IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 23 May 2024, accepted 15 June 2024, date of publication 24 June 2024, date of current version 1 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3418340

RESEARCH ARTICLE

An Enhanced Fake News Detection System With Fuzzy Deep Learning

CHENG XU AND M-TAHAR KECHADI, (Member, IEEE)
School of Computer Science, University College Dublin, Dublin 4, D04 V1W8 Ireland

Corresponding author: Cheng Xu (cheng.xu1@ucdconnect.ie)

This work was supported in part by the Science Foundation Ireland under Grant SFI/12/RC/2289_P2, and in part by the Insight Centre for Data Analytics and China Scholarship Council under Grant 202208340012.

ABSTRACT Addressing the intricate challenge of fake news detection, traditionally reliant on the expertise of professional fact-checkers due to the inherent uncertainty in fact-checking processes, this research leverages advancements in language models to propose a novel fuzzy logic-based network. The proposed model is specifically tailored to navigate the uncertainty inherent in the fake news detection task. The evaluation is conducted on the well-established LIAR dataset, a prominent benchmark for fake news detection research, yielding state-of-the-art results. Moreover, recognizing the limitations of the LIAR dataset, we introduce LIAR2 as a new benchmark, incorporating valuable insights from the academic community. Our study presents detailed comparisons and ablation experiments on both LIAR and LIAR2 datasets and establishes our results as the baseline for LIAR2. The proposed approach aims to enhance our understanding of dataset characteristics, contributing to refining and improving fake news detection methodologies.

INDEX TERMS Deep learning, fuzzy deep learning, fake news, fake news detection, fact-checking, NLP, classification systems, benchmark.

I. INTRODUCTION

The advent of the digital age has transformed the Internet, mainly through social media, into an indispensable source of information and news. This medium offers instantaneous access to vast amounts of information, fostering global connectivity and awareness in our swiftly evolving world. However, the escalating prevalence of fake news poses a significant threat, disrupting social order, undermining media credibility, and jeopardizing democratic processes [1], [2], [3], [4]. Research indicates that fake news spreads rapidly and wields more significant influence than authentic news [5], [6], [7]. The increasing sophistication and ubiquity of disinformation campaigns further compound the challenge, making it progressively challenging to discern trustworthy news sources from deceptive ones [8]. Consequently, there is an urgent need to develop reliable and robust tools and techniques for detecting and mitigating fake news.

Fact-checking stands out among text-related challenges due to its inherent complexity, primarily stemming from the difficulty of enabling machines to comprehend nuanced statements and make informed decisions about their accuracy. The formidable challenge posed by fake news emanates from highly sophisticated organizations employing diverse formats, such as text, images, videos, and audio, with multi-faceted objectives and tactics. Machine learning techniques, such as deep learning, have garnered substantial attention and hold great promise for automated fake news detection [9], [10], [11]. However, existing studies often operate within specific contexts, making comparisons or extensions to different types of misinformation challenging. The intricate nature of language, coupled with the multi-modal aspects of fact-checking, presents a formidable task for automated systems.

Unlike straightforward text classification, fact-checking demands understanding context, intent, and the ability to cross-reference information from diverse sources. It emphasizes the necessity of considering multiple perspectives and verifying claims through various reliable channels, acknowledging that a comprehensive evaluation of the truthfulness of a statement cannot be based on a singular viewpoint. Consequently, addressing the issue of datasets or

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

benchmarks becomes crucial for enabling fair comparisons among proposed approaches and extracting common insights into the challenges inherent in specific benchmarks. Over the past five years, several benchmarks, including FCN-1 [12], Grover-Mega [13], LIAR [14], Politifact [15], COVID-19 Fake News [16], and others, have been introduced. Numerous techniques have been developed and implemented to tackle the distinctive characteristics of each benchmark effectively. The intricate interplay of language nuances, diverse sources, and the dynamic nature of information make fact-checking a uniquely complex challenge that underscores the limitations of simplistic, one-sided news and information verification assessments.

The problem of detecting fake news is inherently complex due to several intertwined factors. Fake news can take many forms, from completely fabricated stories to partially factual statements presented in misleading contexts. The intent behind these statements can vary, ranging from deliberate deception to unintentional misinformation, complicating the task of detection. Moreover, the context in which a statement is made plays a crucial role in determining its veracity, requiring models to understand historical and current events deeply. The dynamic nature of languages and the subtlety of rhetorical devices further add to the difficulty, as does the presence of inherent biases in data labeling and source verification. Especially in political scenarios, different interpretations and the subjective nature of truth in political discourse exacerbate the challenge, as statements often carry multiple layers of meaning and implication. Consequently, any model designed to detect fake news must be able to handle these complexities, necessitating sophisticated algorithms capable of nuanced analysis and contextual understanding.

The complexity of fact-checking is evident in the structure of the model, which must adeptly handle multiple streams of information and leverage these processed representations for fake news detection. Moreover, given the inherently subjective nature of this task, human decision-making can often be ambiguous. For instance, consider a news headline proclaiming, "*Study shows that daily chocolate consumption improves memory.*" While this assertion may have some truth, it could be exaggerated or lacking crucial context. The referenced study might indeed exist and demonstrate a correlation between chocolate intake and memory enhancement. Still, critical details like the extent of the effect, the specific chocolate types and quantities required, and potential confounding variables may not be fully elucidated. Hence, rigid categorization of such news as true or fake is not always ideal; instead, a fuzzy logic approach might offer greater suitability for fake news detection than traditional probabilistic methods. Fuzzy logic provides better interpretability in decision-making inferences, accommodating the nuanced uncertainties in assessing news veracity. However, despite its potential advantages, practical applications of fuzzy logic in fake news recognition remain scarce in current literature.

This study addresses the LIAR benchmark, an influential metric in fake news detection, by adopting a distinctive learning approach. The proposed approach integrates fuzzy logic [17], [18] into deep learning methodologies [19], [20], [21]. Remarkably, the application of hybrid learning to fake news detection has been relatively scarce [22]. In response, we introduce an efficient model termed the **F**uzzy **D**eep **H**ybrid **N**etwork (FDHN[1] [23]), which effectively combines fuzzy logic with deep learning. This model utilizes both news articles and contextual information to enhance the performance of fake news detection. Notably, when evaluated on the LIAR dataset, the FDHN model demonstrates superior results compared to state-of-the-art approaches.

In addition to the LIAR dataset, we developed the LIAR2 dataset,[2] an enhanced and enlarged version of the LIAR benchmark, to improve the accuracy of fake news detection. The LIAR2 dataset incorporates the feedback from the research community on the LIAR dataset in recent years and improves its structure and quality. We also conducted experiments on the LIAR2 dataset using FDHN and released it together as a baseline.

The remainder of the paper is organized as follows: Section II provides a comprehensive review of pertinent studies, setting the foundation for our research. In Section III, we briefly introduce the dataset employed. Following that, Section IV presents the FDHN model, detailing its architecture and conducting a thorough analysis. The model's efficacy is then compared with other state-of-the-art models on the LIAR dataset, with additional experiments reported on the LIAR2 dataset in Section V. Section VI concludes the paper by discussing the strengths and weaknesses of the FDHN model, elucidating the reasons behind its state-of-the-art performance, and exploring potential future directions for both the FDHN model and the LIAR2 dataset in Section VII.

II. RELATED WORK

Fake news refers to false or misleading information presented as news. It encompasses various forms, including misinformation, disinformation, propaganda, and hoaxes [11]. Although false news has existed throughout history, the term "fake news" gained prominence in the 1890s when dramatic reports in newspapers were common [24]. However, the definition of fake news remains fluid and has been broadly applied to any false information presented as news. With the rise of social media like Facebook, especially in political scenarios, which are hardly hit by misleading information, fake news has become more prevalent and can undermine trust in credible media coverage [25]. In summary, we defined the term fake news here: *fake news is intentionally false or misleading information masquerading as legitimate news.*

In the past, fake news detection was done manually by fact-checkers and experts [26]. It was a labor-intensive

[1] The FDHN model is available at https://github.com/chengxuphd/FDHN.
[2] The LIAR2 is available at https://github.com/chengxuphd/liar2.

process that could not keep up with the rapid spread of misinformation on social media. As fake news proliferated rapidly, researchers have dedicated significant efforts to developing detection techniques using machine learning (ML). These techniques leverage natural language processing (NLP), text analysis, and other advanced methods to detect patterns and characteristics of fake news. The origins of automatic fake news detection can be traced back to text classification, a sub-field of NLP. Early work focused on developing classifiers distinguishing fake from authentic news articles based on textual features [27].

As the field evolved, researchers broadened their scope by incorporating additional modalities, such as context and transmission chains, to create more comprehensive detection frameworks. Castillo et al. [28] proposed an automatic approach that combined textual analysis with contextual information, emphasizing the importance of considering broader contextual elements, such as re-posting and citations, in identifying fake news.

The diversification of fake news detection led to the emergence of various methods, evolving into relatively independent system-level applications. Zhou et al. [11] categorized solutions into knowledge-based, style-based, propagation-based, and source-based, representing diverse approaches utilizing different news-related information for fake news detection. Notable works include news environment perception (NEP) [29], att-RNN [30], and EANN [31], which leverage multimodal information for improved results [32], [33], [34].

The infusion of fuzzy logic, a potent computational paradigm, into fake news detection has garnered considerable attention in recent years. Researchers have explored its integration into both machine learning and deep learning frameworks. Das et al. [21] divided approaches that combine deep learning with fuzzy logic into two main categories: integrated models, which embed fuzzy logic within the neural network, exemplified by models like fuzzy restricted Boltzmann machine [35] and Pythagorean fuzzy deep Boltzmann machine [36]; and ensemble models that apply fuzzy logic to both the upstream and downstream of the neural network [37], [38] (fuzzy input) and [39] (fuzzy output). Notably, the fused fuzzy deep neural network (FDNN) [20] considers the fuzzy logic module as an independent entity, combining fuzzy and neural representations to construct a classifier. This concept shares similarities with the approach employed in deep residual networks (ResNet) [40], where input data is fused with output data to enhance the model's robustness and stability. The structural inspiration drawn from these works has played a role in shaping the design of the proposed model.

