IET Information Security

IET The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH PAPER

A hybrid model for fake news detection: Leveraging news content and user comments in fake news

Marwan Albahar 🆔

Department of Computer Science, Umm Al Qura University, Mecca, Saudi Arabia

Correspondence

Marwan Albahar, Department of Computer Science, Umm Al Qura University, Mecca, Saudi Arabia.
Email: mabahar@uqu.edu.sa

Abstract

Nowadays, social media platforms such as Twitter have become a popular medium for people to spread and consume news because of their easy access and the rapid proliferation of news. However, the credibility of the news posted on these platforms has become a significant issue. In other words, written news that contains inaccurate information aiming to mislead readers has been rapidly disseminated on these platforms. In the literature, this news is called fake news. Detecting such news on social media platforms has become a challenging task. One of the main challenges is identifying useful information that is exploited as a way to detect fake news. A hybrid model comprising a recurrent neural network (RNN) and support vector machine (SVM) is incorporated to detect real and fake news. An RNN with bidirectional gated recurrent units was used to encode textual data, including news content and comments, to numerical feature vectors. The encoded features were fed to an SVM with radial basis function kernel to classify the given input of real and fake news. Experiments on the real-world dataset yield encouraging results and demonstrate that the proposed framework outperforms state-of-the-art methods.

1 | INTRODUCTION

Rapid advances in mobile phones and widespread use of the Internet have reformulated social interactions. Because of its particular characteristics—easy access to information, the low cost of generating news, and fast proliferation of news—social media have emerged as an attractive platform for people to disseminate and consume news. Moreover, social media have been shown to capture information related to an ongoing event such as COVID-19 as well as people's interests and opinions. Thus, more people prefer to follow the development of a story or an event on social media than on traditional media such as television or traditional news formats. However, news on social media platforms somehow lacks credibility and reliability compared with news in traditional media. In other words, news published in newspapers is usually considered as coming from trusted sources.

In contrast, the sources for most news posted on social media platforms can be hard to verify, making it easy to manipulate content to achieve various goals [1]. Consequently, such information, disseminated promptly and broadly, can be used to propagate inaccurate news articles [2]. With the absence of rigorous verification, well-meaning users can unintentionally contribute to the fast spread of fake news. The rapid dissemination of fake news can negatively impact society and individuals. Its damage may extend to companies and governments. For example, fake news about an organisation could be spread by malicious users, or spam could cause significant damage to the organisation's image in society. Hence, the detection of fake news has arisen as an important research area. Social media's emergence implies that real and false news is presented similarly and is sometimes difficult to distinguish. First, the content of fake news is combined with both real and false data to draw readers. Second, vast and varied social media information is disseminated—for example, many anonymous users communicate noisy information. The latest research on fake news identification through deep learning algorithms has achieved impressive success [3,4] using numerous social media news features, such as text information, user features, and user feedback. Context learning, however, has not been designed for fake news. Specifically, it is not possible to obtain accurate fake news prediction solely based on textual content [5,6]

because social media content is typically short and fragmented. Second, another study on fake news detection focussed on analysing first news spreader features, thus ignoring the opinions of subsequent user comments [7,8]. It is hard to distinguish fake news from within a massive piece of news. This question leads one to consider different approaches to determine the validity of a news story. In recent years, false news identification process has continued to grow slowly on an evolving research path. We define two key characteristics related to identifying false news, namely, user responses and information content.

The content of the information distributed is primarily what must be classified as true or false. By analysing articles' textual characteristics from different fake news pages, Horne and Adali [9] identified unique features of fake news that differ from true news content and then contrasted those features with articles from credible journalistic websites. Their results indicate that fake news stories are longer, with more capitalised phrases and fewer stop words. Perez-Rosas et al. [10] found that false news articles had more social terms, verbal words, and temporal words, implying that the text tended towards both the present and the future and not on becoming more factual and objective. Newman et al. [11] found that deceptive stories had lower semantic sophistication, fewer phrases, more pessimistic words about feelings, and more words about motion. Silverman [12] notes that 13% of 1600 news articles had incoherent headlines and content, using declaratory headlines coupled with article bodies that were sceptical about the claim.

User responses on social media have auxiliary data that is quite useful in the analysis of false news. They are considered to have stronger identification signals than the content of the information, primarily because user reactions and dissemination patterns are more challenging to exploit than the content of the information and contain obvious veracity data [13]. This secondary knowledge in the form of user interactions (likes, comments, responses, or shares) includes rich information collected in the propagation structure (tree) showing the direction of information flow, timestamp details of interactions, textual information of user interactions, and user profile information of the users engaging in interactions. Zubiaga et al. [14] stated that through their approach, user comments can be differentiated. Four forms are included in the most widely used categorisation: help, reject, question, and comment (which can be neutral or unrelated). They also noted that, based on the phases of dissemination, the essence of user responses differs. In the case of rumours, when a rumour's whole lifespan is considered, most users accept real rumours, and a higher percentage reject fake rumours. Qian et al. [15] found that false news appears to obtain more negative responses and questioning than real news. It was found that users appear to endorse rumours independently of rumour credibility by analysing only the early reactions to rumours. This is apparent because users have trouble assessing integrity in the early stages. However, in practice, most of these methods require an analysis of integrity statements by trained professionals, making it difficult to automate as well as lacking

generalisability over different topics and domains. Below, we list the type of existing methods and their limitations in Table 1. To deal with the above-mentioned limitations of the existing methods, we address the problem of fake news detection in social media by developing a hybrid model. The model incorporates both news content and the potential information in comments and consists of two phases. The embedding process was carried out using bidirectional gated recurrent units (GRUs) and a support vector machine (SVM) with a Gaussian kernel for classification.

