Explainable Detection of Fake News on Social Media Using Pyramidal Co-Attention Network

Fazlullah Khan, *Senior Member, IEEE*, Ryan Alturki, *Senior Member, IEEE*, Gautam Srivastava, *Senior Member, IEEE*, Foziah Gazzawe, *Member, IEEE*, Syed Tauhid Ullah Shah, *Student Member, IEEE*, and Spyridon Mastorakis, *Member, IEEE*

***Abstract*— In today's world, fake news on social media is a universal trend and has severe consequences. There has been a wide variety of countermeasures developed to offset the effect and propagation of Fake News. The most common are linguistic-based techniques, which mostly use deep learning (DL) and natural language processing (NLP). Even government-sponsored organizations spread fake news as a cyberwar strategy. In literature, computational-based detection of fake news has been investigated to minimize it. The initial results of these studies are good but not significant. However, we argue that the explainability of such detection, particularly why a certain news item is detected as fake, is a vital missing element of the studies. In real-world settings, the explainability of the system's decisions is just as important as its accuracy. This article explores explainable fake news detection and proposes a sentence-comment-based co-attention sub-network model. The proposed model uses user comments and news contents to mutually apprehend top-*k* explainable check-worthy user comments and sentences for detecting fake news. The experimental result on real-world datasets shows that our proposed model outperforms state-of-the-art techniques by 5.56% in the *F*1 score. In addition, our model outperforms other baselines by 16.4% in normalized cumulative gain (NDCG) and 22.1% in Precision in identifying top-*k* comments from users, which indicates why a news article can be fake.**

***Index Terms*— Attention, deep learning (DL), fake news detection, long short-term memory (LSTM).**

I. Introduction

SOCIAL media platforms enable users to easily access, share, and generate a wide range of information. More individuals are seeking and receiving prompt news information online due to the growing utilization and comfort of social media. According to the Pew Research Center, roughly 68% of adults in the U.S. acquired news through social media in 2018, compared to 49% in 2012.[1] It indicates that social media exposes users to a plethora of misinformation, including fake news and news items that deliberately incorporate fraudulent content [1], [2]. According to one estimate,[2] all the way to the end of the 2016 presidential election, over 1 million tweets had been tied to the fake news story "Pizzagate." Therefore, we say fake news can damage society due to its extensive distribution. First, it erodes public confidence in journalists and governments. For example, during the 2016 presidential election campaign in the United States, the spread of fake news was, ironically, greater than the reach of the top 20 most-discussed factual stories.[3] In addition, fake news can alter how people react to real news. According to a Gallup investigation,[4] people's confidence in the media has eroded drastically across all political parties and age groups. Also, widespread "online" misinformation and fake news can cause "offline" societal incidents.[5] Fake news stating Barack Obama was wounded in an eruption cost the stock market U.S. $130 billion.[6] Consequently, controlling its propagation on social media has become vital while boosting trust in a wider news ecosystem.

Nevertheless, the detection of fake news on social media brings its own set of difficulties. First, because fake news is purposefully created to deceive readers, it is difficult to detect them solely by their contents. Also, data on social media is occasionally anonymous, large-scale, noisy, primarily user-generated, and multimodal. To address these issues, recent efforts have integrated users' social interactions into news stories to determine whether news pieces are fake [3], [4], with promising early outcomes. Ruchansky *et al.* [3] present a hybrid deep learning (DL) approach for detecting fake news by modeling news content, user responses, and post sources. Guo *et al.* [5] employ a hierarchical network and model social attention with user interactions and relevant user comments for fake news detection. Even though effective DL-enabled techniques are in practice, the existing techniques emphasize the detection of fake news with latent features. However, these cannot justify "why" a portion of the information was flagged

Manuscript received 28 December 2021; revised 11 April 2022, 10 June 2022, 24 July 2022, and 20 August 2022; accepted 13 September 2022. This work was supported in part by the National Science Foundation under Award CNS-2104700, Award CNS-2016714, and Award CBET-2124918; in part by the National Institutes of Health under Grant NIGMS/P20GM109090; and in part by the Nebraska University Collaboration Initiative, and the Nebraska Tobacco Settlement Biomedical Research Development Funds. *(Corresponding authors: Fazlullah Khan; Ryan Alturki.)*

Fazlullah Khan is with the Department of Computer Science, Abdul Wali Khan University Mardan, Mardan 23200, Pakistan (e-mail: fazlullah.mcs@gmail.com).

Ryan Alturki and Foziah Gazzawe are with the Department of Information Science, College of Computer and Information Systems, Umm Al-Qura University, Mecca 24382, Saudi Arabia (e-mail: rmturki@uqu.edu.sa).

Gautam Srivastava is with the Department of Mathematics and Computer Science, Brandon University, Brandon, MB R7A 6A9, Canada, also with the Research Center for Interneural Computing, China Medical University, Taichung 40402, Taiwan, and also with the Department of Computer Science and Mathematics, Lebanese American University, Beirut 1102, Lebanon.

Syed Tauhid Ullah Shah is with the Department of Electrical and Software Engineering, Schulich School of Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada.

Spyridon Mastorakis is with the Department of Computer Science, University of Nebraska at Omaha, Omaha, NE 68182 USA.

Digital Object Identifier 10.1109/TCSS.2022.3207993

[1] https://tinyurl.com/ybcy2foa
[2] https://tinyurl.com/z38z5zh
[3] https://tinyurl.com/y8dckwhr
[4] https://tinyurl.com/y9kegobd
[5] https://tinyurl.com/y9kegobd
[6] https://tinyurl.com/ybs4tgpg

as fake news. It is desirable to justify that news was deemed fake because the developed explanation can deliver practitioners unique insights and previously unattainable details. Furthermore, extracting explainable features from noisy auxiliary data can aid in the detection of fake news. Shu *et al.* [6] proposed dEFEND, an recurrent neural networks (RNN) and co-attention-based technique for detecting explainable news. They employ bidirectional long short-term memorys (LSTMs) and a dual path-way network for explainable fake news detection. RNN networks [LSTM and gated recurrent unit (GRU)], on the other hand, lack generalizability and so performs poorly [7]. Also, co-attention inadequacies global cross-interaction of all features [8]. However, designing a reliable and explainable fake news detection framework remains a research problem.

In this article, we propose a framework to obtain explanations in terms of user comments and news content. First, news articles may contain demonstrably fake content. Journalists, for example, manually review sentences in news pieces on fact-checking platforms like PolitiFact,[7] which is time-consuming and labor-intensive. Also, researchers seek to utilize external sources to determine and explain if a news story is fake or not [9]. In addition, user comments on social media supply plenty of information from the community, such as sentiments, stances, and opinions, which may be used to identify fake news [5]. Furthermore, user comments and news content are inextricably linked and offer key indications to justify whether certain news is fake or true [6]. In the proposed framework, we investigate the issue of detecting fake news by combining information from user comments and news articles. To that aim, we employ a systematic approach to develop an explainable framework for detecting fake news, including: 1) a component for encoding news contents that employs a hierarchical attention network to learn news sentence representation and apprehend grammatical and semantic signals; 2) a component for encoding user comments that employ a word-level attention sub-network to learn the latent representations of user comments; and 3) a sentence-comment-based co-attention component that explains the association between user comments and news contents and helps determine the top-k explainable comments and sentences.

