IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 10, 2021, accepted November 16, 2021, date of publication November 18, 2021,
date of current version November 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3129329

A Comprehensive Review on Fake News Detection With Deep Learning

M. F. MRIDHA[1], (Senior Member, IEEE), ASHFIA JANNAT KEYA[1], MD. ABDUL HAMID[2], MUHAMMAD MOSTAFA MONOWAR[2], AND MD. SAIFUR RAHMAN[1]

[1]Department of Computer Science and Engineering, Bangladesh University of Business and Technology, Dhaka 1216, Bangladesh
[2]Department of Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia

Corresponding author: M. F. Mridha (firoz@bubt.edu.bd)

ABSTRACT A protuberant issue of the present time is that, organizations from different domains are struggling to obtain effective solutions for detecting online-based fake news. It is quite thought-provoking to distinguish fake information on the internet as it is often written to deceive users. Compared with many machine learning techniques, deep learning-based techniques are capable of detecting fake news more accurately. Previous review papers were based on data mining and machine learning techniques, scarcely exploring the deep learning techniques for fake news detection. However, emerging deep learning-based approaches such as Attention, Generative Adversarial Networks, and Bidirectional Encoder Representations for Transformers are absent from previous surveys. This study attempts to investigate advanced and state-of-the-art fake news detection mechanisms pensively. We begin with highlighting the fake news consequences. Then, we proceed with the discussion on the dataset used in previous research and their NLP techniques. A comprehensive overview of deep learning-based techniques has been bestowed to organize representative methods into various categories. The prominent evaluation metrics in fake news detection are also discussed. Nevertheless, we suggest further recommendations to improve fake news detection mechanisms in future research directions.

INDEX TERMS Natural language processing, machine learning, deep learning, fake news.

I. INTRODUCTION

The Internet has changed interaction and communication ways through low cost, simple access, and fast information dissemination. Therefore, social media and online portals have become more popular for news searches and reading for many people rather than traditional newspapers. Social media harms society by influencing major events even though it has become a powerful means of information. Especially after the presidential election of the U.S. in 2016, the issue of online false news has gained more popularity [1], [2]. According to Zhang and Ghorbani [3], voters might be easily controlled by deceptive political statements and claims. Inspection shows that false news or lies propagate more quickly through humans than original information and cause tremendous effects [4].

The terms rumor and fake news are closely interrelated. Fake news or disinformation is intentionally created. On the other hand, rumors are unconfirmed and questionable information that is spread without the aim to deceive [15]. On social media sites, spreaders' intentions might be difficult to determine. As a result, any false or incorrect information is typically branded as misinformation on the Internet. Distinguishing real and fake information is challenging. However, many approaches have been adopted to address this issue. Various machine learning (ML) methods have been used to detect false information spread online in the case of knowledge verification [16], natural language processing (NLP) [16]–[18] and sentiment analysis [19]. Early research concentrated on leveraging textual information derived from the article's content, such as statistical text features [20] and emotional information [21]–[23].

Deep learning (DL) has recently become an emerging technology among the research community and has proven to be more effective in recognizing fake news than traditional ML methods. DL has some particular advantages over ML, such as a) automated feature extraction, b) lightly dependent on data pre-processing, c) ability to

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

TABLE 1. A comparison of existing surveys based on fake news detection.

Survey							Deep Lerning Approaches					
ref	*Year*	*Taxonomy*	*Datasets*	*NLP Techniques*	*Evaluation Metrics*	*Challenges*	*CNN*	*RNN*	*GNN*	*GAN*	*Attention*	*BERT*
[5]	2017	✗	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗
[6]	2018	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗
[7]	2018	✗	✓	✓	✗	✓	✗	✗	✗	✗	✗	✗
[8]	2019	✗	✗	✗	✗	✗	✓	✓	✗	✓	✗	✗
[9]	2019	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗
[10]	2019	✓	✓	✗	✗	✗	✓	✓	✗	✗	✗	✗
[3]	2019	✓	✓	✓	✗	✓	✗	✗	✗	✗	✗	✗
[11]	2019	✓	✓	✓	✗	✗	✓	✓	✗	✓	✗	✗
[12]	2019	✗	✓	✓	✗	✗	✓	✓	✗	✓	✗	✗
[13]	2020	✓	✗	✓	✗	✗	✓	✓	✗	✓	✗	✗
[14]	2021	✓	✗	✓	✗	✓	✓	✓	✓	✓	✗	✗
Ours	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

extract high-dimensional features, and c) better accuracy. Further, the current wide availability of data and programming frameworks has boosted the usage and robustness of DL-based approaches. Hence, in the last five years, numerous articles have been published on fake news detection, mostly based on DL strategies [24]. An enthusiastic effort has been made to review the current literature to compare the extensive amount of DL-based fake news detection research efforts.