2 | RELATED WORKS

While the issue of fake news identification is relatively recent, it has attracted considerable attention. Numerous methods to detect fake news within various datasets have been proposed. This section provides a summary of the current and relevant literature on fake news identification. To identify fake news, there are currently three approaches: propagation-based, source analysis, and content-based research. The propagation-based approach postulates that the distribution patterns of fake news are different from those of trustworthy news. According to the propagation map, these distribution patterns will include news as false or true [17]. The second approach evaluates the news piece's context and patterns, allowing early detection of false news [18]. The content-based research approach relies on the detection of both lexical and syntactic linguistic characteristics. It assumes that fake articles are formulated using syntactic and deceptive language [10,19,20,21,22]. The combination of a multilayer perceptron representation and handcrafted features from the FNC-1 dataset was proposed in [23] for stance detection. The headline and body of each article are encoded with skip-thought vectors. The handmade features involve grams, char-grams, and weighted TF-IDF points between each article and headline body. The authors in [24] proposed the use of neural attention with bidirectional recurrent neural networks (RNNs) for coding an entire news article, the first two sentences of the article, and the article's headline. These representations are then paired with handmade characteristics, as used in [23]. Rubin et al. [25] divided the issue of fake news identification into three categories. These categories are large-scale hoaxes, humorous fake news, and severe fabrication. In [26], Conroy et al. proposed a hybrid approach to detecting false news. Their hybrid approach incorporates network analysis approaches and linguistic cues. For verification of the news, a vector space model was used in [27]. The authors utilised online knowledge to detect misleading information. Dadgar et al. [28] divided news with TF-IDF and SVM into separate categories. In [20], the authors referenced satirical cues to detect misleading or fake news. The SVM-based model evaluated a dataset containing 360 news articles. In [29], Jin et al. established opposing views on social media to cross-check news and validate their concept across a real-world dataset. There has been a detailed discussion of validation methods, data mining algorithms, and datasets for fake news identification [30]. Ahmed et al. employed classification

TABLE 1 Overview of methods and limitations for fake news detection [16]

Method	Limitations
Content-based	Trained professionals are required to analyse statements for integrity, which makes automation difficult.
	Lack of generalisation across languages, topics, and domains does not fully exploit and extract the content's syntactic information.
	It is often challenging to determine veracity by text analysis alone, and thus additional information or fact-checking is necessary.
Propagation-based	Model parameters are not user-specific, and the same rate constants and probabilities are assumed for all users.
Source-based	Network monitoring requires the strategy to be updated based on changing network topologies. Therefore, network monitoring on large-scale networks can be costly because of network complexity.

approaches and n-gram analysis to identify false news and opinion spam [31]. In [32] , Gilda used bounded decision tree, SVM, random forest, gradient boosting, and stochastic gradient algorithms. The author's best results were achieved with the stochastic gradient descent method. Ruchansky et al. [33] used a hybrid algorithm called capture, score, and integrate (CSI) for fake news detection, where three characteristics were combined for more precise prediction. Shu et al. [34] proposed a fake news detection model that considers the association of related user interactions, publisher bias, and news stance.

Moreover, real-world fake news detection datasets were used to verify model efficiency. Long et al. [35] utilised a novel hybrid algorithm focussed on attention-based long short-term memory (LSTM) networks for fake news detection problems. The method was benchmarked against other fake news detection datasets. Figueira and Oliveira [36] reviewed the current state of fake news, suggesting a two-opposite approach for fake news to increase business awareness to automatically detect fake news. A new automated algorithm [10] was introduced by Perez-Rosas et al. The authors displayed a classification model based on combining lexical, syntactic, and semantic details. Buntain and Golbeck [37] introduced an automated method for catching fake news. They used this approach on three publicly accessible datasets. Bessi [38] used online social media to examine the statistical properties of fake news, speculation, and unproven claims. Zu et al. [39] have put forward a competitive approach to reducing the effect of incorrect information that focusses on the correlation between original wrong information and updated information. Shu et al. [40] considered user trust to create a new algorithm for fake news detection. Several other models have been developed for the detection of fake news [32,33,41,42,43]. In [42], the author took a different perspective and approached the issue as a nature language processing (NLP) problem. The author used deep learning based on NLP for the detection of fake news. In addition, the author proposed a new design that incorporated 'attention-like' mechanisms with a convolution network.