In summary, we primarily address the following issues: 1) increasing explainability and detection performance by modeling explainable fake news detection; 2) efficiently extracting explainable comments in the absence of ground labels in the course of model training; and 3) mutually forming the association between user comments and news content for explainable fake news detection? We present a novel solution to these challenges; our primary contributions are detailed below.

1) We investigate and solve a novel and critical problem for detecting explainable fake news using social media platforms.
2) We present a structured method for exploiting user comments and news contents to apprehend explainable user comments for detecting fake news.
3) We performed comprehensive simulation on real-world datasets that exhibit the supremacy of our proposed method over the baseline techniques for explaining and detecting fake news.

The rest of the article is organized according to the following pattern. In Section II related work is discussed followed by preliminaries in Section III. The methodology is given in Section IV, and experimental results are given in Section V. Finally, the article is concluded in Section VI.

II. Related Work

Most methods for detecting fake news rely on social contexts, and news contents [1], [10], [11]. In general, information is extracted from visual and textual elements, which apprehend certain sensational emotions [12] and writing styles [13], both of which are vital in fake news detection. Furthermore, tensor factorization [14] is used to model textual representations, yielding strong results in detecting fake news. The visual elements are derived from visual components to apprehend the diverse features of fake news and misinformation [15]. In this context, recent efforts focus on the fundamental issues in fake news detection through latent representations, including user response generation [16], early detection of fake news using adversarial learning [17], explainable detection meta characteristics [18] and unsupervised and semi-supervised detection [19]. On the other hand, the latent representations are notoriously challenging to interpret, offering just a sliver of information into fake news. Castillo *et al.* [20] proposed social context-based characteristics derived from interaction patterns and user profiles. Other supervised learning techniques rely on social platform-specific characteristics, including retweets, tweets, and likes. Other supervised learning techniques rely on social platform-specific characteristics, including retweets, tweets, and likes [3], [21].

While there has been great improvement in detecting fake news, little focus has been on explainability. Existing techniques train classifiers by extracting features without providing any explanation. They are black boxes in terms of explainability due to their lack of transparency [22]. In this article, we employ a co-attention approach to better identify fake news by apprehending the inherent explainability of user comments and news sentences.

III. Preliminaries

In this section, we discuss the problem formalization and problem definition in detail.

A. Problem Formalization and Notations

Assume A represents a new article with N sentences $\{s_i\}_{i=1}^{N}$, where each sentence $s_i = \{w_1^i, \ldots, w_{M_i}^i\}$ comprises M_i words. $C = \{c_1, c_2, \ldots, c_T\}$ indicates a set of comments T, associated with the news A, with each comment $c_j = \{w_1^j, \ldots, w_{Q_j}^j\}$ comprises Q_j words. Following [1], [23], and [6], we approach the problem as a binary classification problem, where $y = 1$ indicates true while $y = 0$ shows fake news articles. Similarly, we intend to learn a rank list rank of a sentence (RS) and rank of a comment (RC) based

[7] https://www.politifact.com/

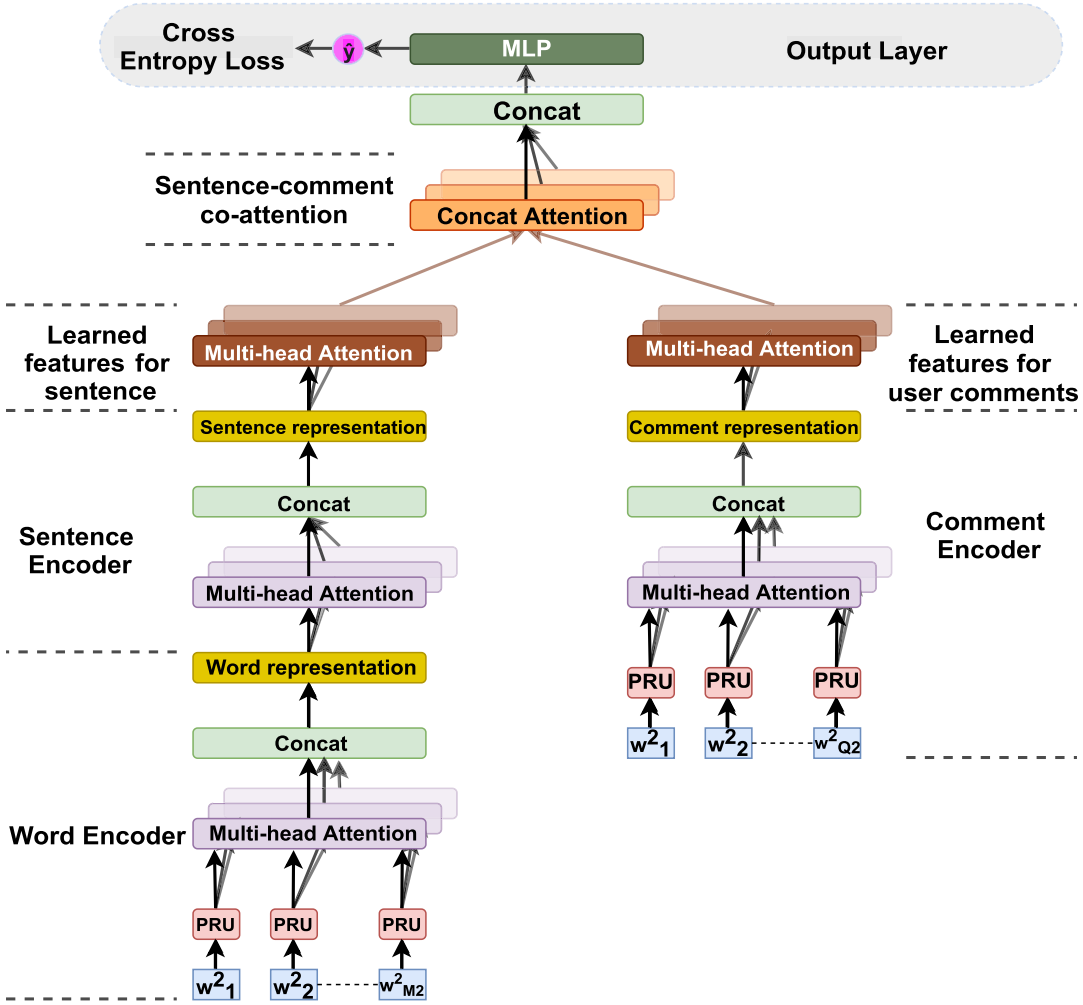

Fig. 1. General architecture of our proposed model.

on the degree of explainability from all sentences $\{s_i\}_{i=1}^{N}$ and comments $\{c_i\}_{j=1}^{T}$, where $\mathrm{RS}_k/\mathrm{RC}_k$ signifies the kth most explainable sentence/comment. The explainability of sentences in news content reflects the extent to which they are check-worthy. Still, the explainability of comments represents the extent to which users believe whether the news is fake or real, which is strongly tied to the primary claims in the news.

B. Problem Definition

Let us assume A and C represent a news article and a set of related comments, respectively. We aim to design a function $f : f(A, C) \rightarrow (\hat{y}, \mathrm{RS}, \mathrm{RC})$, which can accurately predict explainable comments and sentences with high ranks in RC and RS, respectively.