Despite advancements in integrating fuzzy logic into various domains, its application in fake news detection remains relatively unexplored. The subjective nature of assessing news authenticity, which is inherently fuzzy in its clues, makes fuzzy logic more suitable and capable of handling the inherent fuzziness and reducing noisy data [41]. While some attempts, like Yang et al.'s application of fuzzy logic to detect COVID-19 rumors [42] and Gedara et al.'s proposed fuzzy transform-based LSTM network [43], exist, the efficiency and flexibility of traditional fuzzy logic methods remain a concern [44].

In summary, while existing literature addresses the challenges of fake news detection using traditional machine learning or deep learning techniques, these often struggle to model the inherent fuzziness and uncertainty present in natural languages. Motivated by the potential advantages of integrating fuzzy logic and deep learning, this paper introduces the FDHN model, aiming to harness these benefits for effective fake news detection—a critical issue in today's digital society.

In addition to detection models, we must pay attention to datasets. Fake news detection is challenging and requires various datasets for evaluation [45]. Many researchers have worked on this, e.g., the FEVER dataset [46], which consists of 185,445 claims verified against Wikipedia pages. It can be used for a complex pipeline of tasks that involves retrieving evidence, extracting facts, and verifying claims. The FakeNewsNet dataset [47] collects multimodal news data from two sources, BuzzFeed and PolitiFact, and it also provides social media information such as user engagements, comments, and propagation networks. It can be used for source credibility, user behavior, and social network analysis. The FNC-1 dataset [12] is another dataset based on human fact-checking and stance detection, and it contains 75,385 headline and article pairs labeled as agree, disagree, discuss, or unrelated. Apart from these datasets, the most far-reaching is the LIAR dataset, which has been adopted by numerous research institutes mainly due to its high data quality and ordered categorical labeling [48], [49], [50], [51], [52]. Details of the LIAR dataset, as well as our released LIAR2 dataset, are provided in Section III.

III. DATASET AND PROBLEM STATEMENT

We firstly evaluate the proposed model on the LIAR dataset,[3] a widely used real-world benchmark dataset for

[3]https://www.cs.ucsb.edu/ william/data/liar_dataset.zip

TABLE 1. The LIAR and LIAR2 dataset statistics (Num).

Statistics	LIAR	LIAR2
Training set size	10,269	18,369
Validation set size	1,284	2,297
Testing set size	1,283	2,296
Avg. statement length (tokens)	17.9	17.7
Avg. speaker description length (tokens)	\	39.4
Avg. justification length (tokens)	\	94.4
Labels		
Pants on fire	1,050	3,031
False	2,511	6,605
Barely-true	2,108	3,603
Half-true	2,638	3,709
Mostly-true	2,466	3,429
True	2,063	2,585

TABLE 2. An example of random selection from the LIAR and LIAR2 datasets.

#	Field	LIAR (Value)	LIAR2 (Value)
–	id	767.json	737
–	label	Barely-true	barely-true
f_1	statement	Sen. McCain's tax plan provides...	Sen. McCain's tax plan provides...
f_2	date	\	October 2, 2008
f_3	subject	taxes	taxes
f_4	speaker	joe-biden	joe biden
f_5	job_title	U.S. senator	\
f_6	speaker_description	\	Joe Biden is the president of...
f_7	state_info	Delaware	national
f_8	party_affiliation	democrat	\
f_9	true_counts	\	25
f_{10}	mostly_true_counts	16	64
f_{11}	half_true_counts	21	65
f_{12}	barely_true_counts	11	52
f_{13}	false_counts	10	55
f_{14}	pants_on_fire_counts	4	7
f_{15}	context	vice presidential debate in St. Louis	vice presidential debate in St. Louis
f_{16}	justification	\	If you look down the road to 2012...

fake news research, which was initially presented in 2017 by Wang [14]. We then evaluate the same on the proposed LIAR2 dataset. The statistical information of these datasets is shown in Table 1, with an example in Table 2. More detailed information on the LIAR and LIAR2 datasets is provided in Sections III-A and III-B, respectively.

A. LIAR DATASET

The LIAR dataset is composed of $\sim$ 12.8K human-labeled U.S. politics-related short statements by fact-checking experts with their corresponding contextual information, which the author obtained from PolitiFact.[4] In this benchmark, the author partitioned the dataset in the ratio of 8:1:1 into a training set, a validation set, and a test set for the actual task.

For the LIAR dataset, let $\{f_1, f_3, f_4, f_5, f_7, f_8, f_{10}, \cdots, f_{15}\}$ denote the set of features, as illustrated in Table 2. The feature f_1 corresponds to the statement or the news text, representing the primary content subject to truthfulness analysis, forming the core of this study. Features $\{f_3, f_4, f_5, f_7, f_8, f_{15}\}$ encompass textual context information. Specifically, f_3 represents the subject of the statement, providing insight into the topic or area of interest. f_4 denotes the speaker, indicating the statement's source or authorship. The speaker's job title is captured in f_5, adding an extra layer of professional context. Feature f_7 signifies state information, offering geographical context, while the party affiliation of the speaker is encapsulated in f_8. Lastly, f_{15} contains supplementary context, further enriching the understanding of the statement's circumstances.

The set of features $\{f_{10}, \cdots, f_{14}\}$ pertains to numerical context information, specifically the historical truthfulness record of the speaker. These features encompass the count of statements previously labeled as 'Mostly-true' (f_{10}), 'Half-true' (f_{11}), 'Barely-true' (f_{12}), 'False' (f_{13}), and 'Pants-on-fire' (f_{14}), respectively. This historical context serves to underscore the speaker's credibility, providing a crucial perspective in this study.

It is noteworthy that the dataset exhibits a well-balanced distribution based on the truthfulness of the news, categorized into six classes: 'Pants-on-fire', 'False', 'Barely-true', 'Half-true', 'Mostly-true,' and 'True.' This multi-class classification aligns more closely with real-world scenarios than datasets that dichotomize news into only two classes (true and false), serving as a primary motivation for our dataset selection.

B. LIAR2 DATASET

The LIAR dataset has been widely used by fake news detection researchers since its release, and along with a great deal of research, the community has provided a variety of feedback on the dataset to improve it; for example, Alhind et al. proposed the LIAR-PLUS dataset in 2018 [49], which is an extension of the LIAR dataset and has been used several times in subsequent studies [53], [54], [55]. The additional external contextual information (Justification) of the LIAR-PLUS dataset makes the same LIAR dataset models perform better. The justification was taken from the assessment report provided by the fact checker for each statement (f_{16}), which included the reasons why the statement was tagged in a certain category and removed all the words in it that could have led to label leakage.

For example, consider the statement of Donald Trump: "*Goodyear Tires announced a BAN ON MAGA HATS…This is what the Radical Left Democrats do.*" The accompanying Justification for this statement clarifies that "*Trump said Goodyear "announced a ban on MAGA hats" and said that "this is what the Radical Left Democrats do. "This distorts Goodyear's dress code and makes it sound as though the company singled out Trump's campaign messaging. Goodyear has not announced a ban only on hats that display the Trump campaign's MAGA slogan. Rather, the company requires employees to refrain from any political expression that falls "outside the scope of racial justice and*

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

TABLE 3. Comparison table of modalities covered by LIAR and LIAR2, where • stands for perfect coverage, ∘ for imperfect coverage, and × for complete exclusion.

	Statement	Time	Geographic	Speaker	History	Context	Justification
LIAR	•	×	∘	∘	∘	•	×
LIAR2	•	•	•	•	•	•	•

equity issues." Campaign paraphernalia supporting Trump's challenger, former Vice President Joe Biden, is also off-limits by this measure. We rate this statement Mostly False".

After removing keywords such as "*mostly*," "*false*," etc., which can lead to label leakage, the inclusion of such detailed justifications is instrumental in enhancing fake news detection accuracy. By providing the rationale behind each statement's categorization, these justifications enable the model to better understand the nuances of the news content and make more informed decisions. Additionally, they serve as a safeguard against potential biases or misleading interpretations, ensuring that the model's classifications are grounded in factual evidence rather than speculative assumptions. In essence, the incorporation of contextual information through justifications aligns with the need for comprehensive and nuanced analysis, thereby improving the reliability and effectiveness of the detection process."

Kirilin and Strub introduced Speaker2Credit into the LIAR dataset, which is making the reputation history of news producers more complete [56]. The essence is that the original LIAR dataset does not record the historical number of true statements made by speakers (f_9), and in this work, the authors introduced this feature into the LIAR dataset, and then better results were achieved.

We considered these two works to be of interest and adopted them in the development of the LIAR2 dataset, resulting in the features f_{16} and f_9 in Table 2, respectively. In addition to these, we have made the following enhancements to the original LIAR dataset:

- **Comprehensive Speaker Description:** We removed the job_title (f_5) and party_affiliation (f_8) features of the original speaker and replaced them with the speaker's description (f_6). The reason for this is that we have found that over time, many speakers have moved out of a position or changed their party affiliation, which can lead to outdated information about the speaker, so we have used the full description of the speaker. For example, Donald Trump was affiliated with the Democratic Party from 2001 to 2009 and with the Republican Party after 2009. It's not good enough for us to categorize him as a Democrat or a Republican during that time. Another example would be Joe Biden, who was a senator and later held the offices of vice president and president, and it's not objective to use any of those as his job of record. Using full speaker descriptions removes this inaccuracy, such as Joe Biden's description, "*Joe Biden is the president of the United States and is running for re-election in 2024. A Democrat, Biden served as a Delaware senator from 1973 (elected at the age of 29) until 2009. During his time in the Senate, he served as chair of the Foreign Relations Committee, chair of the International Narcotics Control Caucus, and chair of the Judiciary Committee. He subsequently served as Barack Obama's vice president from 2009 to 2016. Biden launched unsuccessful presidential campaigns in 1988 and 2008. He received a degree in history and political science from the University of Delaware and his J.D. from Syracuse School of Law.*" Additionally, we found a high prevalence of missing job_title and party_affiliation in speakers (more than a quarter), implying that these features are not as informative as they could be.
- **Incorporation of Time Information:** With the inclusion of a full speaker description (f_6), figuring out which period of time the statement belongs to the speaker has to incorporate temporal information, so we have included a date feature here (f_2). We believe that this provides further information about the context in which the statement was made.
- **Correction of Geographic Information:** We changed the basis for state_info (f_7). In the original LIAR dataset, the feature appeared as information about where the speaker is or has been working, e.g., Barack Obama was a senator from Illinois, so his record for the feature is Illinois. On the basis of this, it means that the information for this feature largely overlaps with the speaker. This makes the feature less informative. So, we decided to change the feature to the state information that the statement is about, e.g., a statement about a national policy would have its record value as "*national*," whereas a statement about a policy in New York State would be recorded as "*New York*". This seemed to make more sense and better suited to reflect the context of the statement.
- **Correction of Error Cases:** Furthermore, we have fixed some errors that appeared in the original LIAR dataset. For example, in the PolitiFact database, there exists another rating system called Flip-O-Meter,[5] which is used to respond to whether or not a speaker's attitude shifts before or after the same thing, which is not part of the dataset we need to cover, but it was erroneously included in the original LIAR dataset, and we filtered out such instances.