Furthermore, the author documented the results of comparing different neural networks (such as recurrent, LSTM, GRU, and attention-augmented convolutional). Accordingly, the RNN architecture with GRUs outperformed LSTM. The author also evaluated the dataset on other classifiers, such as vector support, stochastic gradient decrease etc. In [44], the authors used a detector that was automated with deep learning methods through a hierarchical network of three attention levels (3HAN). 3HAN has been used to construct a news vector with three attention levels corresponding to the input of a news article's sentences, words, and headlines following a hierarchical bottom-up manner. A distinguishing feature of a fake news article is its headline, so relatively few phrases and words in an article are more important than the rest. Owing to its three layers of attention, 3HAN provides differential importance to parts of an article. The authors observed the efficacy of 3HAN with an accuracy of 96.77% through experiments on a large real-world dataset.

3 | DATASET

We used a public dataset, FakeNewsNet [13], that is collected especially for fake news detection. FakeNewsNet contains labelled news from two websites, politifact.com and gossipcop. com. The content includes linguistic and visual information, all the tweets and retweets for each news item, and corresponding Twitter user information. In addition, the dataset contains annotated statements and incorporates spatiotemporal information and information on the social context. In addition, the authors introduced a pipeline of continually updating data for current fake news. Detailed statistics of the FakeNewsNet repository are shown in Figure 1 [13].

3.1 | Preprocessing

In order to make the dataset _t for the model, the sentences were tokenised, and punctuations and stop words were removed. Stop words are less important words and do not define any context. For example, consider the following line:

The FBI also investigated liberal groups that had progressive in their names . . . the FBI was basically looking at everybody.

After tokenising and removing punctuation and stop words, we obtain the following:

	Category	Features	PolitiFact		GossipCop	
			Fake	Real	Fake	Real
News Content	*Linguistic*	# News articles	432	624	5,323	16,817
		# News articles with text	420	528	4,947	16,694
	Visual	# News articles with images	336	447	1,650	16,767
Social Context	*User*	# Users posting tweets	95,553	249,887	265,155	80,137
		# Users involved in likes	113,473	401,363	348,852	145,078
		# Users involved in retweets	106,195	346,459	239,483	118,894
		# Users involved in replies	40,585	18,6675	106,325	50,799
	Post	# Tweets posting news	164,892	399,237	519,581	876,967
	Response	# Tweets with replies	11,975	41,852	39,717	11,912
		# Tweets with likes	31692	93,839	96,906	41,889
		# Tweets with retweets	23,489	67,035	56,552	24,955
	Network	# Followers	405,509,460	1,012,218,640	630,231,413	293,001,487
		# Followees	449,463,557	1,071,492,603	619,207,586	308,428,225
		Average # followers	1299.98	982.67	1020.99	933.64
		Average # followees	1440.89	1040.21	1003.14	982.80
Spatiotemporal Information	*Spatial*	# User profiles with locations	217,379	719,331	429,547	220,264
		# Tweets with locations	3,337	12,692	12,286	2,451
	Temporal	# Timestamps for news pieces	296	167	3,558	9,119
		# Timestamps for response	171,301	669,641	381,600	200,531

FIGURE 1   Statistics of the FakeNewsNet repository

< FBI >, < investigated >, < liberal >, < group >, < progressive >, < names>, < FBI >, < basically >, < looking >, < everybody >

4 | NEWS CONTENT AND COMMENT EMBEDDING

In the following sections, we discuss the process of embedding the words, sentences, and comments separately. The embedded sentences and comments are concatenated before feeding them to the SVM.

4.1 | Word embedding

To encode a word, the RNN-based word encoder is used to learn sentence representation. In an RNN, when the sequence becomes longer, the old memory gets faded away. Therefore, to ascertain long-term RNN dependencies and ensure persistent memory, GRUs were adopted in the RNN. Moreover, the bidirectional GRU [neural machine translation by jointly learning to align and translate] was incorporated to capture the annotations' contextual information. There are two GRUs in a bidirectional GRU; one, Gf, is forward, and the other, Gb, is backward. Forward GRU Gf reads the ith sentence from word x_1^i to x_m^i, while backward GRU G_b reads the ith sentence from word x_m^i to x_1^i.

$$O_f = GRU\left(x_t^i\right),\ t \in \{1, \ldots, m_i\}$$

$$O_b = GRU\left(x_t^i\right),\ t \in \{m_i, \ldots, 1\}$$

The annotation of word x_t^i was obtained by concatenating both forward and backward output states $O_t^i = <O_f,\ O_b>$. The notation O_t^i includes all the information of the whole sentence based on each word x_t^i.

4.2 | Sentence embedding

The same approach used for word encoding is used to encode sentences. The RNN with GRU units is utilised to encode each sentence of news. The annotated word vector O_t^k is used to learn the sentence representation S_k by using the bidirectional GRU units. It can be shown mathematically as follows:

$$s_f^k = GRU\left(O_t^k\right), k \in \{1, 2 \ldots, N\}$$

$$s_b^k = GRU\left(O_t^k\right), k \in \{N, \ldots, 2, 1\}$$

Both forward and backward annotations were concatenated to obtain the sentence annotation $S_k = <S_f^k, S_b^k>$ The notation S_k apprehends the contextual information from the sentences that lie in the locality of sentence k.

4.3 | User comments embedding

User comments are important in this scenario, as people express their opinions towards different news types by posting comments, reactions, and sharing sceptical opinions. Therefore, comments potentially contain useful semantic information to discriminate between real and fake news. The RNN with bidirectional GRU was used to encode such comments and is the same method applied for news in the previous section.