IV. Methodology

This section goes through a news content encoder component, a user comment encoder component, a sentence-comment co-attention component, and a fusion component. Fig. 1 illustrates the architecture of our proposed framework. The news content encoder component uses the hierarchical chat-level and discussion-level encoding to establish latent features of linguistic news. The user comment encoder component then demonstrates how to employ word-level attention networks to extract latent features from comments. In addition, the sentence-comment co-attention component for learning feature representations represents the common impacts among user comments and news sentences. The explainability degree of comments and sentences is learned via attention weights. Lastly, the fusion (fake news prediction component) demonstrates how to classify fake news by combining news content and user comment elements.

A. Encoding of News Contents

The purpose of articles having fake news is to propagate misinformation, which frequently uses limited and dazzling language that can aid in detecting fake news. Furthermore, a news document includes linguistic indicators at various levels, giving varying degrees of value to explaining why news can be fake. For instance, fake news claiming, "Did Trump Send This Phallic Christmas Card in 2021?" Here, the word "Phallic" gives better cues than other terms in the sentence in determining if the news claim is fake or not. In such scenario, hierarchical attention neural networks (HANs) [24], [25], which emphasize influential words or sentences for classification. These have been efficient and fruitful in learning document representations. They employ a hierarchical neural network for modeling sentence-level word-level representations through self-attention techniques. Influenced by the

success of self-attention, we proposed a hierarchical framework for learning news content representations. First, we learn the sentence representation with the word encoder component and sentence vectors with the sentence encoder component.

1) Word Encoders: We train a word encoder based on Pyramidal Recurrent Units (PRUs) to learn the sentence representation. RNNs can capture long-term dependency in principle, but their high dimension space limits their generalizability [7]. PRUs are employed to provide a more in-depth and improved learning experience to address generalizability. Unlike [26] and [6], we use PRU to encode the sequence of words and employ bidirectional PRU to model word sequences from both sides of words to apprehend contextual information from annotations better. PRU applies Pyramidal transformation to the input vector and uses Grouped linear transformation to the context vector. Then, they combine them under the umbrella of PRU and feed them as input to the LSTM cell. Given the input sentence s, the PRU subsample it into K pyramidal levels to obtain various representations with different scales as

$$\mathbf{s}^k \in \mathbb{R}^{\frac{N}{2^{k-1}}} \tag{1}$$

where 2^{k-1} denotes the sampling rate and $k = \{1, \ldots, K\}$. For each $k = \{1, \ldots, K\}$, the PRU learns a scale-specific transformation as

$$\mathbf{W}^k \in \mathbb{R}^{\frac{N}{2^{k-1}} \times \frac{M}{K}}. \tag{2}$$

Then, PRU concatenates the transformed sub-samples to get the pyramidal output $\overline{\mathbf{y}} \in \mathbb{R}^M$ as

$$\overline{\mathbf{y}} = \mathcal{F}_P(\mathbf{s}) = \left[\mathbf{W}^1 \cdot \mathbf{s}^1, \ldots, \mathbf{W}^K \cdot \mathbf{s}^K\right] \tag{3}$$

where $[\cdot, \cdot]$ denotes the concatenation operation. Given an input vector s, PRU sub-sample it using a kernel k with $2e + 1$ elements as

$$\mathbf{s}^k = \sum_{i=1}^{N/s} \sum_{j=-e}^{e} \mathbf{x}^{k-1}[si]\kappa[j] \tag{4}$$

where s denotes the stride operation while $k = \{2, \ldots, K\}$.

At the same time, given the context vector $\mathbf{h} \in \mathbb{R}^N$, PRU split it into g smaller groups as

$$h = \left\{\mathbf{h}^1, \ldots, \mathbf{h}^g\right\}, \quad \forall \mathbf{h}^i \in \mathbb{R}^{\frac{N}{g}}. \tag{5}$$

Then, through a linear transformation $\mathcal{F}_L : \mathbb{R}^{(N/g)} \to \mathbb{R}^{(M/g)}$, PRU transforms h^i into $\mathbf{z}^i \in \mathbb{R}^{(M/g)}$ for each $i = \{1, \ldots, g\}$. The final output vector is then formed by concatenating the resulting g output vectors z^i as

$$\overline{\mathbf{z}} = \mathcal{F}_G(\mathbf{h}) = \left[\mathbf{W}^1 \cdot \mathbf{h}^1, \ldots, \mathbf{W}^g \cdot \mathbf{h}^g\right]. \tag{6}$$

Then, at a given time t, PRU combines input and context vectors through a transformation function using the following equation:

$$\hat{\mathcal{G}}_v(\mathbf{s}_t, \mathbf{h}_{t-1}) = \hat{\mathcal{F}}_P(\mathbf{s}_t) + \mathcal{F}_G(\mathbf{h}_{t-1}) \tag{7}$$

where $v \in \{f, i, c, o\}$ indicates the forget, input, and output gates of the vanilla LSTM. $\hat{\mathcal{F}}_P(\cdot)$ denotes the pyramidal while $\mathcal{F}_G(\cdot)$ represent the grouped linear transformations. The resultant $\mathcal{G}_v$ is then fed to the vanilla LSTM architecture to model PRU.

The bidirectional PRU consists of a forward PRU $\vec{f}$, reading sentences s_i from word w_1^i to $w_{M_i}^i$ and a backward PRU $\overleftarrow{f}$, reading sentences s_i from word $w_{M_i}^i$ to w_1^i

$$\begin{aligned} \overrightarrow{\mathbf{h}_t^i} &= \overrightarrow{\text{PRU}}\left(\mathbf{w}_t^i\right), \quad t \in \{1, \ldots, M_i\} \\ \overleftarrow{\mathbf{h}_t^i} &= \overleftarrow{\text{PRU}}\left(\mathbf{w}_t^i\right), \quad t \in \{M_i, \ldots, 1\}. \end{aligned} \tag{8}$$

We then concatenate $\overrightarrow{\mathbf{h}_t^i}$ and $\overleftarrow{\mathbf{h}_t^i}$, i.e., $\mathbf{h}_t^i = [\overrightarrow{\mathbf{h}_t^i}, \overleftarrow{\mathbf{h}_t^i}]$, to get an annotation having the information of the entire sentences, revealing around the word w_t^i. Because not every word contributes equally to conveying the meaning of a sentence. To better embrace the importance of weights in measuring words, we use a multihead attention approach and compute the sentence vector $\alpha^i \in \mathbb{R}^{2d \times 1}$ as follows:

$$\begin{aligned} \alpha_t^i &= \text{Concat}\left(a_t^1, \ldots, a_t^m\right) \\ \text{where } a_t^i &= \text{softmax}\left(\tanh\left(\left[\mathbf{h}_t^i\right]\mathbf{W}_1\right)\mathbf{w}_2\right) \end{aligned} \tag{9}$$

where $\mathbf{W}_1$ and $\mathbf{w}_2$ learnable parameters. α_t^i denotes the significance of the word tth in the sentence s_i.

2) Sentence Encoder: We use PRUs and multihead attention in the same way as word encoders to encode every sentence in the news. We use the α^i to learn the sentence representations $\mathbf{h}^i$ by apprehending context details at the sentence level. In particular, we employ bidirectional PRUs to encode and learn the sentence representations $\mathrm{s}^i \in \mathbb{R}^{2d \times 1}$ using multihead attention as follows:

$$\begin{aligned} \overrightarrow{\mathbf{h}^i} &= \overrightarrow{\text{PRU}}\left(\alpha^i\right), \quad i \in \{1, \ldots, N\} \\ \overleftarrow{\mathbf{h}^i} &= \overleftarrow{\text{PRU}}\left(\alpha^i\right), \quad i \in \{N, \ldots, 1\}. \end{aligned} \tag{10}$$

The learned sentence representation apprehends the context from neighboring sentences around sentence s_i.