A number of research works has been published on the survey of fake news detection [5], [25], [26]. Our investigation reveals that existing studies do not provide a thorough overview of deep learning-based architectures for detecting fake news. The existing survey papers mostly cover the ML strategies in detecting fake news, scarcely exploring the DL strategies [3], [9], [10]. We provide a complete list of NLP techniques as well as describe their benefits and drawbacks. In what follows, in this survey, we performed an in depth analysis of current DL-based studies. Table 1 provides a brief overview of the existing survey papers and our research contributions. The present study aims to address the previous research's weaknesses and strengths by conducting a systematic survey on fake news detection. First, we divide existing fake news detection research into two main categories: (1) Natural Language Processing (NLP) and (2) Deep Learning (DL). We discuss the NLP techniques such as data pre-processing, data vectorizing, and feature extraction. Second, we analyze the fake news detection architectures based on different DL architectures. Finally, we discuss used evaluation metrics in fake news detection. Figure 1 depicts an overall taxonomy of fake news detection approaches. We also include a table 2, including acronyms used throughout the survey to assist researchers when encountering issues due to acronyms.

The rest of the paper is organized as follows. Section II highlights the consequences of fake news. Section III describes the used datasets. Section IV explains the Natural Language Processing techniques in fake news detection. Section V contains an in-depth analysis of deep learning strategies. Section VI presents the evaluation metrics used in previous studies. Section VII narrates the challenges and future research direction. Finally, Section VIII concludes the paper.

II. FAKE NEWS CONSEQUENCES

There has always been fake news since the beginning of human civilization. However, the spread of fake news is increased by modern technologies and the conversion of the global media landscape. The major consequences on social, political, and economic environments may be caused by fake news. Fake information and fake news have various faces. As information molds our view toward the world, fake news has a huge impact. We make critical decisions based on the information. By obtaining information, we develop an impression about a situation or people. We cannot obtain good decisions if we find fake, false, distorted, or fabricated information on the Internet. The primary impacts of fake news are as follows:

Impact on Innocent People: Rumors can have a major impact on specific people. These people may be harassed by social media. They may also face insults and threats that may have real-life consequences. People must not believe in invalid information on social media or judge a person.

Impact on Health: The number of people searching for health-related news on the Internet is continuously increasing. Fake news in health has a potential impact on people's lives [36]. Therefore, this is one of the major challenges today. Misinformation about health has had a tremendous impact in the last year [37]. Social media platforms have made some policy changes to ban or limit the spread of health misinformation as they face pressure from doctors, lawmakers, and health advocates.

Financial Impact: Fake news is currently a crucial problem in industries and the business world. Dishonest businessmen spread fake news or reviews to raise their profits. Fake information can cause stock prices to fall. It can ruin the fame of a business. Fake news also has an impact on customer expectations. Fake news can create an unethical business mentality.

Democratic impact: The media has discussed the fake news phenomenon significantly because fake news played a

TABLE 2. The table contains the acronyms used in this survey.

Acronym	Meaning	Acronym	Meaning
ML	Machine Learning	dEFEND	Explainable Fake News Detection
DL	Deep Learning	GCN	Graph Convolutional Network
NLP	Natural Language Processing	RvNN	Recursive Neural Networks
BoW	Bag of Words	PGNN	Propagation Graph Neural Network
TF-IDF	Term Frequency Inverse Document Frequency	SAGNN	Simplified Aggregation Graph Neural Network
SVM	Support Vector Machine	GANs	Generative Adversarial Networks
NB	Naive Bayes	SeqGAN	Sequence GAN
KNN	K-Nearest Neighbour	RL	Reinforcement Learning
GloVe	Global Vectors for Word Representation	EANN	Event Adversarial Neural Network
GI	Gini Coefficient	GCAN	Graph-aware Co-Attention Networks
IG	Information Gain	3HAN	Three-level Hierarchical Attention Network
1v1I	Mutual Information	att-RNN	attention on RNN
PCA	Principal Component Analysis	ACT	Automatic fake news Classification Through self-attention
CHI	Chi-Square Statistics	BERT	Bidirectional Encoder Representations for Transformers
TI-CNN	Text and Image information based convolutional neural network	BDANN	BERT-based Domain-Adaption Neural Network
DNNs	Deep Neural Networks	MLM	Mask Language Model
RF	Random Forest	NSP	Next Sentence Prediction
CNN	Convolutional Neural Network	exBAKE	BERT with extra unlabeled news corpora
RNN	Recurrent Neural Network	1d-CNN	One-dimensional Convolutional Neural Network
MLP	Multilayer Perceptron	A	Accuracy
MCNN	Multilevel CNN	P	Precision
TFW	Sensitive Word's Weight Calculating Method	R	Recall
LSTM	Long Short Term Memory Networks	F1	F1-score
GRU	Gated Recurrent Unit	ROC	Receiver Operating Characteristics
Bi-LSTM	Bidirectional LSTM	FPR	False Positive Rate
Bi-GRU	Bidirectional GRU	AUC	Area Under the ROC curve
CSI	Capture, Score, and Integrate	DBN	Deep Belief Network
FDML	Fake News Detection Multi Task Learning	GPT	Generative Pre-trained Transformer
AI	Artificial Intelligence	XAI	Explainable Artificial Intelligence