5 | THEORETICAL ANALYSIS

Suppose we are given a news article N that contains S sentences $\{s_i\}_{i=1}^s$. Each sentence s_i consists of Ki words $\{w_j^i\}_{j=1}^{k_i}$. Let C be the set of T comments regarding news N and each comment consist of Qj words. We treat the
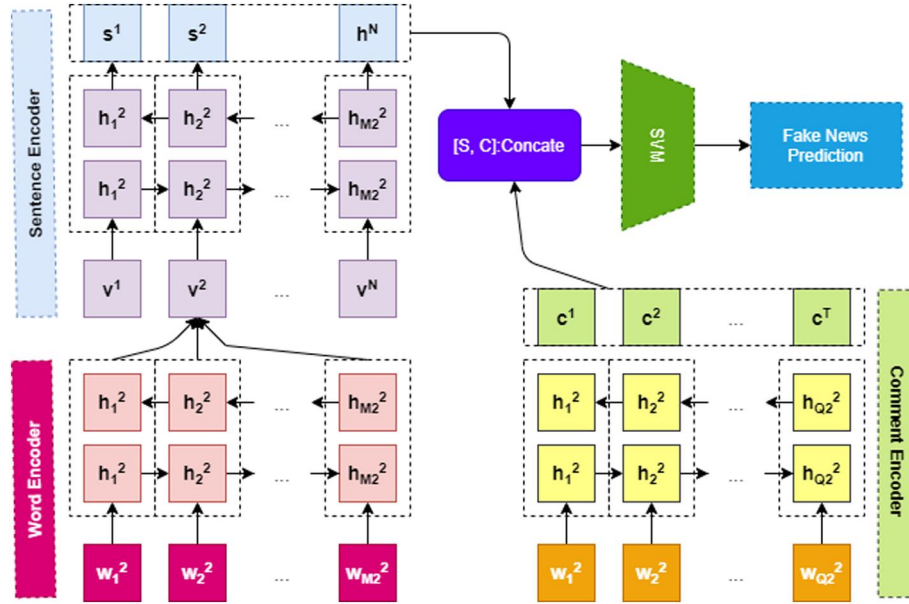
FIGURE 2 Proposed model based on recurrent neural network with bidirectional gated recurrent units and support vector machine. Comments are embedded using the same method used for word embedding

fake news detection problem as a binary classification. Therefore, each news piece may be true $y = 1$ or fake $y = 0$. Further, we apply the embedding process that was explained in Section 4. After embedding, news content and comments were concatenated for each corresponding news article and given as input to the kernelised SVM. The output of the GRUs was limited to 512 units, and after concatenation of the comments, the size was extended to 1024 units. Hence, the size of the feature vector is limited to 1024 units. The SVM with a Gaussian kernel takes the input and transforms it into a high-dimensional space where the input becomes more separable. News can be easily classified in corresponding classes with high precision. Comments add more semantic information that has the potential to assist in fake news detection. This is because people express their emotions or opinions towards fake news via social media posts in the form of doubting opinions, asking related questions, stating their own explanations, and reacting sensationally. Therefore, these traits add an extra dimension to the detection task. As shown in Figure 2, the encoded comments are combined with the news content's sentences to make a vector length of 1024 units, which further enhances fake news detection. The embedding method is the same as for sentence embedding, but logically, the process accumulates the complementing information that is potentially present in the comments. The feature vector's size is set to 1024 units, as this vector becomes the input to the SVM for further processing. The performance of the SVM on this size is more accurate than it is for 512 or 128 units. Further, the input feature is again transformed to a higher dimension; therefore, increasing the size to more than 1024 units increases the processing overhead and thus has a negligible effect on accuracy.

6 | MODEL

In this study, SVM and RNN with bidirectional GRUs are incorporated to make a hybrid model to detect rumours in a given tweet. SVM is an efficient algorithm that works well on high-dimensional data, while RNN is a variant of neural networks that incorporate previous time-stamp data. Both SVM and RNN are sequential models, and the combination of both outperformed the existing algorithms. In the first step, the labelled tweets are fed to the RNN model with bidirectional GRU units and word and sentence embedding. In the second phase, the encoded words, sentences, and extracted document features are fed to the SVM. The SVM model was trained on the interim features, and the results of the model were recorded.

6.1 | Architecture

In our approach, RNN and SVM are introduced for detecting real and fake news. RNN with bidirectional GRUs was used for embedding the content and yields output in one-hot encoding as a 2D matrix. This output was given as input to the SVM to train it to detect real and fake news in the inputs. The SVM is incorporated with a Gaussian kernel, as shown in Equation (3), to enhance the detection power of the SVM:

$$K(x, x^-) = \exp\left(- \frac{\left\| x - x^- \right\|^{-2}}{2\sigma^2} \right)$$

To determine the feature vector's size and the maximum number of words, we choose the possible longest sentences in the content. Those sentences having a length equal to or less than the maximum number of words were padded to make

Datasets	Measure	Our model	HAN	TCNN–URG	HPA–BLSTM	CSI
PolitiFact	Accuracy	**0.912**	0.837	0.712	0.846	0.827
	Precision	**0.910**	0.824	0.711	0.894	0.847
	Recall	**0.961**	0.896	0.941	0.868	0.897
	F1 score	**0.932**	0.860	0.810	0.881	0.871
GossipCop	Accuracy	**0.802**	0.742	0.736	0.753	0.772
	Precision	0.730	0.655	0.715	0.684	**0.732**
	Recall	**0.790**	0.689	0.521	0.662	0.638
	F1 score	**0.762**	0.672	0.603	0.673	0.682

Abbreviations: CSI, capture, score, and integrate; HAN, hierarchical attention network; HPA–BLSTM, hypothalamic-pituitary-adrenocortical–bidirectional long short-term memory; TCNN–URG, transfer convolutional neural network–user response generator.