B. Encoding of User's Comments

People use social media posts to articulate their sentiments or sentences about fake news, such as skeptical opinions, comments, sensational reactions, etc. Such textual details are relevant to the scope of actual news stories and possess worthwhile semantic information, which can assist in detecting fake news. We employ bidirectional PRUs to encode the comments of users to latent representations directly because the comments gained from social media are generally brief text. Precisely, we map every word w_t^j in a given comment c_j in words $w_t^j, t \in \{1, \ldots, Q_j\}$ into the word vector $\mathbf{w}_t^j \in \mathbb{R}^d$ through an embedding matrix.

Then, the same as word embeddings, we get the $\overrightarrow{\mathbf{h}}_t^j$ and $\overleftarrow{\mathbf{h}}_t^j$ as

$$\begin{aligned} \overrightarrow{\mathbf{h}}_t^j &= \overrightarrow{\text{PRU}}\left(\mathbf{w}_t^j\right), \quad t \in \left\{1, \ldots, Q_j\right\} \\ \overleftarrow{\mathbf{h}}_t^j &= \overleftarrow{\text{PRU}}\left(\mathbf{w}_t^j\right), \quad t \in \left\{Q_j, \ldots, 1\right\}. \end{aligned} \tag{11}$$

Furthermore, we concatenate $\overrightarrow{\mathbf{h}}_t^j$ and $\overleftarrow{\mathbf{h}}_t^j$, expressed as $\mathbf{h}_t^j = [\mathbf{h}_t^j, \mathbf{h}_t^j]$, to get the annotation of word w_t^j. Then, we utilize multihead attention to learn the importance of measuring the weights of every word to get the comment vector $\mathbf{c}^j \in \mathbb{R}^{2d}$ as

follows:

$$\begin{aligned} \beta_t^j &= \text{Concat}\left(a_t^1, \ldots, a_t^n\right) \\ \text{where } a_t^j &= \text{softmax}\Big(\tanh\Big(\Big[\mathbf{h}_t^j\Big]\mathbf{W}_1\Big)\mathbf{w}_2\Big) \end{aligned} \tag{12}$$

where β_t^j describes the significance of tth word for the comment c_j.

C. Sentence-Comment Co-Attention

Not every sentence is fake in the news contents, and as a matter of fact, few are genuine, but only to help fake ones [27]. As a result, news sentences must not be equally relevant in explaining whether news articles are fake or not. At the same time, user comments can provide details regarding the crucial components that explain why a portion of news is fake, but they can also be slightly noisy and informative.

Hence, we strive to choose user comments and news sentences that illustrate why a fraction of the news is fake. They should aid in detecting fake news since they give a clear explanation, indicating us to employ attention mechanisms to better understand the weight representations of comments and news sentences, which will aid in fake news identification. It motivates us to model attention mechanisms that provide better weights for representing sentences and comments in information, resulting in detecting fake news. In particular, we employ another round of multihead attention networks on top of the learned sentence and comment representations to better comprehend fine-grained semantic features as

$$\begin{aligned} H^s &= \text{Concat}\left(a_t^1, \ldots, a_t^m\right) \\ \text{where } a_t^i &= \text{softmax}\big(\tanh\big(\big[\alpha_t^i\big]\mathbf{W}_1\big)\mathbf{w}_2\big) \\ H^c &= \text{Concat}\left(b_t^1, \ldots, b_t^n\right) \\ \text{where } b_t^i &= \text{softmax}\Big(\tanh\Big(\Big[\beta_t^j\Big]\mathbf{W}_3\Big)\mathbf{w}_4\Big) \end{aligned} \tag{13}$$

where H^s and H^c represent the final sentence and comment representation, respectively. Next, we employ co-multihead attention to capture the semantic association between sentences and integrate them as follows:

$$\begin{aligned} \mathbf{H} &= \text{Concat}\left(a^1, \ldots, a^m\right) \\ \text{where } a_t^i &= \text{softmax}\big(\tanh\big(\big[\mathbf{H^s}; \mathbf{H}^c\big] \cdot \mathbf{W}_1\big) \cdot \mathbf{W}_2\big) \end{aligned} \tag{14}$$

where $\mathbf{W}_1, \mathbf{W}_2 \in \mathbb{R}^{2d \times 2\ d}$ are learnable parameters.

D. Fusion and Training

We have covered how to model the hierarchical structure of news material at the word and sentence levels, encode comments by employing multihead attention and combine comments and sentence representations. Then, we pass the combined representation $\mathbf{H}$ to a multilayer perceptron (MLP) to predict fake news as

$$\hat{y} = \sigma\left(\mathbf{W}_f \cdot (\mathbf{W} \cdot [\mathbf{H}] + \mathbf{b}) + \mathbf{b}_f\right) \tag{15}$$

where $\hat{y}$ indicates the predicted probability, where $\hat{y} = 1$ (fake news) and $\hat{y} = 0$ (real news), respectively. Hence, we strive to minimize the cross-entropy loss function for each news item as

$$\mathcal{L}(\theta) = -y \log(\hat{\mathbf{y}}_1) - (1 - y) \log(1 - \hat{\mathbf{y}}_0) \tag{16}$$

TABLE I
SIMULATION CONDITIONS OF OUR EXPERIMENTAL FAKENEWSNET DATASET

Dataset	Users	Comments	Candidate news	True news	Fake news
PolitiFact	68,523	89,999	415	145	270
GossipCop	156,467	231,269	5,816	3,586	2,230

where θ signifies the network parameters, we employ Adam [28], which delivers faster convergence than the SGD and avoids the challenge of adjusting the learning rate [29].

V. EXPERIMENTS

In this section, we perform experiments to assess the significance of our proposed method. We specifically intend to answer the following questions.

1) **Q1**: Can our proposed method, which models user comments and news content simultaneously, enhance fake news classification performance?
2) **Q2**: What influence do news articles and user comments have on the detection performance of our proposed method?
3) **Q3**: Is it possible for our proposed approach to apprehend the user comments and news sentences that indicate why a fraction of the news is fake?

A. Datasets

We use FakeNewsNet [1], [30], which is an exhaustive fake news detection benchmark dataset. The dataset is created from two fact-checking systems, PolitiFact and GossipCop, which provide news with social context information and labels. The meta properties of the news are included in news content, and the relevant user social interactions of news items are included in the social context. Table I displays the comprehensive statistics for dataset. We only preserve news items with at least three comments in our experiments.

B. Compared Methods

We compare our proposed method with the following representative baseline fake news detection algorithms:

1) **Rhetorical Structure Theory (RST)** [31] creates a tree configuration to describe rhetorical relationships among words and drags news style features to a vector space by associating the frequencies of rhetorical connections.[8]
2) **Linguistic Inquiry and Word Count (LIWC)** [32] extracts psycholinguistic lexicons and learns a feature vector based on deception and psychology perspectives.
3) **HAN** [24] utilizes a hierarchical attention network for fake news detection and encodes news content and every sentence with sentence-level word-level attention.
4) **Text-Convolutional Neural Networks (CNN)** [33] model news contents through a CNN and uses various convolution filters to capture various granularities of text characteristics.
5) **Two-Level Convolutional Neural Network with User Response Generator (TCNN-URG)** [16] employ a CNN and a conditional variational auto-encoder to learn news content and user comments representations.