Based on these changes, we built the LIAR2 dataset, a new benchmark dataset of $\sim$ 23k manually labeled by professional fact-checkers for fake news detection tasks. The informative enhancement of LIAR2 with respect to LIAR can be summarized more compactly in Table 3, where the enhancement of LIAR2 is in almost all modalities.

[5]https://www.politifact.com/article/2008/aug/05/introducing-flip-o-meter/

TABLE 4. The Truth-O-Meter rating used by PolitiFact.

Rating	Short Description
True	Factually accurate and truthful statements.
Mostly True	Mostly accurate but may have minor inaccuracies.
Half True	Partially True but contains some misleading elements.
Mostly False	Mostly False with elements of truth.
False	False or significantly inaccurate statements.
Pants on Fire	Outright false, often with an intent to deceive.

We have used a split ratio of 8:1:1 to distinguish between the training set, the test set, and the validation set, details of which are provided in Table 1. Similar to the LIAR dataset, let $\{f_1, \cdots, f_4, f_6, f_7, f_9, \cdots, f_{16}\}$ denote the set of the LIAR2 dataset features, as shown in Table 2. The feature f_1 represents the speaker's statement. The features from $\{f_2, f_3, f_4, f_6, f_7, f_{15}\}$ consist of the textual context information. The set of features $\{f_9, \cdots, f_{14}\}$ represents the numerical context information that contains the credibility history of the speaker. And f_{16} denotes the justification for classifying the statement.

During the development of the LIAR2 dataset, we used the same data source as the LIAR, i.e., in collaboration with PolitiFact, and then got the data directly from their database. PolitiFact is a renowned fact-checking organization that plays a crucial role in scrutinizing the accuracy of statements made by public figures, politicians, and media outlets. Established as a non-partisan initiative, PolitiFact employs a rigorous methodology to evaluate the truthfulness of claims, assigning ratings ranging from "True" to "Pants on Fire" based on the veracity of the information, and the labels are detailed in the Table 4. To arrive at these ratings, PolitiFact's team of experienced journalists and fact-checkers meticulously researched and verified claims using a variety of reliable sources. The organization places a strong emphasis on providing readers with detailed explanations and evidence supporting their ratings, fostering an informed public discourse by presenting the facts behind the statements in question. This commitment to transparency and thorough analysis distinguishes PolitiFact as a leading authority in the field of fact-checking, their specific evaluation flow[6] is summarized as follows:

1) **Ask for Evidence:** When PolitiFact fact-checks a claim, they directly ask the person making the statement for evidence. Whether it's a politician or an advocate, they inquire about the specific basis for the claim. If the claim is viral or online, they carefully examine available information, including citations or fine print on images and political ads.
2) **Check Previous Fact-Checks:** PolitiFact leverages the wealth of fact-checking resources available today. They search for earlier fact-checks on similar topics, which provide sources that can be independently confirmed. Previous work often offers essential background information and pointers to experts or other evidence. Google's Fact-check Explorer is a valuable tool for this purpose, focusing on stories by established fact-checkers.
3) **Basic and Advanced Internet Search:** PolitiFact starts with a basic Google search to ensure they see what most people would find. However, for claims related to obscure topics, they delve deeper. Advanced searches involve specific phrases, numbers, languages, and elaborated search operators. They explore multiple search engines, including Bing and DuckDuckGo, to uncover relevant information.
4) **Review Multiple Sources:** PolitiFact's fact-checking process includes reviewing transcripts, speeches, news stories, press releases, campaign brochures, and social media content. They consult experts and examine publications to ensure thorough reporting. Their goal is to provide citizens with accurate information to make informed decisions in a democracy.
5) **Truth-O-Meter Ratings:** PolitiFact assigns ratings using their "Truth-O-Meter." These ratings include True, Mostly True, Half True, Mostly False, False, and Pants on Fire.
6) **Campaign Promise Tracking:** PolitiFact also tracks campaign promises made by politicians, holding them accountable for their statements over time.

The LIAR2 dataset is an upgrade of the LIAR dataset, which inherits the ideas of the LIAR dataset, refines the details and architecture, and expands the size of the dataset to make it more responsive to the needs of fake news detection tasks. We believe that with the help of the LIAR2 dataset, it will be able to perform better fake news detection tasks. More analysis and baseline information about the LIAR2 dataset is provided in Section V-D.

IV. PROPOSED MODEL

The FDHN model employs a tripartite input structure, encompassing three distinct modalities: a news text modality that incorporates the written content of the news article, a textual context modality that integrates any accompanying textual context information, and a numerical context modality that captures relevant numerical data about the news article. The detailed architecture of the model is illustrated in Fig. 1.

To establish a clear and organized framework for the FDHN model, we compartmentalize it into three main components, each corresponding to one of the input modalities: News Text (Section IV-A), Textual Context (Section IV-B), and Numerical Context (Section IV-C). Additionally, the final Output Layer is detailed in Section IV-E. Recognizing the Fuzzy Layer as a core component of the FDHN model, we present it separately in Section IV-D.

The second-to-last layer of the FDHN model produces four distinct feature representations (News Text, Textual Context, Numerical Context, and Fuzzy Layer output), which are concatenated before passing through the final layer. The ultimate Output Layer determines the plausible class of the

[6]https://www.politifact.com/article/2022/mar/31/politifacts-checklist-thorough-fact-checking/

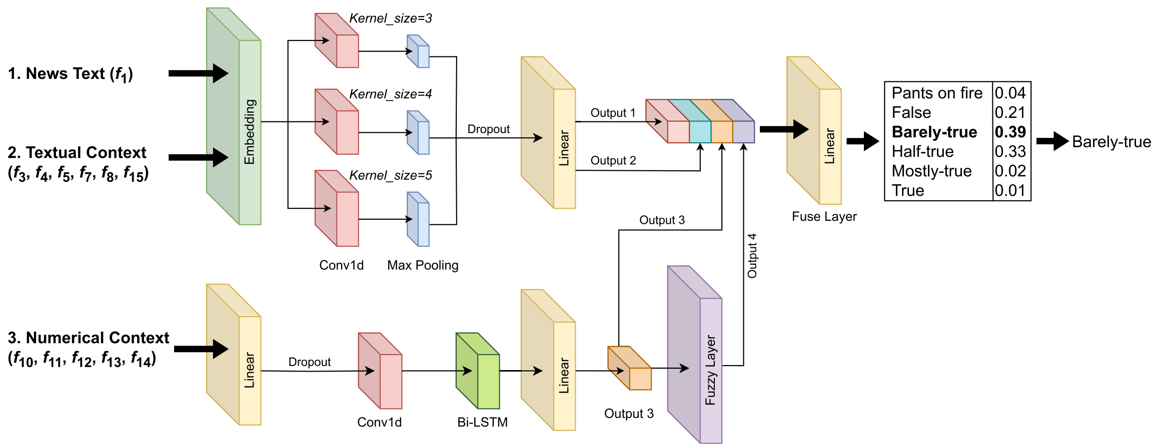

FIGURE 1. **The FDHN model architecture. The input comprises the news text, textual context, and numerical context. The news text and textual context are inputs of two separate TextCNNs, while the numerical context is processed by CNN and Bi-LSTM before going through a Fuzzy Layer. The final result is four outputs; output 1 is derived from the 1. News Text, output 2 is derived from the 2. Textual Context, output 3 is derived from the 3. Numerical Context and Output 4 is the Fuzzy Layer processed Output 3. The four output representations of the four modules are concatenated and fused in the final layer. The figure is illustrated with the LIAR dataset.**

input data. Further insights into the complexity of the FDHN model are discussed in Section IV-F.

A. NEWS TEXT

The text data component of news serves as the primary source of information for the detection of fake news, posing a significant challenge in effective processing. Traditional approaches involve text preprocessing techniques such as stemming, punctuation, and stop word removal, followed by tokenization for downstream tasks. However, contemporary deep learning-based NLP methods tend to preserve the original text information and build end-to-end models, such as BERT [57] and GPT [58]. These models leverage powerful NLP techniques, including convolutional neural networks (CNN) [59], [60], [61], attention mechanisms [62], and recurrent neural networks (RNN) [63], [64], effectively capturing complex semantic relationships within the text. These techniques have demonstrated superior performance in various NLP tasks, including fake news detection, and have become widely adopted in the field.

The News Text module of the FDHN model conducts text classification by extracting salient features to accurately differentiate between true and fake news. To achieve this, we employ the TextCNN model introduced in 2014 [65], which adapts CNN from computer vision and natural language processing [66]. This approach has proven effective in capturing important features and patterns in text data, leading to improved performance in various NLP-based applications, including text classification. It is important to note that the feature representation is based on the processed news text, which is subsequently fed into the downstream components of the model for further processing.

B. TEXTUAL CONTEXT

Textual context information, although still in a textual format, diverges from the main news article text by not directly relating to the article's content, encompassing aspects such as author identity and political affiliation. The direct encoding and integration of such information into the model responsible for processing the main news article text may result in information loss and could potentially interfere with the training of numerical features. To tackle this challenge, we adopt the TextCNN architecture for processing the textual context, deploying a distinct model tailored specifically for handling this contextual information. This approach ensures independent processing and refinement of the contextual information without impacting the main news text model. Subsequent to the TextCNN refinement of the textual contextual information, we obtain a feature representation that seamlessly integrates with other features, facilitating effective classification in downstream components of the proposed model.

