [[PubMed](#)]
53. Sadeghi, F.; Bidgoly, A.J.; Amirkhani, H. Fake news detection on social media using a natural language inference approach. *Multimed. Tools Appl.* **2022**, *81*, 33801–33821. [[CrossRef](#)]

-
54. Vosoughi, S.; Roy, D.; Aral, S. The spread of true and false news online. *Science* **2018**, *359*, 1146–1151. [[CrossRef](#)]
 55. Wang, S.; Pang, M.S.; Pavlou, P.A. Cure or Poison? Identity Verification and the Posting of Fake News on Social Media. *J. Manag. Inf. Syst.* **2021**, *38*, 1011–1038. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.