[8]https://github.com/jiyfeng/DPLP

TABLE II
SIMULATION CONDITIONS OF OUR EXPERIMENTAL FAKENEWSNET DATASET

Method	Datasets							
	PolitiFact				GossipCop			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
LIWC	0.769	0.843	0.794	0.818	0.736	0.756	0.461	0.572
RST	0.607	0.625	0.523	0.569	0.531	0.534	0.492	0.512
HAN	0.837	0.824	0.896	0.86	0.742	0.655	0.689	0.672
TCNN-URG	0.837	0.775	0.812	0.764	0.787	0.695	0.442	0.538
text-CNN	0.653	0.678	0.863	0.76	0.739	0.707	0.477	0.569
HPA-BLSTM	0.846	0.894	0.868	0.881	0.753	0.684	0.662	0.673
CSI	0.827	0.847	0.897	0.871	0.772	0.732	0.638	0.682
dEFEND	0.904	0.902	0.956	0.928	0.808	0.729	0.782	0.755
Proposed	0.936	0.938	0.977	0.951	0.843	0.805	0.819	0.797

6) **Hierarchical Network with Social Attention and Bi-Directional Recurrent Neural Networks (HPA-BLSTM)** [5] learns news representation through different levels of user arrangements on social media and extract post features to model the attention weights.
7) **CSI** [3] employ an LSTM and Doc2Vec [34] based hybrid network to model the user comments and news contents using information from text, source and response.
8) **dEFEND** [6] employ a co-attention fake news detection algorithm to learn the relationship between sentences in the source article and user profiles.

For a fair comparison, we used the baseline methods with the following perspectives: 1) LIWC, RST, HAN, and text-CNN for only news contents; 2) HPA-BLSTM only for user comments; and 3) dEFEND, CSI, and TCNN-URG for both user comments and news content. We use a variety of learning algorithms, including, Decision Tree, Random Forest, Naive Bayes, and Logistic Regression, for LIWC and RST and present the best outcomes. For our proposed and baseline models, we employ scikit-learn [35] and Pytorch[9] for implementation. We selected 256 as the batch size, 200 as the epoch, and 0.001 as the learning rate. For all the methods, we use the max sentence length as 120, max sentence count as 50, max comment count as 150, max comment length as 120, and use Glove [36] as the word embedding technique for the baseline methods, set the d and k as 100 and 80, respectively, while keeping the vocabulary size at 20 000.

C. Q1. Fake News Classification Performance

We utilize the following measures to assess the efficiency of fake news detection algorithms: Recall, Precision, Accuracy, and $F1$. Randomly, we select 75% of news articles for training and the rest for testing. The procedure is repeated five times, with Table II showing the average performance. We draw the following observation from the results illustrated in Table II.

1) HAN outperformed the other news content-based framework for both datasets, i.e., LIWC and RST, indicating that it can efficiently capture semantic and syntactic signals through HANs to distinguish between fake and true news. Also, LIWC performs better than RST, indicating it can better model the linguistic features. The better LIWC results show that fake news differs from actual news in selecting words that reflect psychometric features.
2) Furthermore, approaches combining user comments and news material outperform methods that rely only on news content and solely on user comments. This implies that characteristics generated from news material and accompanying user comments include complementary information, which improves detection efficiency.
3) Furthermore, user comment-based approaches are marginally more promising than news content-based approaches. For both the Gossipcop and PolitiFact datasets, HPA outperforms both HAN and BLSTM in terms of $F1$ and Accuracy ratings. It demonstrates that features pulled from user comments have a higher discriminative potential for predicting fake news than features taken from news text alone.
4) In general, on both datasets and methods based on user comments and news content, we can observe that our proposed technique consistently beats dEFEND, TCNN-URG, and CSI in terms of all evaluation metrics. Compared to dEFEND in terms of $F1$ and Accuracy score, our proposed method yields an average relative improvement of 5.56%, 4.33% on Gossipcop, and 2.47%, respectively, 3.53% on PolitiFact. It demonstrates the value of using PRUs and modeling the co-attention of user comments and news sentences in identifying fake news.

D. Q2. Influence of News Contents and User Comments

Apart from the news contents, we also combine it with the features extracted from the user comments through a co-attention mechanism and examine the influence of these components even further by developing three variations of our proposed method.

1) *Proposed$_N$:* This version of our work does not consider news content information. It begins by learning comments features through the comment encoder and then feeding them to an MLP and a softmax layer for classification.
2) *Proposed$_{Co}$:* This variation does not use multihead attention on top of the sentence and comment learned representations and concatenates them directly by feeding them to an MLP and a softmax layer for classification, rather than using comment-sentence co-attention.

[9]https://pytorch.org/

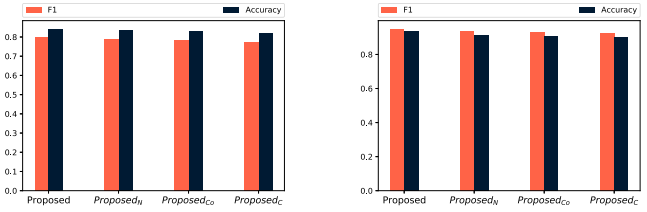

Fig. 2. Influence of various features on our proposed methods for detecting fake news. (a) GossipCop. (b) PolitiFact.

3) *Proposed$_C$:* This version of our work does not take into account details from user comments. It begins by encoding news contents with word-level attention and then feeds the consequent sentence characteristics into an MLP and a softmax layer for classification.

We used cross-validation to identify the parameters in all variations, and the best outcomes are reported in Fig. 2. From the outcomes, we derive the following.

1) The performance of Proposed$_N$ falls behind that of our proposed method when we remove the impact of news content. For instance, the Accuracy and $F1$ scores dropped by 1.2% and 1.33% on GossipCop and 3.6% and 2.9% on PolitiFact, suggesting that news contents are essential for good performance.
2) The performance is degraded when co-attention for user comments and news information is removed, suggesting that it is essential to model their correlation and apprehend the joint effect between user comments and news contents.
3) The performance of Proposed$_C$ falls behind our proposed method when we remove the impact of user comments, which indicates that it is essential to keep the user comments features for fine-grained fake news detection.

Following the above results, we discovered that user comments and news content help our method enhance its fake news detection performance. Both must be modeled since they provide supportive information.

E. Q3. Explainability Evaluation

To meet Q3, in this section, we assess the performance of our proposed method's explainability in terms of user comments and news content. Note that, except dEFEND in Section V-B, all other baseline techniques are intended for fake news detection. Neither is originally introduced to uncover explainable user comments or news sentences. To assess the performance of our proposed method for explainability, we select dEFEND for both user comments and news sentences and HAN as a baseline for news sentence explainability.