Note: Best results are highlighted in bold

them 30-word sentences. All these unique-sized sentences were mapped to GRU layers to yield embedded sentences in the form of 2D one-hot encoding. Finally, this output was given as input to the SVM to detect true and false news classes. An illustration of our proposed model is shown in Figure 2.

6.2 | Fine-tuning

We choose accuracy, precision, recall, and F1 score performance metrics to evaluate and compare the model's performance. In the training process, we selected 75% of the samples randomly and evaluated the performance of the remaining 25% of the sample. This process was repeated five times, and the average measures are described in Table 2. The parameters for SVM were set to 0.001 and 1.0, respectively. These values were experimentally proved to be the best parameter values.

7 | RESULTS AND COMPARISON

To further compare our proposed model, we chose the existing representative models described in the discussion that follows. For comparison with these models, various measures such as accuracy, f-score, precision, and sensitivity are computed, as shown in Table 2. However, it is often hard to measure the model's performance using accuracy, recall, and precision, so we need to examine the AUC-ROC curve, which allows a false positive rate because it plots the true positive rate against a false positive rate. Figures 3 and 4 show the micro and macro average and class-wise AUC scores achieved with the proposed model and show consistent AUC scores, indicating stable predictions of the proposed model. In the following methods, hypothalamic-pituitary-adrenocortical–bidirectional long short-term memory (HPA–BLSTM) [45] and CSI [33] incorporate the news content and consider user comments, while other methods work based only on news content. Results in terms of the accuracy and F1 score of our proposed model were compared with the existing state-of-the-art models given
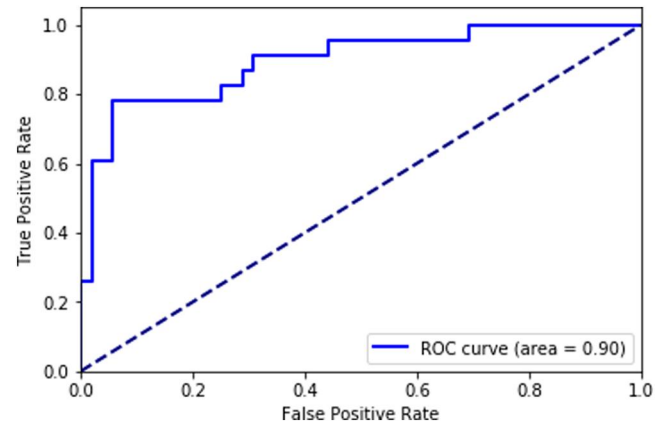
**FIGURE 3** AUC-ROC curve for PolitiFact dataset
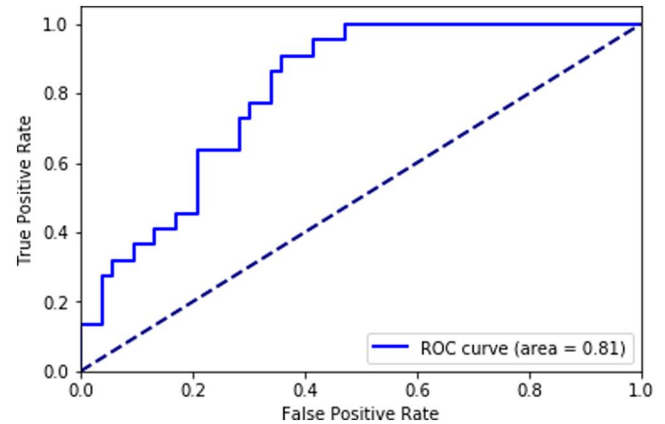
FIGURE 4 AUC-ROC curve for GossipCop dataset

in Figures 5 and 6. In Table 2, various performance measures indicate that our work outperforms the existing models.