C. NUMERICAL CONTEXT

CNN and Bi-LSTM are popularly used in the fake news detection task because of their ability to capture both local and global features of the input data [67], [68], [69], [70], [71]. Convolutional Neural Networks (CNNs) are effective in extracting local features from the input data, while Bidirectional Long Short-Term Memory (Bi-LSTM) networks are good at capturing global features [72]. The combination of these two architectures has been shown to be effective in various natural language processing tasks, including fake news detection [73]. CNNs are used to extract features from the input data, and the resulting feature maps are then fed into a Bi-LSTM network to capture the temporal dependencies between the features. This approach has been shown to be effective in capturing the complex relationships between different parts of the input data, which is crucial for fake news detection.

In the proposed model, we integrate a CNN block followed by a Bi-LSTM [74], collectively referred to as the CNNBiLSTM module, to proficiently process numerical context data. This hybrid model draws inspiration from the approach outlined in the original LIAR paper [14]. Subsequently, the output feature representation undergoes further refinement through a fuzzy layer, as elaborated in Section IV-D. The culmination of this modality yields two distinct representations: the feature representation derived from CNNBiLSTM and the feature representation obtained through the fuzzy layer.

D. FUZZY LAYER

We introduce a novel Fuzzy Logic-based layer designed to transform the output feature representation of numerical context data into a fuzzy membership degree space. This transformation enhances the model's ability to capture underlying patterns while mitigating the impact of noise. The Fuzzy Layer takes a tensor as input, shaped (*batch_size*, *feature_dim*), where *batch_size* denotes the number of instances in a batch, and *feature_dim* is the number of features in each instance. The layer produces an output tensor of shape (*batch_size*, *membership_num*), where *membership_num* represents the number of membership functions synonymous with the number of classes. This innovative layer facilitates a more nuanced and robust representation of numerical context data within the model architecture.

We use the Gaussian membership function for the fuzzy layer, which consists of two parameters: σ and m to control the shape of the function. According to this setting, the fuzzy layer will have *feature_dim* $\times$ *membership_num* fuzzy membership functions, and the equation for each membership function μ_G is

$$\mu_G(x; m, \sigma) = e^{-\frac{1}{2}(\frac{x-m}{\sigma})^2} \tag{1}$$

In this function, m controls the centre of the curve, and σ controls the spread of the curve. To obtain the membership value $\mu_{i,j}$ of instance i for class j, we compute the average of the fuzzy membership values of all features, which is similar to a t-norm like operation. $\mu_{i,j}$ is given by the following equation:

$$\mu_{i,j} = \frac{1}{feature\_dim} \sum_{k=1}^{feature\_dim} \mu_G(x_{i,k}; m_{j,k}, \sigma_{j,k}) \tag{2}$$

where $x_{i,k}$ is the k^{th} feature of the i^{th} instance, $m_{j,k}$ and $\sigma_{j,k}$ are the parameters of the membership function that outputs the membership value of class j to which feature k belongs, respectively.

E. OUTPUT LAYER

The responsibility of the output layer is to fuse the outputs of the individual modalities for the final output, and its structure is a basic linear layer with the equation of $y = xA^T + b$, where x is the input data, y is the output data, A and b are learning parameters. Its input is the output representation of all the modalities concatenated together. Through training, its internal parameters can adaptively assign different weights and biases to minimize the loss.

F. COMPLEXITY ANALYSIS

In this section, we present a comprehensive analysis of the complexity of the FDHN model, encompassing both computational and memory aspects. Our analysis delves into the number of operations executed in each layer and the size of the parameters employed. We denote the size of the dataset by A, the length of the news text by L, the feature dimension by D, the length of all textual context features by T, the number of numerical context features by N, and the number of membership functions (classes) by C.

- **News Text:** In the TextCNN component of the News Text module, convolution operations are the primary contributor to computational complexity. Given that we have M different kernel sizes and each kernel size has K filters, the total computational complexity is $O(A \cdot M \cdot K \cdot L \cdot D)$. The total number of parameters is $O(M \cdot K \cdot L \cdot D)$.
- **Textual Context:** Similarly, the computational complexity in the TextCNN module for the textual context is $O(A \cdot M \cdot K \cdot T \cdot D)$, and the total number of parameters is $O(M \cdot K \cdot T \cdot D)$.
- **Numerical Context:** In the CNNBiLSTM module for the numerical context, the complexity of CNN operations is $O(A \cdot M \cdot K \cdot N \cdot D)$ with a total parameter size of $O(M \cdot K \cdot N \cdot D)$. The complexity of Bi-LSTM operations is $O(A \cdot N \cdot 4 \cdot D^2)$ with a total parameter size of $O(4 \cdot D^2)$.
- **Fuzzy Layer:** The computational complexity of the fuzzy layer comes from the computation of the Gaussian membership function, which is $O(A \cdot N \cdot C)$, and the total number of parameters is $O(N \cdot C)$.
- **Final Output Layer:** Finally, the output layer consists of a linear operation with complexity $O(A \cdot D)$, and its total number of parameters is $O(D)$.

In summary, the overall computational complexity of the FDHN model is $O(A \cdot M \cdot K \cdot (L + T + N) \cdot D + A \cdot N \cdot 4 \cdot D^2 + A \cdot N \cdot C + A \cdot D)$. The total number of parameters, which corresponds to the memory requirement, is $O(M \cdot K \cdot (L + T + N) \cdot D + 4 \cdot D^2 + N \cdot C + D)$. However, since all but A are fixed hyperparameters, the computational complexity is $O(A)$ in the actual computation, which means it is linear.

The analysis presented herein underscores the manageable computational complexity and memory requirements of the FDHN model, rendering it feasible for training on contemporary hardware, even when dealing with extensive datasets. Notably, the model's adaptability is further demonstrated by the ease with which its complexity can be controlled through adjustments in parameters such as the number of filters, kernel sizes in TextCNN, and the number of units in the Bi-LSTM. This adaptability allows the model to cater to specific resource constraints, enhancing its utility across diverse computational environments.

It is crucial to emphasize that the model's complexity is predominantly influenced by the input size and

TABLE 5. Comparison of various fake news detection models on LIAR dataset with or without external data sources. The FDHN (CNN+LSTM+Fuzzy) model achieves the highest accuracy among the compared models. Note: In cases where multiple submodels within the same model exhibit omitted features, it indicates that these features adhere to the identical settings as employed in the first submodel of that specific model.

Author	Base Model	Metadata	External Data	Accuracy
Guo [75]	GNN	Yes	None	0.268
Wang [14]	CNN+LSTM	Yes	None	0.274
Yang [76]	CofCED	No	Justification	0.295
Alhindi [49]	LSTM	Yes	None	0.250
	LR		Justification	0.370
Karimi [77]	MMFD	No	None	0.291
		Yes		0.348
			Verdict Reports	0.388
Liu [78]	BERT	No	Pre-training	0.345
		Yes		0.406
Long [79]	LSTM	Yes	None	0.399
	LSTM+Attention			0.415
Atanasova [55]	DistilBERT	Yes	Justification+Pre-training	0.443
Kirilin [56]	LSTM	Yes	None	0.415
			Speaker2Credit	0.457
FDHN	CNN+LSTM	Yes	None	0.437
	CNN+LSTM+Fuzzy			**0.465**

dimensionality. This characteristic endows the FDHN model with versatility, enabling its application to a broad spectrum of input configurations and diverse practical scenarios. The model's ability to strike a balance between computational efficiency and performance makes it well-suited for a variety of applications, providing a practical and scalable solution for fake news detection and potentially extending its utility to other domains.

V. EXPERIMENTAL RESULTS

In this section, we present the experimental setup and results of the FDHN model to assess its effectiveness in detecting fake news. Initially, in Section V-A, we detail the development environment and hyperparameter settings employed in the experiments. Subsequently, in Section V-B, we conduct a comparative analysis of our FDHN model against other state-of-the-art models on the LIAR dataset. The FDHN model demonstrates superior performance, achieving noteworthy accuracy and F1 scores. Furthermore, in Section V-C, we undertake an ablation study on the components of the FDHN model to discern the contribution of each element to its overall performance. Lastly, we assess our proposed LIAR2 dataset with the FDHN model in Section V-D, which involves comparisons with the LIAR dataset and ablation experiments on the new dataset.

A. EXPERIMENTAL SETTINGS

The implementation of the model is carried out using PyTorch,[7] executed on an NVIDIA Tesla A100 GPU. In this section, we elaborate on the specific configuration of the model employed in the experiments. The output sequence length of each module is set to 6, the dropout rate is configured at 0.5, and the embedding depth is defined as 128, incorporating zero-padding where necessary.

The TextCNN module, responsible for processing both news text and textual context, comprises an embedding layer followed by three parallel CNN layers. These CNN layers employ kernel sizes of (3, 4, 5) with a depth of 128. The output of each CNN layer undergoes a MaxPooling operation to extract the maximum value for each feature map. Subsequently, the resulting feature maps are concatenated and fed into a linear layer with dropout.

The CNNBiLSTM module is dedicated to processing numerical context. It initiates with passage through a linear layer featuring dropout, followed by a CNN layer with an output channel of 32 and a kernel size of 1. The output then traverses a one-layer Bi-LSTM network with dropout, concluding with another linear layer for the final output.

B. PERFORMANCE COMPARISON

After conducting experiments with our proposed fuzzy method, we attained a test set accuracy of 0.465 on the LIAR dataset without external data. To comprehensively assess the performance of our model, we conducted comparisons with representative models, as outlined in Table 5.

Wang [14] introduced the LIAR dataset and proposed a hybrid model based on CNN and LSTM to fuse news data with contextual information for classification. Their model achieved an accuracy of 0.274 on the LIAR dataset, serving as a baseline model for studying this dataset.

Guo et al. [75] presented a model based on heterogeneous graph neural networks, combining data and its corresponding metadata, namely news and context information in the LIAR datasets. Their model achieved comparable results to the baseline model, with an accuracy of 0.268.

Alhindi et al. [49] introduced external justification information into the LIAR dataset, resulting in the creation of the LIAR-PLUS dataset. The model they proposed achieved an accuracy of 0.37 on their dataset. Yang [76] and Atanasova [55] also achieved accuracies of 0.295 and 0.443, respectively, on the LIAR-PLUS dataset.