F. Explainability for News Sentence

We conducted experiments to assess the performance of our model for the explainability rank list, i.e., RS, in the context of news sentences. We are particularly interested in seeing if our method's top-ranked explainable sentences are highly possible to be connected to important sentences in fake news that are worth checking. Hence, for a portion of news content,

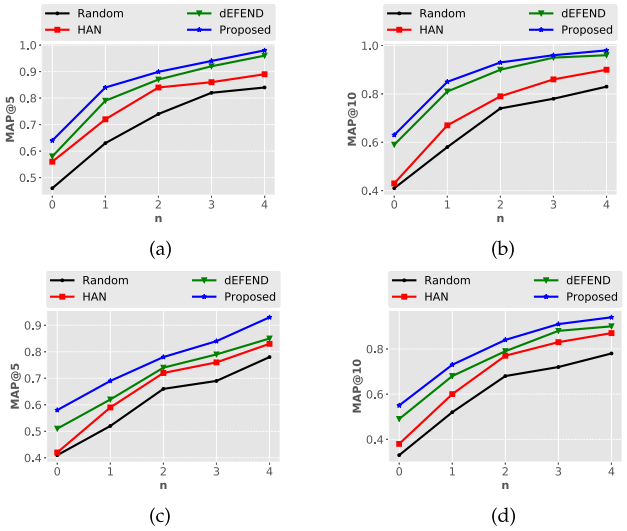

Fig. 3. Comparison of proposed and baselines methods for explainability of sentences on MAP at 5 and 10 regarding the neighborhood threshold n. MAP at (a) 5 on PolitiFact, (b) 10 on PolitiFact, (c) 5 on GossipCop, and (d) 10 on GossipCop.

we employ ClaimBuster [37] to create an $\hat{RS}$ of all check-worthy sentences. ClaimBuster presents a scoring model that employs multiple language elements and assigns a "check-worthiness" score between 1 and 0 based on hundreds of thousands of sentences tagged by human coders from previous general election discussions. Higher scores indicate that the sentence holds check-worthy factual sentences, while the lower describes the sentence as non-factual. We use the evaluation metric Mean Average Precision MAP@k to compare the top-k rank lists for explainable sentences produced by the proposed method (RS^1) and dEFEND (RS^2) and also with HAN (RS^3) with the top-k rank lists, $\hat{RS}$, produced by ClaimBuster. We set k as 5 and 10. In addition, while comparing the sentences in RS^1, RS^2, and RS^3 with $\hat{RS}$, we use the threshold n, which determines the size of window that permits n nearby sentences to be examined. From the results in Fig. 3, we derive the following.

1) On both datasets, we can observe that our proposed technique beats dEFEND, HAN, and Random in terms of discovering check-worthy sentences in news material, demonstrating that our method's sentence-comment co-attention component efficiently chooses more check-worthy sentences.
2) As n increases, we loosen the need to match ground truth and check worthy sentences, and therefore the MAP performance improves.
3) When setting $n = 1$, our proposed method's performance on MAP at 5 and 10 surpasses 0.8 for PolitiFact, demonstrating that our proposed method can find check-worthy statements even in the case of a single neighboring sentence from the $\hat{RS}$ ground truth sentences.

G. Explainability for User Comments

We use Amazon Mechanical Turk (AMT)[10] to launch multiple tasks, to assess the explainability of ranked lists of

[10]https://www.mturk.com/

comments RS for fake news. We use the following configurations for 50 fake news articles to launch AMT tasks. We start by removing items that are fewer than 50 words long from each news article. Furthermore, we offered only the first 500 words of long articles with more than 500 words to limit the portion of rendering required by workers. Because the initial three to four paragraphs of a news article outline the scope. Similarly, the first 500 words are generally enough to convey the story's substance. Following that, familiar with the articles and with an approval percentage of greater than 0.95, we hired AMT workers in the United States. We left with three lists of top-k comments for each news piece to evaluate the explainability of user comments: $L^{(1)} = (L_1^{(1)}, L_2^{(1)}, \ldots, L_k^{(1)})$ for utilizing our proposed method, $L^{(2)} = (L_1^{(2)}, L_2^{(2)}, \ldots, L_k^{(2)})$ and $L^{(3)} = (L_1^{(3)}, L_2^{(3)}, \ldots, L_k^{(3)})$ for HPA-BLSTM. We select the top-k comments, rank them from high to low based on attention weights, and set $k = 5$ to test the model's ability to choose the most explainable comments.

We compare lists one by one in Task 1. Workers are asked to choose between $L^{(1)}$, $L^{(2)}$, and $L^{(3)}$, a list that is collectively better. To eliminate position bias, we allocate $L^{(1)}$, $L^{(2)}$, and $L^{(3)}$ to workers in a random order, top, and bottom. For each news item, we allow each worker to choose the best list from $L^{(1)}$, $L^{(2)}$, and $L^{(3)}$. We made sure that each news item was reviewed by three workers and collected 150 findings based on the worker's selections. We calculate the number of workers who select $L^{(1)}$, $L^{(2)}$, and $L^{(3)}$ at the worker level, as well as the winning ratio (WR) for them. Furthermore, we conduct a majority vote across all three workers to determine if workers prefer $L^{(1)}$, $L^{(2)}$, or $L^{(3)}$ at the news level. We also calculate the worker-level options for each news by calculating $L^{(1)}$, $L^{(2)}$ and $L^{(3)}$. From the results in Fig. 4, we derive the following.

1) Both at the news worker and levels, our proposed method can choose superior-top-k explainable comments than dEFEND and HPA-BLSTM. At the first step and at worker level, 122 of 150 workers with a WR $=$ 0.81 select $L^{(1)}$ over $L^{(2)}$ and 134 of 150 workers with a WR $=$ 0.89 select $L^{(1)}$ over $L^{(3)}$. At the news level, our model outperforms dEFEND in 41 of 50 news articles with a WR $=$ 0.82 and HPA-BLSTM in 44 of 50 with a WR $=$ 0.88.
2) We can witness more news articles in which three workers vote unequivocally for $L^{(1)}$, i.e., 3 versus 0, compare to 0 versus 3 for both dEFEND and HPA-BLSTM for their explainability. In the same fashion, we encountered other instances where two workers voted for our proposed model than both dEFEND and HPA-BLSTM.

We conduct the item-wise evaluation for Task 2. We direct workers to select a score from 0 to 4 for every comment in $L^{(1)}$, $L^{(2)}$, and $L^{(3)}$, where 0 represents "not explainable at all," 1 symbolizes "not explainable," 2 describes "somewhere in between," 3 depicts "somewhat explainable," and 4 portrays "highly explainable." We rearrange the comments in $L^{(1)}$, $L^{(2)}$, and $L^{(3)}$ to prevent bias induced by distinct user criteria and direct workers to rate how explainable each remark is. We employ Precision@k and normalized cumulative

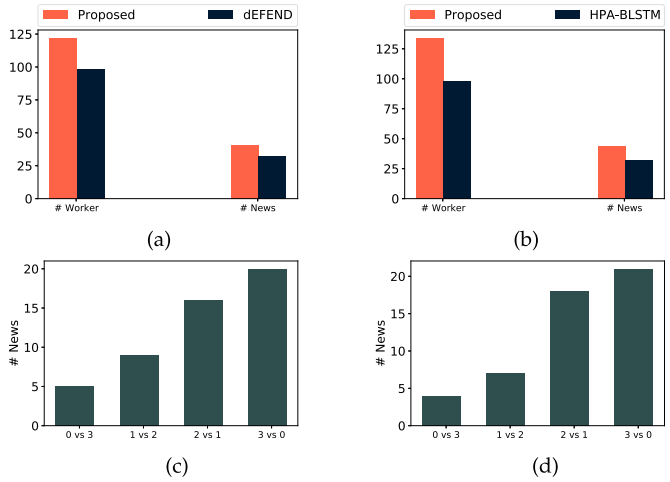

Fig. 4. Human-level evaluation of explainable comment list for Task 1. Winning count (a) proposed versus dEFEND and (b) proposed versus HPA-BLSTM. Worker voting ratio (c) proposed versus dEFEND and (d) proposed versus HPA-BLSTM.