- **HAN** [46]: To detect fake news, the hierarchical attention network (HAN) applies the framework of hierarchical attention neural networks on news content. In this

FIGURE 5 Comparison of our model with existing models in terms of accuracy and F1 score for PolitiFact dataset
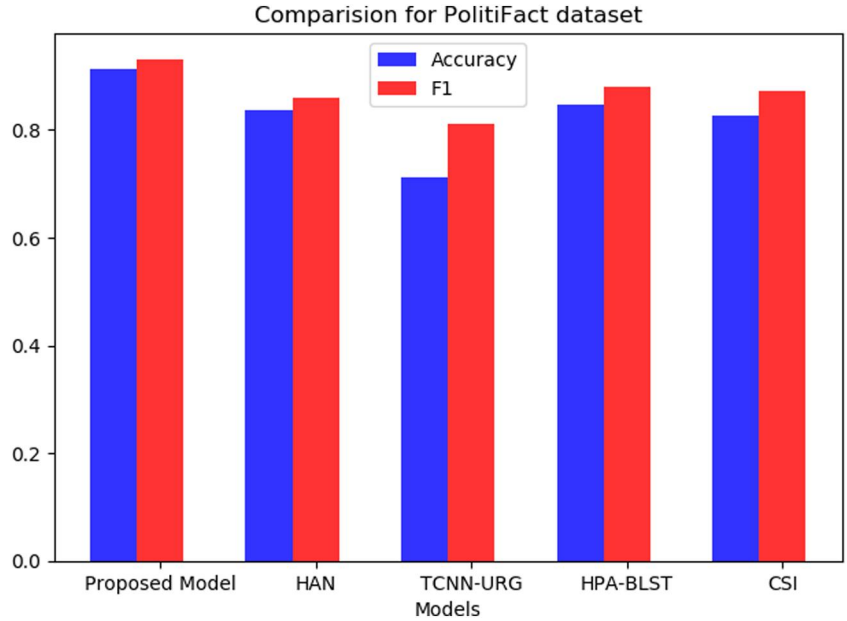
framework, word-level attention is used on each sentence, while sentence-level attention is used on each document.

- **TCNN–URG** [15]: The transfer convolutional neural network–user response generator (TCNN–URG) framework contains two major components—a convolutional neural network and a conditional variational autoencoder. The convolutional neural network is responsible for learning representations from news content, while the variational auto-encoder is used to record attributes from user comments.
- **HPA–BLSTM** [45]: HPA–BLSTM utilises the HAN framework based on a neural network. It learns news embedding of word-level, postlevel, and sub-event-level user involvement on social media.
- **CSI** [33]: CSI is an LSTM-based hybrid deep learning framework that models news representations based on Doc2Vec embedding by taking news content and user comments as inputs.

8 | DISCUSSION AND LIMITATIONS

To consider only news content–based methods, semantic and syntactic cues can efficiently be captured through the HAN framework based on hierarchical attention neural networks. Our work is based on news content, but in addition, it incorporates comment data. From the results in Table 2, it can be observed that the comments contain complementary information; this helped us boost the detection performance of the method. The performance of our proposed method is higher than that of the HPA–BLSTM and CSI frameworks. HPA–BLSTM and CSI are user comments–based methods, and the performance of such methods is higher than that of the news content–based methods. Considering the incorporation of

both comments and news content revealed that user comments have more discriminative power than either news content or comments alone. However, because our proposed model exploits user comments to improve the detection rate, this makes it highly vulnerable to attacks via adversarial comments (such as UNITRIGGER [47], HOTFLIP [48], TextBugger [49], and MALCOM [50]). In particular, generating high-quality comments and coherency to articles by using a less diverse set of highly relevant words can mislead our model.

Besides these attacks, GPT-2 [51] is a modern method that can produce fake texts. GPT-2 is a massive transformer-based language model with 1.5 billion parameters with the basic goal of predicting the next word in any text from previous words [51]. The terms 'real text' and 'fake text' refer only to whether the text was generated by a machine or a human. However, GPT-2 does not make a clear argument about the veracity of a piece of content. There are instances where a language model produces a valid expression, and there are also instances where a human writes a false statement. We argue that machine-generated text is not the same as false text. Identifying whether a part of the content is likely to be machine-generated may help gauge credibility. For instance, incorporating linguistic patterns can effectively detect human-written comments to identify a malicious comment generated by a machine. While this may improve the model's performance for detecting fake news, it also raises the bar for potential attacks.

In a different direction, a few studies [52,53,54] have suggested that fake news detection should not be simplified as supervised approaches that focus on individual binary labels. Instead, it should be modelled and studied as a continuous spectrum of values. Particularly, we should consider the complexity of embracing manipulation and deception to spot suspicious coordination. Consequently, the potential solution
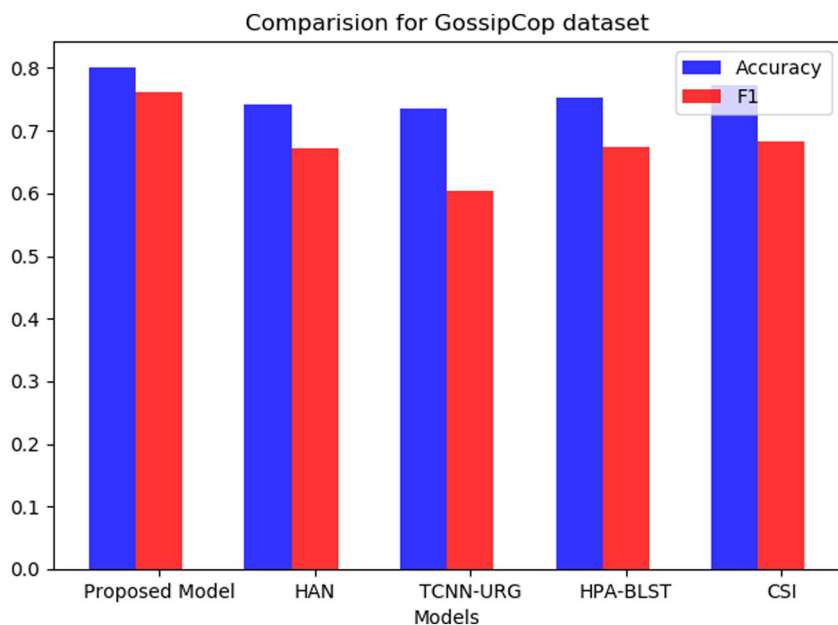
should not include oversimplistic linear labels but should instead produce multifaceted questionable cooperation measurements.