Karimi et al. [77] and Kirilin [56] introduced verdict reports and Speaker2Credit data into the LIAR dataset, achieving an accuracy of 0.388 and 0.457, respectively.

[7] https://pytorch.org/

TABLE 6. Performance and computational efficiency comparison of different sub-models of the FDHN on validation and test datasets. The models are evaluated based on accuracy, F1-macro, F1-micro, mean, and the average time taken for training one epoch. Three types of models are considered: (i) models with only (TC) or ($TC + FZ$) for news text data, (ii) models with multiple components ($BT + TC + CB$) and their fuzzy counterpart ($BT + TC + CB + FZ$), (iii) models with multiple components ($TC + TC + CB$) and their fuzzy counterpart ($TC + TC + CB + FZ$). The best results on each metric are highlighted in bold.

Model	Validation			Test			Mean	Avg. Epoch Time (Seconds)
	Accuracy	F1-Macro	F1-Micro	Accuracy	F1-Macro	F1-Micro		
TC	0.252	0.177	0.232	0.243	0.193	0.243	0.223	3.096
$TC + FZ$	0.252	0.155	0.248	0.246	0.145	0.246	0.215	3.179
$BT + TC + CB$	0.434	0.421	0.421	0.408	0.415	0.408	0.418	233.941
$BT + TC + CB + FZ$	0.444	0.429	0.438	0.424	0.425	0.424	0.431	234.190
$TC + TC + CB$	0.459	0.442	0.441	0.437	0.444	0.437	0.443	3.779
$TC + TC + CB + FZ$	**0.467**	**0.449**	**0.458**	**0.465**	**0.470**	**0.465**	**0.462**	3.944

Liu et al. [78] and Atanasova also utilized the pre-trained BERT model for this task, with Liu ultimately obtaining an accuracy of 0.406. Similarly, Long et al. [79] experimented with the attention mechanism combined with LSTM, achieving an accuracy of 0.415.

Our FDHN model demonstrated an accuracy of 0.465 on the LIAR dataset, surpassing all the models listed above, including competitors that incorporated external data. To provide a more in-depth understanding of the contributions of individual components of our model, we conducted ablation experiments, the results of which are presented in Section V-C.

C. COMPONENT ANALYSIS

In this section, we present a comprehensive component analysis of our proposed models designed for the task of fake news detection. The performance evaluation of FDHN models is conducted using diverse metrics, encompassing accuracy, F1-macro, and F1-micro, applied to both validation and test sets. Notably, all model training phases are completed within 10 epochs.

Table 6 encapsulates the performance summary of our models on both sets, incorporating four fundamental components: TextCNN (TC), BERT (BT), CNNBiLSTM (CB), and Fuzzy (FZ). The initial two rows depict the performance of the LIAR dataset using solely the TextCNN module for news data processing, both with and without our proposed fuzzy method. In this configuration, only news text data is utilized, devoid of contextual data. The results reveal a marginal improvement in model performance with the inclusion of the fuzzy method on this dataset.

Subsequently, we compare the models utilizing the BERT base pre-trained language model (uncased) for news text data, TextCNN for textual context, and CNNBiLSTM for numeric context, with and without the fuzzy method. The introduction of the fuzzy method demonstrates enhancement in both accuracy and F1 metrics on both the validation and test sets.

Finally, we assess the model performance by substituting BERT with TextCNN for news text data processing, incorporating and excluding the fuzzy method. The outcomes highlight that models incorporating the fuzzy method consistently outperform those without it across all metrics. The best performance is observed in the $TC + TC + CB + FZ$ model, achieving an accuracy of 0.465, F1-macro of 0.470, and F1-micro of 0.465 on the test set.

Our experiment underscores the criticality of incorporating contextual information for accurate fake news detection. The FDHN model exhibits a substantial performance boost, nearly doubling its accuracy on this dataset with the introduction of contextual information. Furthermore, the fuzzy module proves most effective when applied to numerical data, showing less impact on textual data and, in some cases, leading to performance degradation.

Additionally, our experiments reveal that the utilization of BERT did not yield a notable improvement in model performance. This is attributed to the relatively small dataset and concise statements used in our study. Notably, the resource-intensive nature of training with large language models, such as BERT, is highlighted, with the BERT-based model taking approximately 230 seconds per epoch compared to the TextCNN-based model's mere 4 seconds. This resource consumption challenge is also acknowledged in [80].

The integration of fuzzy layers contributes to additional performance improvements. The FDHN model's fuzzy logic simulation effectively addresses the inherent fuzziness and uncertainty associated with the dataset labels, leading to enhanced performance. We posit that fuzzy logic excels when dealing with multi-labeled, fuzzy, and uncertain target task labels. Notably, we refrained from directly replacing the CNNBiLSTM output with the fuzzy layer output, inspired by ResNet [40], as this combined output better preserves the distinctions between different models.

In contrast to existing fake news detection models, such as the hybrid network in the LIAR paper and the competitors listed in Section V-C, the FDHN model demonstrates superior efficiency in fusing data from multiple modalities for classification representations. This enhanced efficiency stems from the tailored data processing approach employed for individual channel data, facilitating a more effective fusion of data across modalities. Furthermore, the incorporation of the Fuzzy module in the FDHN model introduces a more realistic simulation of human decision-making processes, thereby enhancing interpretability. This unique aspect distinguishes the FDHN model from its competitors, as it enables a nuanced understanding of the classification outcomes. The integration of fuzzy layers not only contributes to additional performance

TABLE 7. Performance metrics for experiments on the split LIAR2 dataset. Numbers in parentheses indicate the proportion of the segment that is split. The original in the first line means the unmodified official version of the LIAR dataset, while without the original, the LIAR dataset has been structured by our enhancement. The mix marking in the last line means that the LIAR and NEW parts are all mixed together before splitting, i.e., experiments on the full LIAR2 dataset.

Dataset		Validation			Test			Mean
Train	Test & Val.	Accuracy	F1-Macro	F1-Micro	Accuracy	F1-Macro	F1-Micro	
LIAR (Original .8)	LIAR (Original .2)	0.467	0.449	0.458	0.465	0.470	0.465	0.462
LIAR (.8)	LIAR (.2)	0.614	0.589	0.583	0.608	0.614	0.608	0.603
NEW (.8)	NEW (.2)	0.768	0.671	0.747	0.759	0.695	0.759	0.733
LIAR (1.)	NEW (1.)	0.736	0.696	0.730	0.736	0.706	0.736	0.723
NEW (1.)	LIAR (1.)	0.558	0.520	0.520	0.558	0.563	0.558	0.546
LIAR (1.) + NEW (.557)	NEW (.443)	0.787	0.725	0.770	0.783	0.739	0.783	0.764
LIAR (.635) + NEW (1.)	LIAR (.365)	0.631	0.612	0.607	0.621	0.627	0.621	0.620
LIAR (.8) + NEW (.8)	LIAR (.2) + NEW (.2)	0.698	0.667	0.677	0.689	0.683	0.689	0.684
LIAR + NEW (Mix .8)	LIAR + NEW (Mix .2)	0.697	0.657	0.668	0.702	0.696	0.702	0.687

improvements but also effectively addresses the inherent fuzziness and uncertainty associated with the dataset labels, leading to enhanced overall performance. We posit that fuzzy logic is particularly adept at handling multi-labeled, fuzzy, and uncertain target task labels.

In conclusion, our component analysis underscores the significant performance enhancement achieved through the integration of the fuzzy method, particularly in models such as $TC + TC + CB$ and $TC + TC + CB + FZ$, for the task of fake news detection. Additionally, our findings indicate that, on this dataset, utilizing TextCNN directly outperforms the use of BERT.

D. EXPERIMENT ON LIAR2

In this section, We conduct two types of experiments on the LIAR2 dataset to analyze its structure and quality. The first type is based on segmentation, where we compare the LIAR2 dataset with the LIAR dataset in terms of the information content and distribution. The second type is based on ablation, where we examine the contribution of each feature in the LIAR2 dataset to the fake news detection performance.

Note that, unlike the experiments using FDHN on the LIAR dataset, due to the high number of Token of the justification (f_{16} in Table 2), in order not to interfere with the classification of the statement and textual context, we set up a separate TextCNN module for processing the justification modality, whose structure is the same as that of the TextCNN used for processing the statement and textual context in Section IV-A and IV-B. Therefore, the final number of outputs used for the output layer is changed from 4 to 5.

For the first experiment, we conducted a total of five groups of experiments, and the results are presented in Table 7. To make the comparison easy, we split the LIAR2 dataset into two parts: $LIAR2 = LIAR + NEW$. LIAR refers to the original LIAR dataset after we enhance the structure, and NEW denotes the data that we expand on the original LIAR dataset, i.e., the incremental part. The following insights were obtained from the experimental observations in Table 7:

- **Performance Improvement with Dataset Enhancement:** From the first group of the experiment, we compared the original LIAR dataset with the improved LIAR dataset, and we observed significant improvements in various performance metrics. Notably, the accuracy, F1-Macro, and F1-Micro scores all show substantial increases. In other words, we achieved $\sim 15\%$ improvement in overall performance on the same dataset by only changing the structure of the dataset. This suggests that the efforts to enhance the dataset's structure have resulted in a more robust and informative dataset for training fake news detection models.
- **Effectiveness of the NEW Part:** From the last row of the first group of the experiment, the NEW part, representing an incremental addition to the LIAR dataset, demonstrates impressive performance across all metrics. It consistently outperforms both the original LIAR dataset and the improved LIAR dataset. This underscores the importance of continually updating and expanding datasets to adapt to evolving challenges in fake news detection.
- **The NEW Part has Additional Knowledge:** From the second group of the experiment, it can be seen that the LIAR dataset is more informative compared to the NEW part, which is understandable since the LIAR dataset itself makes up a larger portion of the LIAR2 dataset. However, by analyzing the second and third groups of the experiment together, it is easy to see that the NEW part contains knowledge that the LIAR dataset does not. This means that the LIAR2 dataset was created to allow the model to learn more in order to make the model more robust.
- **Complementary Benefits of Combining Datasets:** In the third, forth and fifth groups of the experiment, We combined the LIAR dataset with the NEW part (LIAR + NEW), and observed a notable boost in performance. This indicates that the synergy between different datasets can lead to enhanced model performance, potentially capturing a broader range of fake news patterns and characteristics.
- **Model Robustness with FDHN:** Throughout the experiments, the FDHN model consistently delivers robust results across all datasets and configurations. This underscores the model's ability to effectively tackle the challenge of fake news detection and adapt to varying dataset compositions. The consistent performance of FDHN implies its potential as a reliable choice for fake news detection tasks.