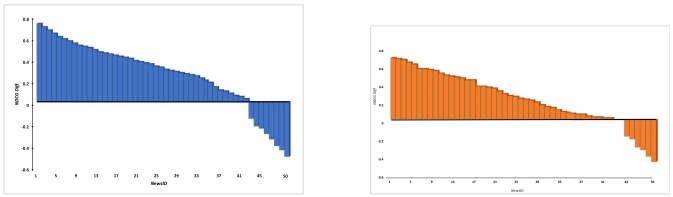

Fig. 5. Discrepancy histograms with respect to mean (a) NDCG and (b) mean Precision@5.

gain (NDCG) [38] as assessment standards to judge the rank-aware explainability of comments. NDCG evaluates the quality of ranks by comparing it to the ideal rank lists as determined through user feedback. The MAP@k signifies the percentage of relevant suggested elements in the top-k collection. Likewise, for each technique, we ensure that every news item is assessed by three workers, resulting in a total of 750 worker ratings. Fig. 5 shows the results arranged in descending order by the discrepancy in metrics among the three techniques. We only exhibit the results of MAP@5 with similar MAP@10 results, and we noticed the following.

1) In the item-wise evaluation of 50 fake news pieces, our proposed technique achieves higher NDCG ratings than dEFEND and HPA-BLSRM in 42 cases. Our proposed method, dEFEND and HPA-BLSRM, have overall mean NDCG scores of 0.84, 0.71, and 0.55 across 50 instances, respectively.
2) Precision@5 yields similar outcomes. On 42 fake news pieces, our model outperforms dEFEND and HPA-BLSTM, while two articles are tied. Our proposed method, dEFEND and HPA-BLSRM, has overall mean Precision@5 scores of 0.84, 0.67, and 0.51 across 50 instances.

H. Case Study

Fig. 6 illustrates a comparison between our proposed method and dEFEND, as well as the explainable comments that we properly scored high but that dEFEND did not

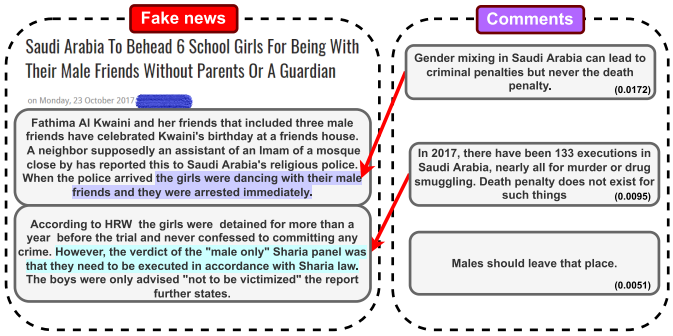

Fig. 6. Explainable comments captured by our proposed method.

recognize. The figure shows that the proposed method is capable of ranking a high number of explainable comments than non-explainable ones. For instance, the comment "Gender mixing in Saudi Arabia can lead to criminal penalties but never the death penalty" is top-ranked. It accurately explains why the following sentence from the news content is fake. "When the police arrived, the girls were dancing with their male friends, and they were arrested immediately." For this purpose, we can also provide high weights to explainable comments rather than intrusive or irrelevant comments. It can help select more relevant comments to detect fake news and other misinformation. For instance, comment 3 is irrelevant and has a lower attention weight of 0.0051 than explainable comment 2, which has a higher attention weight of 0.0095. The second comment is chosen as an essential attribute for fake news prediction.

VI. Conclusion

In recent years, there has been increasing interest in fake news detection. Nevertheless, it is equally critical to comprehend why a news element can be predicted as fake. We explore the issues of detecting explainable fake news as: 1) greatly enhancing detection performance and 2) uncovering explainable user's chats and discussions in the form of comments and news sentences to better understand why a news article is flagged as fake. The proposed methods for fake news detection to learn various representations and discover explainable comments and sentences. Experimental results on real-world datasets show that the proposed method is effective and explainable. In the future, fact-checking websites will be included to steer the learning mechanism better to acquire check-worthy news sentences. In addition, we will check the trustworthiness of those creating explainable comments to boost the effectiveness of fake news detection.

References

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

[2] H. C. Hughes and I. Waismel-Manor, "The Macedonian fake news industry and the 2016 US election," *Political Sci. Politics*, vol. 54, no. 1, pp. 19–23, Jun. 2021.

[3] N. Ruchansky, S. Seo, and Y. Liu, "CSI: A hybrid deep model for fake news detection," in *Proc. ACM Conf. Inf. Knowl. Manag.*, 2017, pp. 797–806.

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

[5] H. Guo, J. Cao, Y. Zhang, J. Guo, and J. Li, "Rumor detection with hierarchical social attention network," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manag.*, 2018, pp. 943–951.

[6] K. Shu, L. Cui, S. Wang, D. Lee, and H. Liu, "Defend: Explainable fake news detection," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2019, pp. 395–405.

[7] S. Mehta, R. Koncel-Kedziorski, M. Rastegari, and H. Hajishirzi, "Pyramidal recurrent unit for language modeling," 2018, *arXiv:1808.09029*.

[8] L. Wu and Y. Rao, "Adaptive interaction fusion networks for fake news detection," 2020, *arXiv:2004.10009*.

[9] Z. Guo, M. Schlichtkrull, and A. Vlachos, "A survey on automated fact-checking," *Trans. Assoc. Comput. Linguistics*, vol. 10, pp. 178–206, Feb. 2022.

[10] X. Zhou, R. Zafarani, K. Shu, and H. Liu, "Fake news: Fundamental theories, detection strategies and challenges," in *Proc. 12th ACM Int. Conf. Web Search Data Mining*, 2019, pp. 836–837.

[11] X. Zhou and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," *ACM Comput. Surv.*, vol. 53, no. 5, pp. 1–40, 2020.

[12] X. Zhang, J. Cao, X. Li, Q. Sheng, L. Zhong, and K. Shu, "Mining dual emotion for fake news detection," 2019, *arXiv:1903.01728*.

[13] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein, "A stylometric inquiry into hyperpartisan and fake news," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguist. Conf. (ACL)*, vol. 1, 2018, pp. 231–240.

[14] S. Hosseinimotlagh and E. E. Papalexakis, "Unsupervised content-based identification of fake news articles with tensor decomposition ensembles," in *Proc. Workshop Misinf. Misbehav. Mining Web (MIS)*, Feb. 2018, pp. 1–8.

[15] Z. Jin, J. Cao, Y. Zhang, J. Zhou, and Q. Tian, "Novel visual and statistical image features for microblogs news verification," *IEEE Trans. Multimedia*, vol. 19, no. 3, pp. 598–608, Mar. 2017.

[16] F. Qian, C. Gong, K. Sharma, and Y. Liu, "Neural user response generator: Fake news detection with collective user intelligence," in *Proc. IJCAI*, vol. 18, 2018, pp. 3834–3840.

[17] Y. Wang *et al.*, "EANN: Event adversarial neural networks for multi-modal fake news detection," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2018, pp. 849–857.