Finally, one SVM limitation is that its performance depends on feature vector size. As the feature vector becomes larger, performance goes up, but performance decreases for small feature vectors. Therefore, the feature vector's minimum size (output of the GRUs) was limited to 512 units.

9 | CONCLUSION

We have proposed a hybrid model based on an RNN and a SVM to detect rumours in news content in FakeNews datasets composed of two subparts, the PolitiFact and GossipCop datasets. RNN was utilised for encoding news content and comments to feature representation. The features were given as input to an SVM with a Gaussian kernel to detect rumours (real or fake news) in the input data. Results in terms of the accuracy and F1 score of our proposed model were compared with existing state-of-the-art models. Various performance measures show that our work has outperformed those existing models.

Researchers need to devote more attention to grasping fake news structures to understand their patterns and dissemination across the digital universe. The online presentation of fake news has adapted in line with digital growth and continues to acquire new formats that are increasingly significant. Future research should discuss the effect of fake news and misinformation broadly for the latest forms. The phenomenon of fake news with a broader range of sources in various scenarios will also be essential in future studies.

CONFLICT OF INTEREST
The author declares that I have no conflict of interest.

ORCID
Marwan Albahar https://orcid.org/0000-0003-3586-3423

REFERENCES
1. Allcott, H., Gentzkow, M.: Social media and fake news in the 2016 election. J. Econ. Perspect. 31, 211–36 (2017)
2. Zhang, Y., et al.: Rumour and authoritative information propagation model considering super spreading in complex social networks. Phys. Stat. Mech. Appl. 506, 395–411 (2018)
3. Girgis, S., Amer, E., Gadallah, M.: Deep learning algorithms for detecting fake news in online text. 13th IEEE International Conference on computer Engineering and Systems (ICCES) (2018)
4. Liu, Y., fang Brook Wu, Y.: Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. AAAI. 354–361 (2018)
5. Popat, K.: Assessing the credibility of claims on the web. In: Proceedings of the 26th International Conference on World Wide Web Companion, pp. 735–739.International World Wide Web Conferences Steering Committee (2017)
6. Gupta, A., et al.: TweetCred: real-Time credibility Assessment of content on Twitter, pp. 228–243.Springer International Publishing (2014)
7. Yang, F., et al.: Automatic detection of rumour on sina weibo. In: Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics, vol. 12. MDSAssociation for Computing Machinery, New York (2012)
8. Castillo, C., Mendoza, M., Poblete, B.: Information credibility on twitter. In: Proceedings of the 20th International Conference on World Wide Web, vol. 11, pp. 675–684.WWWAssociation for Computing Machinery, New York (2011)
9. Horne, B.D., Adali, S.: This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news, (2017) arXiv, eprint:1703.09398
10. P_erez-Rosas, V., et al.: Automatic detection of fake news. In: Coling '18, pp. 3391–3401 (2018)
11. Newman, M.L. et al.: Lying words: predicting deception from linguistic styles. Pers. Soc. Psychol. Bull. 29(5), 665–675 (2003)
12. Silverman, C.: Lies, Damn lies, and Viral content: How news websites spread (and Debunk) online rumours, Unverified claims, and misinformation. Tow Centre for Digital Journalism. Columbia University (2015)
13. Shu, K., et al.: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big Data. 8(3), 171–188 (2020)

14. Zubiaga, A., et al.: Detection and resolution of rumours in social media. ACM Comput. Surv. 51, 1–36 (2018)

15. Qian, F. et al.: Neural user response generator: fake news detection with collective user intelligence. Int. Joint Conf. Artif. Intell. 18, 3834–3840 (2018)

16. Zhou, X., Zafarani, R.: A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Comput. Surv. 53 (2020)

17. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science. 359(6380), 1146–1151 (2018)

18. Baly, R., et al.: Predicting factuality of reporting and bias of news media sources. CoRR. abs/1810, 01765 (2018)

19. Afroz, S., Brennan, M., Greenstadt, R.: Detecting hoaxes, frauds, and deception in writing style online. In: 2012 IEEE Symposium on Security and Privacy, pp. 461–475 (2012)

20. Rubin, V., et al.: Fake news or truth? using satirical cues to detect potentially misleading news. In: Proceedings of the Second Workshop on Computational Approaches to Deception Detection, pp. 7–17.Association for Computational Linguistics, San Diego, CA (2016)

21. Rashkin, H., et al.: Truth of varying shades: analysing language in fake news and political fact-checking. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 2931–2937.Association for Computational Linguistics, Copenhagen (2017)