TABLE 8. Ablation experiment results on the LIAR2 dataset.

Feature	Validation			Test			Mean
	Accuracy	F1-Macro	F1-Micro	Accuracy	F1-Macro	F1-Micro	
Statement	0.317	0.196	0.312	0.320	0.238	0.320	0.284
Date	0.291	0.188	0.291	0.308	0.178	0.308	0.261
Subject	0.324	0.231	0.318	0.327	0.227	0.327	0.292
Speaker	0.328	0.225	0.317	0.331	0.246	0.331	0.297
Speaker Description	0.332	0.244	0.325	0.328	0.244	0.328	0.300
State Info	0.293	0.158	0.295	0.298	0.152	0.298	0.249
Credibility History	0.501	0.470	0.499	0.506	0.466	0.506	0.491
Context	0.298	0.182	0.298	0.313	0.179	0.313	0.264
Justification	**0.596**	**0.566**	**0.583**	**0.612**	**0.597**	**0.612**	**0.594**
All without							
Statement	**0.708**	**0.673**	**0.682**	**0.718**	**0.711**	**0.718**	**0.702**
Date	0.693	0.657	0.668	0.708	0.699	0.708	0.689
Subject	0.700	0.658	0.668	0.708	0.701	0.708	0.691
Speaker	0.694	0.665	0.676	0.704	0.694	0.704	0.690
Speaker Description	0.689	0.664	0.674	0.717	0.707	0.717	0.695
State Info	0.707	0.663	0.673	0.710	0.702	0.710	0.694
Credibility History	0.603	0.572	0.590	0.619	0.605	0.619	0.601
Context	0.701	0.662	0.672	0.704	0.697	0.704	0.690
Justification	0.529	0.490	0.515	0.534	0.515	0.534	0.519
Statement +							
Date	0.343	0.254	0.334	0.338	0.251	0.338	0.310
Subject	0.355	0.276	0.351	0.338	0.258	0.338	0.319
Speaker	0.362	0.286	0.354	0.348	0.264	0.348	0.327
Speaker Description	0.358	0.281	0.353	0.367	0.289	0.367	0.336
State Info	0.332	0.237	0.329	0.333	0.236	0.333	0.300
Credibility History	0.507	0.474	0.508	0.524	0.500	0.524	0.506
Context	0.336	0.268	0.339	0.346	0.256	0.346	0.315
Justification	0.602	0.558	0.580	0.618	0.603	0.618	0.596
All	**0.697**	**0.657**	**0.668**	**0.702**	**0.696**	**0.702**	**0.687**

In the second experiment, we present an in-depth analysis of the ablation experiments for each feature in the LIAR2 dataset. Ablation experiments involve systematically removing specific features or components of the dataset to assess their individual contributions to the model's performance. We conducted a total of three sets of experiments: (1) All features were taken individually for testing. (2) All features were included, and one of those features was removed individually. (3) Statements were retained, and the rest of the features were added individually. The results, as shown in Table 8, offer valuable insights into the significance of different features in fake news detection and the overall effectiveness of the LIAR2 dataset. From this experiment, we obtained the following findings:

- The justification feature is the most important and informative feature for fake news detection, as it has the highest scores when used alone and the lowest scores when removed. This indicates that the justification provides strong evidence for verifying the truthfulness of the statement.
- The credibility history feature is the second most important and informative feature for fake news detection, as it has the second highest scores when used alone and the second lowest scores when removed. This suggests that the credibility history reflects the trustworthiness and consistency of the speaker.
- The statement feature, which is the news itself in fake news detection, is informative, as can be seen by comparing the first and third sets of experiments. However, in the second set of experiments, removing it will make the model perform better and even outperform the model that uses all the features. This suggests that the statement feature has redundant information compared to contextual features, and the noise it introduces on its own makes the model perform less well.
- The other features, such as date, subject, speaker, speaker description, state info, and context, have relatively low scores when used alone or removed. This means that they have less influence on fake news detection, and they may be redundant or noisy.

These findings provide essential guidance for the development of effective fake news detection models and underline the value of the LIAR2 dataset as a versatile resource for research and experimentation in this domain. Intriguingly, our experiments on the LIAR2 dataset unveiled a noteworthy and somewhat counter-intuitive observation. Specifically, we found that excluding the statement feature from our model's input led to an improvement in overall classification accuracy. Further analysis suggests that the statement feature may, in certain instances, introduce redundancy and noise when considered in combination, potentially affecting the model's performance negatively. This outcome underscores the unique characteristics of the LIAR2 dataset, where contextual features, such as subject, speaker, credibility history, and context, provide robust cues to distinguish between fake and genuine news. While this observation challenges the conventional wisdom that the statement itself should be the primary determinant of news veracity, it also highlights the need for a balanced approach to fake news detection. We recognize the importance of the statement feature in fake news detection tasks, as it encapsulates the core content. However, relying solely on this attribute might not capture the nuances and subtleties inherent in news articles. We aspire to harness the statement's significance in our fake news detection task while extending its impact by integrating it effectively with contextual information. This approach acknowledges that the statement, as the news itself, lacks robustness and interpretability when utilized in isolation for classifying news articles solely based on contextual information.

VI. CONCLUSION AND LIMITATION

This paper introduces FDHN, an innovative deep-learning model incorporating fuzzy logic principles to address the complex problem of fake news detection. Our approach distinguishes itself from conventional methods by integrating a classification algorithm rooted in fuzzy set theory, inherently equipped to handle uncertainties and imprecisions inherent in natural language contexts. Experimental results on the LIAR dataset showcase the superior performance of our approach compared to existing state-of-the-art models, validating the robustness and efficacy of FDHN.

The study emphasizes the pivotal role of fuzzy logic in enhancing the performance of multi-label tasks, particularly those involving fuzzy labels. Additionally, experiments highlight that the application of fuzzy logic in our model does not significantly escalate computational demands, demonstrating efficiency in resource utilization. This finding is significant, affirming the practical feasibility of incorporating fuzzy logic into intricate tasks.

While the integration of fuzzy logic contributes to a moderate improvement in model accuracy, its true value lies not only in potential performance enhancements but also in the novel avenues it opens for exploration. Fuzzy logic provides unique interpretability compared to traditional probabilistic logic, a characteristic of particular significance in fields where interpretability holds critical importance, such as social sciences and medicine. This attribute positions FDHN as a valuable tool not only for achieving superior performance but also for advancing the understanding and interpretability of complex tasks like fake news detection.

Apart from the benefits that fuzzy logic can bring, our research has unveiled a notable limitation in the application of fuzzy logic within our model, particularly when extending it to text-based modalities. While the advantages of fuzzy logic are evident when applied to numerical modalities, we have observed that its effect on text-based modalities is not only insignificant but, in certain instances, leads to a degradation of performance. This intriguing observation underscores the intricate interplay between data characteristics and the efficacy of fuzzy logic. The inherently complex, unstructured, and noisy nature of text-based data may already encompass inherent uncertainty, rendering the additional fuzzification less advantageous. Consequently, our findings emphasize the importance of meticulous feature engineering and the strategic placement of fuzzy layers to ensure their maximum effectiveness. This limitation serves as a compelling invitation for further research and experimentation, aiming to refine the utilization of fuzzy logic within complex deep learning models and, in particular, addressing the nuances of text-based data to improve fake news detection methodologies.

In addition to proposing the LIAR2 dataset, which is an improved and extended version of the original LIAR dataset, we have made several contributions to the field of fake news detection. We identified that the original modalities used to collect the LIAR dataset were not enough and proposed efficient modalities that complemented the original ones. As a consequence, the accuracy of the dataset improved significantly. We also collected more data and mainly augmented the initial data so that it becomes representative enough to learn from it. Furthermore, we proposed an efficient learning technique based on Deep Learning to analyze the new dataset and build an efficient classifier for fake news detection. The proposed system achieved around 70% accuracy, which is far better than the existing systems.

While the substantial improvement of LIAR2 on the LIAR dataset makes it a promising and potentially useful asset in the field of fake news detection, there are still some limitations. Since the LIAR/LIAR2 dataset focuses exclusively on political statements, the ability of the FDHN model to detect fake news in other domains remains untested despite the significant proportion of fake news attributed to political content. The dataset used in this study contains a relatively small number of political statements compared to the vast number of statements made daily, presenting a limitation due to its reliance on extensive contextual information and pre-labeled rationales to determine factual accuracy. While our experiments, detailed in Table 8, demonstrate that the modular construction of the FDHN model allows for the omission of certain features without rendering the model unusable, it does lead to a degradation in performance. Additionally, despite the rigorous efforts of the PolitiFact team to label each statement scientifically, the inherent human subjectivity in the labeling process cannot be completely eliminated. This subjectivity poses a challenge not only to our study but also to the broader field of fake news detection research.

Our comprehensive analysis and experimentation demonstrated the LIAR2 dataset's versatility and the intricate interplay between various features in the context of news veracity classification. While the unexpected observation regarding the statement feature may challenge conventional assumptions, it highlights the dataset's uniqueness and the importance of considering the multifaceted nature of news itself. Our findings underscore the need for a balanced approach that leverages both statement and contextual information to enhance the robustness and interpretability of fake news detection models. The LIAR2 dataset not only serves as a benchmark for assessing model performance but also opens doors to innovative research avenues, encouraging further exploration into the nuanced challenges of distinguishing fact from fiction in the realm of news media. As the landscape of misinformation continues to evolve, the LIAR2 dataset remains a crucial asset for researchers seeking to advance the frontiers of fake news detection.