[18] F. Yang *et al.*, "XFake: Explainable fake news detector with visualizations," in *Proc. World Wide Web Conf.*, 2019, pp. 3600–3604.

[19] G. B. Guacho, S. Abdali, N. Shah, and E. E. Papalexakis, "Semi-supervised content-based detection of misinformation via tensor embeddings," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2018, pp. 322–325.

[20] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on Twitter," in *Proc. 20th Int. Conf. World Wide Web*, 2011, pp. 675–684.

[21] S. Volkova, K. Shaffer, J. Y. Jang, and N. Hodas, "Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on Twitter," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics*, vol. 2, 2017, pp. 647–653.

[22] M. Du, N. Liu, and X. Hu, "Techniques for interpretable machine learning," *Commun. ACM*, vol. 63, no. 1, pp. 68–77, Dec. 2019.

[23] Z. Jin, J. Cao, Y. Zhang, and J. Luo, "News verification by exploiting conflicting social viewpoints in microblogs," in *Proc. AAAI Conf. Artif. Intell.*, vol. 30, 2016, pp. 1–7.

[24] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2016, pp. 1480–1489.

[25] H. Chen, M. Sun, C. Tu, Y. Lin, and Z. Liu, "Neural sentiment classification with user and product attention," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 1650–1659.

[26] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, Y. Bengio and Y. LeCun, Eds., San Diego, CA, USA, May 2015, pp. 1–19.

[27] A. Alharbi, H. Dong, X. Yi, Z. Tari, and I. Khalil, "Social media identity deception detection: A survey," *ACM Comput. Surv.*, vol. 54, no. 3, pp. 1–35, 2021.

[28] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, Y. Bengio and Y. LeCun, Eds., San Diego, CA, USA, May 2015, pp. 1–10.

[29] S. T. U. Shah, J. Li, Z. Guo, G. Li, and Q. Zhou, "DDFL: A deep dual function learning-based model for recommender systems," in *Proc. Int. Conf. Database Syst. Adv. Appl.* Cham, Switzerland: Springer, Sep. 2020, pp. 590–606.
[30] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "FakeNewsNet: A data repository with news content, social context and spatialtemporal information for studying fake news on social media," 2018, *arXiv:1809.01286*.
[31] V. L. Rubin, N. J. Conroy, and Y. Chen, "Towards news verification: Deception detection methods for news discourse," in *Proc. Hawaii Int. Conf. Syst. Sci.*, Jan. 2015, pp. 5–8.
[32] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The development and psychometric properties of LIWC2015," Univ. Texas Austin', Austin, TX, USA, 2015.
[33] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, A. Moschitti, B. Pang, and W. Daelemans, Eds., Oct. 2014, pp. 1746–1751.
[34] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 1188–1196.
[35] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12 no. 10, pp. 2825–2830, 2012.
[36] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors for word representation," in *Proc. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1532–1543.
[37] N. Hassan, F. Arslan, C. Li, and M. Tremayne, "Toward automated fact-checking: Detecting check-worthy factual claims by claimbuster," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2017, pp. 1803–1812.
[38] K. Järvelin and J. Kekäläinen, "Cumulated gain-based evaluation of IR techniques," *ACM Trans. Inf. Syst.*, vol. 20, no. 4, pp. 422–446, Oct. 2002.

Fazlullah Khan (Senior Member, IEEE) is currently a Faculty Member with the Department of Computer Science, Abdul Wali Khan University Mardan, Mardan, Pakistan. He has published his research work in top-notch journals and conferences. His research has been published in IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS (TII), IEEE INTERNET OF THINGS JOURNAL, IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING (TNSE), IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS (TCSS), IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING (TGCN), IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS (ITS), IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS (JBHI), *Future Generation Computer Systems* (Elsevier), *Journal of Network and Computer Applications* (Elsevier), *Mobile Networks and Applications* (Springer), and *Neural Networks and Applications* (Springer). His research interests include intelligent and robust protocol designs, security and privacy of wireless communication systems, the Internet of Things, machine learning, artificial intelligence, and intelligent transportation systems. Recently, he has been involved in the latest developments in the field of the Internet of Vehicles security and privacy issues, software-defined networks, fog computing, and big data analytics.

Dr. Khan had been a recipient of various prestigious scholarships during his Ph.D. studies.

Ryan Alturki (Senior Member, IEEE) received the Ph.D. degree from the University of Technology, Sydney, NSW, Australia, in 2019.

He is currently an Assistant Professor with the Department of Information Science, College of Computers and Information Systems, Umm Al-Qura University, Mecca, Saudi Arabia. He has published several publications in high-ranked international journals, conferences, and chapter of books. His research interests include eHealth, mobile technologies, the Internet of Things (IoT), artificial intelligence, cloud computing, deep learning, natural language processing, and cybersecurity.

Gautam Srivastava (Senior Member, IEEE) received the B.Sc. degree from Briar Cliff University, Sioux City, IA, USA, in 2004, and the M.Sc. and Ph.D. degrees from the University of Victoria, Victoria, BC, Canada, in 2006 and 2012, respectively.

He is currently an Associate Professor with Brandon University, Brandon, MB, Canada. His areas of interest are data mining, big data, and security.

Foziah Gazzawe (Member, IEEE) received the B.Sc. degree in information science from Umm Al-Qura University, Mecca, Saudi Arabia, in 2013, the master's degree from Claremont Graduate University, Claremont, CA, USA, in 2016, and the Ph.D. degree in software engineering from Loughborough University, Loughborough, U.K., in 2020.

She is currently an Assistant Professor with the Department of Information Science, Umm Al-Qura University. Her current research interests include software engineering, socio-technical systems, conceptual modeling, ontology, computing, deep learning, and natural language processing.

Syed Tauhid Ullah Shah (Student Member, IEEE) received the master's degree in computer applied technology from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2020. He is currently pursuing the Ph.D. degree with the Department of Electrical and Software Engineering, Schulich School of Engineering, University of Calgary, Calgary, AB, Canada.

His work has been published in various IEEE conferences, journals, and TRANSACTIONS. His research interests include machine learning, deep learning, natural language processing, recommender systems, and the Internet of Things.

Spyridon Mastorakis (Member, IEEE) received the M.Eng. degree in electrical and computer engineering from the National Technical University of Athens, Athens, Greece, in 2014, and the M.S. and Ph.D. degrees in computer science from the University of California at Los Angeles, Los Angeles, CA, USA, in 2017 and 2019, respectively.

Since August 2019, he has been an Assistant Professor in computer science with the University of Nebraska at Omaha, Omaha, NE, USA. His research has been published at premier conferences and journals, including the IEEE International Conference on Pervasive Computing and Communications (PerCom), the IEEE International Conference on Computer Communications (INFOCOM), the IEEE International Conference on Distributed Computing Systems (ICDCS), the IEEE INTERNET OF THINGS JOURNAL, and the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS. His research interests include network systems and protocols, Internet architectures, the Internet of Things (IoT) and edge computing, and security.

Dr. Mastorakis is a member of Association for Computing Machinery (ACM) and a fellow of the U.S. National Strategic Research Institute (NSRI). He has served as a TPC Member for IEEE International Conference on Network Protocols (ICNP) 2022, IEEE Global Communications Conference (GLOBECOM) 2022, and IEEE International Conference on Communications (ICC) 2023, and an Associate Editor for the IEEE INTERNET OF THINGS JOURNAL.