22. Potthast, M., et al.: A stylometric inquiry into hyperpartisan and fake news. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Long Papers, vol. 1 (2018)

23. Bhatt, G., et al.: Combining neural, statistical and External features for fake news stance identification. Companion of the the Web Conference 2018 on the Web Conference. WWW '18 (2018)

24. Borges, L., Martins, B., Calado, P.: Combining similarity features and deep representation learning for stance detection in the context of checking fake news. J. Data Infor. Qual. 11 (2019)

25. Rubin, V.L., Chen, Y., Conroy, N.K.: Deception detection for news: three types of fakes. Proc. Assoc. Infor. Sci. Technol. 52(1), 1–4 (Jan. 2015)

26. Conroy, N.K., Rubin, V.L., Chen, Y.: Automatic deception detection: methods for finding fake news. Proc. Assoc. Infor. Sci. Technol. 52(1), 1–4 (2015)

27. Chen, Y., Rubin, V.L., Conroy, N.: Towards news verification: deception detection methods for news discourse. In: Proceedings of the 48th European Conference on Principles of knowledge Discovery and data mining (PKDD), 48th ed. (2015)

28. Dadgar, S.M.H., Araghi, M.S., Farahani, M.M.: A novel text mining approach based on TF-IDF and support vector machine for news classification. In: 2016 IEEE International Conference on Engineering and Technology (ICETECH), Coimbatore, pp. 112–116 (2016)

29. Jin, Z., et al.: News verification by exploiting conflicting social viewpoints in microblogs. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI, vol. 16, pp. 2972–2978.AAAI Press (2016)

30. Shu, K., et al.: Fake news detection on social media. SIGKDD Expl. Newsl. 19(1), 22–36 (2017)

31. Ahmed, H., Traore, I., Saad, S.: Detecting opinion spams and fake news using text classification. Secur. Privacy. 1(1), e9 (2017)

32. Gilda, S.: Notice of Violation of IEEE Publication Principles: Evaluating machine learning algorithms for fake news detection. 2017 IEEE 15th student Conference on research and development. (SCOReD) (2017)

33. Ruchansky, N., et al.: A hybrid deep model for fake news detection. Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (2017)

34. Shu, K., Wang, S., Liu, H.: Exploiting tri-relationship for fake news detection. ArXiv. vol. abs/1712, 07709 (2017)

35. Long, Y., et al.: Fake news detection through multi-perspective speaker profiles. In: Proceedings of the Eighth International Joint Conference on Natural Language Processing, pp. 252–256.Asian Federation of Natural Language Processing (2017)

36. Figueira, Á., Oliveira, L.: The current state of fake news: challenges and opportunities. Procedia Comput. Sci. 121, 817–825 (2017)

37. Buntain, C., Golbeck, J.: Automatically identifying fake news in popular twitter threads. In: 2017 IEEE International Conference on Smart Cloud, pp. 208–215 (2017)

38. Bessi, A.: On the statistical properties of viral misinformation in online social media. Phys. Stat. Mech. Appl. 469, 459–470 (2017)

39. Azam, N., Yao, J.: Comparison of term frequency and document frequency-based feature selection metrics in text categorisation. Expert Syst. Appl. 39(5), 4760–4768 (2012)

40. Shu, K., Wang, S., Liu, H.: Understanding user profiles on social media for fake news detection. In: 2018 IEEE Conference on Multimedia information processing and Retrieval, (MIPR), Miami, pp. 430–435 (2018)

41. Zhang, J., et al.: Fake news detection with deep diffusive network model. ArXiv. vol. abs/1805, 08751 (2018)

42. Bajaj, S.: The pope has a new baby! fake news detection using deep learning (2017)

43. Trovati, M., Hill, R., Bessis, N.: A non-genuine message detection method based on unstructured datasets. In: 2015 10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, (3PGCIC), Krakow, pp. 597–600 (2015)

44. Singhania, S., et al.: A deep neural network for fake news detection. In Neural Information Processing, pp. 572–581.Springer International Publishing (2017)

45. Guo, H., et al.: Rumour detection with hierarchical social attention network. Proceedings of the 27th ACM International Conference on Information and knowledge Management (2018)

46. Yang, Z., et al.: Hierarchical attention networks for document classification. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics. Human Language Technologies (2016)

47. Cer, D., et al.: Universal Sentence Encoder (2018)

48. Ebrahimi, J., et al.: HotFlip: White-Box adversarial examples for text classification. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (2018)

49. Li, J., et al.: TextBugger: generating adversarial text against real-world applications. Proceedings 2019 Network and Distributed System Security Symposium (2019)

50. Le, T., et al.: Generating malicious comments to attack neural fake news detection models, pp. 2020 (2009) ArXiv, abs/.01048

51. Radford, A, et al.: Language models are unsupervised multitask learners (2019)

52. Starbird, K.: Disinformation's spread: bots, trolls and all of us. Nature. 571, 449–449 (2019)

53. Cresci, S.: A decade of social bot detection. Commun. ACM. 63, 72–83 (2020)

54. Nizzoli, L., et al.: Coordinated behaviour on social media in 2019 UK general election (2020) arXiv e-prints, p. arXiv:2008.08370