VII. FUTURE WORK

Guided by the findings and insights from our current research, we envisage the following prospective avenues for future work:

- **Exploring Alternative Fuzzy Logic-Based Algorithms:** We will investigate alternative fuzzy logic algorithms like fuzzy clustering and fuzzy neural networks to enhance our understanding and application of fuzzy logic in fake news detection.
- **Application to Diverse Datasets and Real-World Scenarios:** To validate the effectiveness and universality of our model, we plan to apply it to a diverse range of datasets and situate it within real-world scenarios. This will provide valuable insights into the model's adaptability and performance across different contexts.
- **Incorporation of Additional Features:** We aim to enhance our model by integrating a broader set of features such as sentiment analysis, linguistic cues, and visual indicators, ultimately evolving it into a comprehensive multimodal fake news detection system.
- **Emphasising Interpretability and Explainability:** Further research efforts will focus on delving deeper into the interpretability and explainability aspects of our approach. This emphasis is crucial for building trust in the model and ensuring that it produces results that are comprehensible and trustworthy.
- **Performance Stability with Varied Scales of Input Data:** A key aspect of our future endeavors is to investigate how changes in the volume, velocity, and variety of data impact the performance of our model. This exploration is essential for ensuring the scalability and robustness of the model in diverse, real-world settings.
- **Exploring the Application of News and Tweets in Financial Economics:** Recent trends in financial economics involve the use of news articles and tweets to develop measures such as uncertainty, risk, polarity, and sentiment. These techniques offer significant potential for advancing our understanding of financial markets. Our approach may serve as a foundational model for future research, with implications for addressing and mitigating the impact of fake news in the financial sector.

Our current research represents a significant advancement in the ongoing pursuit of effectively detecting fake news through the application of machine learning and deep learning techniques. These outlined avenues for future work aim to further refine and expand the capabilities of our proposed model, contributing to the broader field of fake news detection.

In the landscape of NLP, the integration of large language models (LLMs), exemplified by models like GPT-3 (Generative Pre-trained Transformer) [58], [81], has marked a paradigm shift, showcasing remarkable capabilities. However, it is imperative to acknowledge that fact-checking remains a notable challenge for generalized LLMs [82], [83], [84], introducing issues such as model illusion, forgetting non-parametric memories, and neglecting long-tail knowledge [85], [86], [87]. This study strategically focuses on addressing these challenges within LLMs, recognizing that despite the overarching prowess of such models in various tasks, the specific nuances of fact-checking are pivotal for a comprehensive understanding. In the context of contemporary LLMs, notably the commercially-driven GPT-4 [88], which is not open-sourced, dedicated efforts to enhance fact-checking mechanisms become increasingly pertinent. While the deployment of LLMs may obfuscate the importance of task-specific problem-solving, we contend that tackling intricacies such as fact-checking not only aids in immediate applications but also contributes significantly to our broader pursuit of artificial general intelligence in the future. This alignment with specific challenges enriches our understanding of language model intricacies, thereby fostering advancements in the quest for more intelligent and responsible AI systems.

Our pioneering approach, rooted in fuzzy logic principles, not only contributes to the existing body of research but also catalyzes further exploration and innovation within this pivotal domain. We envision our work as a catalyst for continued advancements, reinforcing the critical importance of tackling misinformation in our evolving information landscape.

REFERENCES

[1] D. Pogue, "How to stamp out fake news," *Sci. Amer.*, vol. 316, no. 2, p. 24, Jan. 2017.

[2] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *J. Econ. Perspect.*, vol. 31, no. 2, pp. 211–236, May 2017.

[3] R. Zafarani, X. Zhou, K. Shu, and H. Liu, "Fake news research: Theories, detection strategies, and open problems," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, Jul. 2019, pp. 3207–3208.

[4] Y. M. Rocha, G. A. de Moura, G. A. Desidério, C. H. de Oliveira, F. D. Lourenço, and L. D. de Figueiredo Nicolete, "The impact of fake news on social media and its influence on health during the COVID-19 pandemic: A systematic review," *J. Public Health*, vol. 31, no. 7, pp. 1007–1016, Jul. 2023.

[5] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, Mar. 2018.

[6] C. Silverman, *This Analysis Shows How Viral Fake Election News Stories Outperformed Real News on Facebook*. New York, NY, USA: BuzzFeed News, 2016.

[7] C. Xu and N. Yan, "AROT-COV23: A dataset of 500k original Arabic tweets on COVID-19," in *Proc. 4th Workshop Afr. Natural Lang. Process.*, 2023, pp. 1–9.

[8] C. Colomina, H. S. Margalef, R. Youngs, and K. Jones, *The Impact of Disinformation on Democratic Processes and Human Rights in the World*. Brussels, Belgium: European Parliament, 2021.

[9] 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.

[10] X. Zhang and A. A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion," *Inf. Process. Manage.*, vol. 57, no. 2, Mar. 2020, Art. no. 102025.

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

[12] J. Shang, J. Shen, T. Sun, X. Liu, A. Gruenheid, F. Korn, A. D. Lelkes, C. Yu, and J. Han, "Investigating rumor news using agreement-aware search," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2018, pp. 2117–2125.

[13] R. Zellers, A. Holtzman, H. Rashkin, Y. Bisk, A. Farhadi, F. Roesner, and Y. Choi, "Defending against neural fake news," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 9054–9065.

[14] W. Wang, "'Liar, liar pants on fire': A new benchmark dataset for fake news detection," in *Proc. 55th Annu. Meeting ACL (Short Papers)*, vol. 2. Vancouver, BC, Canada, Jul. 2017, pp. 422–426.

[15] N. Vo and K. Lee, "Where are the facts? Searching for fact-checked information to alleviate the spread of fake news," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2020, pp. 7717–7731.

[16] P. Patwa, S. Sharma, S. Pykl, V. Guptha, G. Kumari, M. Akhtar, A. Ekbal, A. Das, and T. Chakraborty, "Fighting an infodemic: COVID-19 fake news dataset," in *Proc. Int. Workshop Combating Line Hostile Posts Regional Lang. During Emergency Situation*. Cham, Switzerland: Springer, 2021, pp. 21–29.

[17] L. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, 1965.

[18] L. Zadeh, *Fuzzy Logic*. New York, NY, USA: Springer, 2023, pp. 19–49.

[19] J.-S. R. Jang, "ANFIS: Adaptive-network-based fuzzy inference system," *IEEE Trans. Syst. Man, Cybern.*, vol. 23, no. 3, pp. 665–685, Jun. 1993.

[20] Y. Deng, Z. Ren, Y. Kong, F. Bao, and Q. Dai, "A hierarchical fused fuzzy deep neural network for data classification," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 4, pp. 1006–1012, Aug. 2017.

[21] R. Das, S. Sen, and U. Maulik, "A survey on fuzzy deep neural networks," *ACM Comput. Surv.*, vol. 53, no. 3, pp. 1–25, May 2020.

[22] F. Olan, U. Jayawickrama, E. O. Arakpogun, J. Suklan, and S. Liu, "Fake news on social media: The impact on society," *Inf. Syst. Frontiers*, vol. 26, pp. 443–458, Jan. 2022.

[23] C. Xu and M.-T. Kechadi, "Fuzzy deep hybrid network for fake news detection," in *Proc. 12th Int. Symp. Inf. Commun. Technol.*, New York, NY, USA, Dec. 2023, pp. 118–125.

[24] A. Grignolio, M. Morelli, and M. Tamietto, "Why is fake news so fascinating to the brain?" *Eur. J. Neurosci.*, vol. 56, no. 11, pp. 5967–5971, Dec. 2022.

[25] T. Yerlikaya and S. Toker, "Social media and fake news in the post-truth era: The manipulation of politics in the election process," *Insight Turkey*, vol. 22, pp. 177–196, Jun. 2020.

[26] J. Alghamdi, Y. Lin, and S. Luo, "A comparative study of machine learning and deep learning techniques for fake news detection," *Information*, vol. 13, no. 12, p. 576, Dec. 2022.

[27] Y. Yang, J. G. Carbonell, R. D. Brown, T. Pierce, B. T. Archibald, and X. Liu, "Learning approaches for detecting and tracking news events," *IEEE Intell. Syst.*, vol. 14, no. 4, pp. 32–43, Jul. 1999.

[28] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on Twitter," in *Proc. 20th Int. Conf. World wide Web*, New York, NY, USA, Mar. 2011, pp. 675–684.

[29] Q. Sheng, J. Cao, X. Zhang, R. Li, D. Wang, and Y. Zhu, "Zoom out and observe: News environment perception for fake news detection," in *Proc. 60th Annu. Meeting Assoc. Comput. Linguistics (Long Papers)*, vol. 1, 2022, pp. 4543–4556.

[30] Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo, "Multimodal fusion with recurrent neural networks for rumor detection on microblogs," in *Proc. 25th ACM Int. Conf. Multimedia*, New York, NY, USA, Oct. 2017, pp. 795–816.

[31] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, L. Su, and J. Gao, "EANN: Event adversarial neural networks for multi-modal fake news detection," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, Jul. 2018, pp. 849–857.

[32] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proc. 28th Int. Conf. Mach. Learn. (ICML)*, 2011, pp. 689–696.

[33] V. Radu, C. Tong, S. Bhattacharya, N. D. Lane, C. Mascolo, M. K. Marina, and F. Kawsar, "Multimodal deep learning for activity and context recognition," *ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 4, pp. 1–27, Jan. 2018.

[34] T. Kim, B. Kang, M. Rho, S. Sezer, and E. G. Im, "A multimodal deep learning method for Android malware detection using various features," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 3, pp. 773–788, Mar. 2019.

[35] C. L. P. Chen, C.-Y. Zhang, L. Chen, and M. Gan, "Fuzzy restricted Boltzmann machine for the enhancement of deep learning," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 6, pp. 2163–2173, Dec. 2015.

[36] Y.-J. Zheng, W.-G. Sheng, X.-M. Sun, and S.-Y. Chen, "Airline passenger profiling based on fuzzy deep machine learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 12, pp. 2911–2923, Dec. 2017.

[37] Y. Wang, Z. Wu, and J. Zhang, "Damaged fingerprint classification by deep learning with fuzzy feature points," in *Proc. 9th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2016, pp. 280–285.

[38] H. A. Chopade and M. Narvekar, "Hybrid auto text summarization using deep neural network and fuzzy logic system," in *Proc. Int. Conf. Inventive Comput. Informat. (ICICI)*, Nov. 2017, pp. 52–56.

[39] S. Zhou, Q. Chen, and X. Wang, "Fuzzy deep belief networks for semi-supervised sentiment classification," *Neurocomputing*, vol. 131, pp. 312–322, May 2014.

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

[41] B. Mohamed, H. Haytam, and F. Abdelhadi, *Applying Fuzzy Logic and Neural Network in Sentiment Analysis for Fake News Detection: Case of COVID-19*. Cham, Switzerland: Springer, 2022, pp. 387–400.

[42] Y. Yang, S. Nazir, and W. Khalil, "A probabilistic approach toward evaluation of Internet rumor on COVID," *Soft Comput.*, vol. 26, no. 16, pp. 8077–8088, Aug. 2022.

[43] T. M. H. Gedara, V. Loia, and S. Tomasiello, "Using fuzzy transform for sustainable fake news detection," *Appl. Soft Comput.*, vol. 151, Jan. 2024, Art. no. 111173.

[44] B. D. Ravichandran and P. Keikhosrokiani, "Classification of COVID-19 misinformation on social media based on neuro-fuzzy and neural network: A systematic review," *Neural Comput. Appl.*, vol. 35, no. 1, pp. 699–717, Jan. 2023.

[45] A. D'Ulizia, M. C. Caschera, F. Ferri, and P. Grifoni, "Fake news detection: A survey of evaluation datasets," *PeerJ Comput. Sci.*, vol. 7, p. e518, Jun. 2021.

[46] J. Thorne, A. Vlachos, C. Christodoulopoulos, and A. Mittal, "FEVER: A large-scale dataset for fact extraction and VERification," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol. (Long Papers)*, vol. 1, 2018, pp. 809–819.

[47] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "FakeNewsNet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media," *Big Data*, vol. 8, no. 3, pp. 171–188, Jun. 2020.

[48] S. Hakak, M. Alazab, S. Khan, T. R. Gadekallu, P. K. R. Maddikunta, and W. Z. Khan, "An ensemble machine learning approach through effective feature extraction to classify fake news," *Future Gener. Comput. Syst.*, vol. 117, pp. 47–58, Apr. 2021.

[49] T. Alhindi, S. Petridis, and S. Muresan, "Where is your evidence: Improving fact-checking by justification modeling," in *Proc. 1st Workshop Fact Extraction VERification (FEVER)*, 2018, pp. 85–90.

[50] M. H. Goldani, S. Momtazi, and R. Safabakhsh, "Detecting fake news with capsule neural networks," *Appl. Soft Comput.*, vol. 101, Mar. 2021, Art. no. 106991.

[51] S. Yang, K. Shu, S. Wang, R. Gu, F. Wu, and H. Liu, "Unsupervised fake news detection on social media: A generative approach," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 1, 2019, pp. 5644–5651.

[52] N. Aslam, I. U. Khan, F. S. Alotaibi, L. A. Aldaej, and A. K. Aldubaikil, "Fake detect: A deep learning ensemble model for fake news detection," *Complexity*, vol. 2021, pp. 1–8, Apr. 2021.

[53] W. Ostrowski, A. Arora, P. Atanasova, and I. Augenstein, "Multi-hop fact checking of political claims," in *Proc. 13th Int. Joint Conf. Artif. Intell.*, Aug. 2021, pp. 3892–3898.

[54] D. Stammbach and E. Ash, "e-FEVER: Explanations and summaries for automated fact checking," in *Proc. Truth Trust Online (TTO)*, 2020, pp. 32–43.

[55] P. Atanasova, J. G. Simonsen, C. Lioma, and I. Augenstein, "Generating fact checking explanations," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 7352–7364.

[56] A. Kirilin and M. Strube, "Exploiting a speaker's credibility to detect fake news," in *Proc. Data Sci., Journalism Media Workshop KDD (DSJM)*, 2018, pp. 1–7.

[57] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, vol. 1. Minneapolis, MI, USA: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186.

[58] A. Radford and K. Narasimhan, "Improving language understanding by generative pre-training," OpenAI, San Francisco, CA, USA, Tech. Rep., 2018.

[59] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. J. Lang, "Phoneme recognition using time-delay neural networks," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 3, pp. 328–339, Mar. 1989.

[60] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, Dec. 1989.

[61] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[62] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 5998–6008.

[63] J. Elman, "Finding structure in time," *Cognit. Sci.*, vol. 14, no. 2, pp. 179–211, Jun. 1990.

[64] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[65] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, Doha, Qatar, 2014, pp. 1746–1751.

[66] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing [review article]," *IEEE Comput. Intell. Mag.*, vol. 13, no. 3, pp. 55–75, Aug. 2018.

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

[68] P. Bahad, P. Saxena, and R. Kamal, "Fake news detection using bi-directional LSTM-recurrent neural network," *Proc. Comput. Sci.*, vol. 165, pp. 74–82, Jan. 2019.

[69] A. Wani, I. Joshi, S. Khandve, V. Wagh, and R. Joshi, "Evaluating deep learning approaches for COVID19 fake news detection," in *Combating Online Hostile Posts in Regional Languages During Emergency Situation*, T. Chakraborty, K. Shu, H. R. Bernard, H. Liu, and M. S. Akhtar, Eds. Cham, Switzerland: Springer, 2021, pp. 153–163.

[70] A. Roy, K. Basak, A. Ekbal, and P. Bhattacharyya, "A deep ensemble framework for fake news detection and multi-class classification of short political statements," in *Proc. 16th Int. Conf. Natural Lang. Process.*, 2019, pp. 9–17.

[71] A. Abedalla, A. Al-Sadi, and M. Abdullah, "A closer look at fake news detection: A deep learning perspective," in *Proc. 3rd Int. Conf. Adv. Artif. Intell.*, New York, NY, USA, Oct. 2019, pp. 24–28.

[72] J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, T. Darrell, and K. Saenko, "Long-term recurrent convolutional networks for visual recognition and description," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2625–2634.

[73] J. P. C. Chiu and E. Nichols, "Named entity recognition with bidirectional LSTM-CNNs," *Trans. Assoc. Comput. Linguistics*, vol. 4, pp. 357–370, Dec. 2016.

[74] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997.

[75] J. Guo, L. Du, W. Bi, Q. Fu, X. Ma, X. Chen, S. Han, D. Zhang, and Y. Zhang, "Homophily-oriented heterogeneous graph rewiring," in *Proc. ACM Web Conf.*, New York, NY, USA, Apr. 2023, pp. 511–522.

[76] Z. Yang, J. Ma, H. Chen, H. Lin, Z. Luo, and Y. Chang, "A coarse-to-fine cascaded evidence-distillation neural network for explainable fake news detection," in *Proc. 29th Int. Conf. Comput. Linguistics*, Oct. 2022, pp. 2608–2621.

[77] H. Karimi, P. Roy, S. Saba-Sadiya, and J. Tang, "Multi-source multi-class fake news detection," in *Proc. 27th Int. Conf. Comput. Linguistics*, 2018, pp. 1546–1557.

[78] C. Liu, X. Wu, M. Yu, G. Li, J. Jiang, W. Huang, and X. Lu, "A two-stage model based on BERT for short fake news detection," in *Knowledge Science, Engineering and Management*, C. Douligeris, D. Karagiannis, and D. Apostolou, Eds. Cham, Switzerland: Springer, 2019, pp. 172–183.

[79] Y. Long, Q. Lu, R. Xiang, M. Li, and C.-R. Huang, "Fake news detection through multi-perspective speaker profiles," in *Proc. Int. Joint Conf. Natural Lang. Process.*, Nov. 2017, pp. 252–256.

[80] L. Bates and I. Gurevych, "Like a good nearest neighbor: Practical content moderation and text classification," 2023, *arXiv:2302.08957*.

[81] L. Floridi and M. Chiriatti, "GPT-3: Its nature, scope, limits, and consequences," *Minds Mach.*, vol. 30, no. 4, pp. 681–694, Dec. 2020.

[82] I.-C. Chern, S. Chern, S. Chen, W. Yuan, K. Feng, C. Zhou, J. He, G. Neubig, and P. Liu, "FacTool: Factuality detection in generative AI—A tool augmented framework for multi-task and multi-domain scenarios," 2023, *arXiv:2307.13528*.

[83] V. Rawte, A. Sheth, and A. Das, "A survey of hallucination in large foundation models," 2023, *arXiv:2309.05922*.

[84] C. Xu, S. Guan, D. Greene, and M.-T. Kechadi, "Benchmark data contamination of large language models: A survey," 2024, *arXiv:2406.04244*.

[85] A. Mallen, A. Asai, V. Zhong, R. Das, D. Khashabi, and H. Hajishirzi, "When not to trust language models: Investigating effectiveness of parametric and non-parametric memories," in *Proc. 61st Annu. Meeting Assoc. Comput. Linguistics (Long Papers)*, vol. 1, 2023, pp. 9802–9822.

[86] M. Sap, E. Horvitz, Y. Choi, N. A. Smith, and J. Pennebaker, "Recollection versus imagination: Exploring human memory and cognition via neural language models," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 1970–1978.

[87] Q. Wang, L. Ding, Y. Cao, Z. Tian, S. Wang, D. Tao, and L. Guo, "Recursively summarizing enables long-term dialogue memory in large language models," 2023, *arXiv:2308.15022*.

[88] *GPT-4 Technical Report*, OpenAI, San Francisco, CA, USA, 2023.

CHENG XU received the B.Eng. degree in data science and big data technology from Huaibei Normal University, Huaibei, China, in 2022. He is currently pursuing the Ph.D. degree in computer science with University College Dublin, Dublin, Ireland. From 2018 to 2021, he was a Research Assistant with the Data Mining Laboratory, Huaibei Normal University. His current research interests include the development of a multimodal fuzzy system for fake news detection, generative models, and reinforcement learning.

M-TAHAR KECHADI (Member, IEEE) received the M.Sc. and Ph.D. degrees in computer science from the University of Lille 1, France. He is currently a Full Professor in data science with the School of Computer Science, University College Dublin. He is also a Principal Investigator with the Insight Centre for Data Analytics. His research interests include big data analytics and its applications to real-world applications (digital healthcare and digital agriculture), distributed mining techniques and models and their execution environments and applications, cloud computing and services for supporting data access, management, mining processes, digital forensics, cybercrime investigations, and cybersecurity. He has been a Visiting Professor at many universities (Liverpool, Fuzhou, and Dalian). He is a member of ACM.