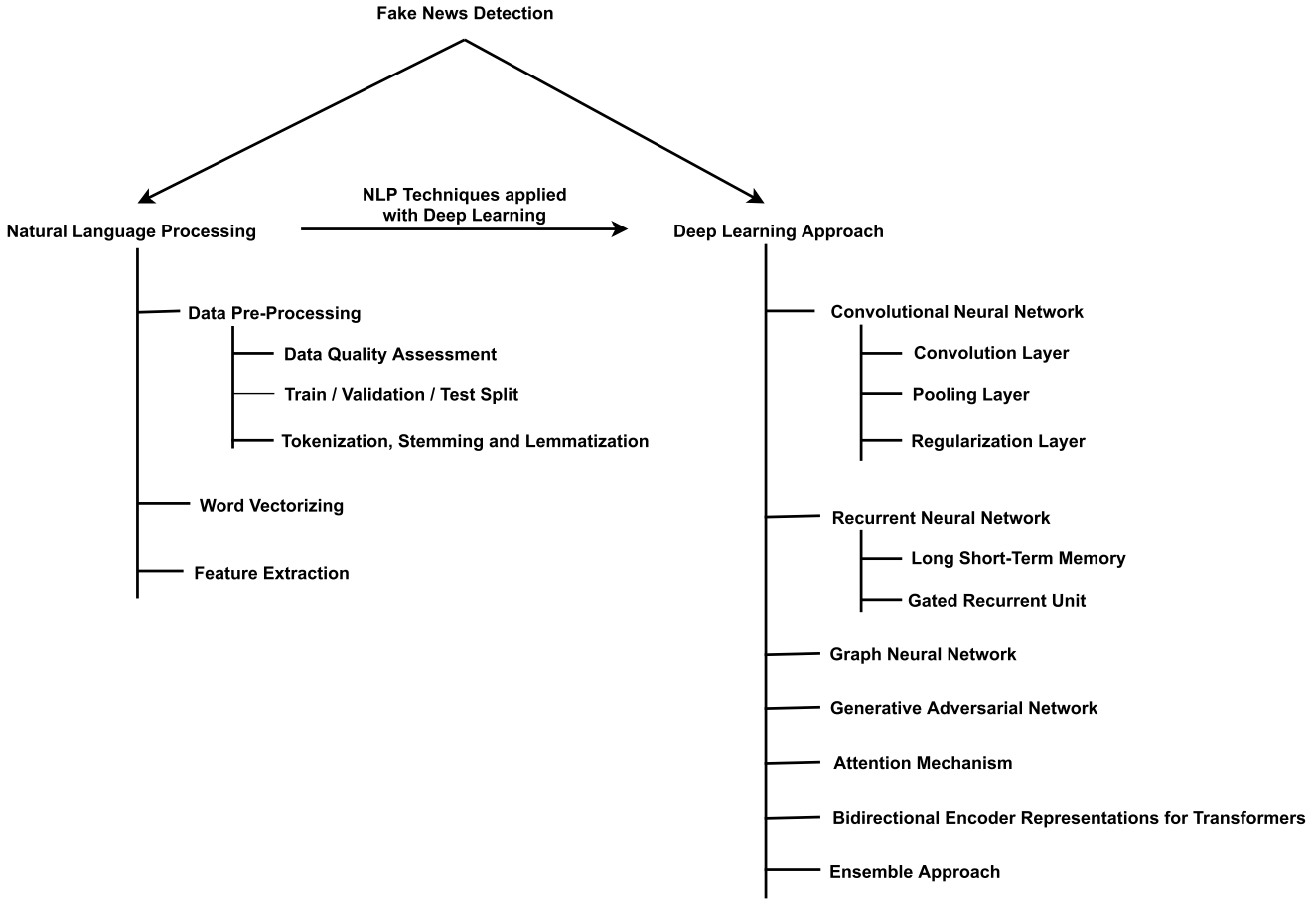


FIGURE 1. A taxonomy of deep learning-based fake news detection.

vital role in the last American presidential election. This is a major democratic problem. We must stop spreading fake news as it has a real impact.

III. BENCHMARK DATASET

In this section, we discuss the datasets used in various studies. For both training and testing, benchmark datasets

TABLE 3. The table provides details of publicly available datasets and corresponding URLs.

Dataset	Modality	Size	Labels	Type	URL
Fake news	Text	20,800	Unreliable, reliable	News articles	https://www.kaggle.com/c/fake-news/data.
Weibo [27]	Text & image	40k tweets	Rumor, Non-rumor	Social media data	https://drive.google.com/file/d/14VQ7EWPiFeGzxp3XC2DeEHi-BEisDINn/view
Twitter15 [28]	Propagation trees	1,381 propagation trees, 276,663 users	Unverified, true, false, non-rumor	Social media data	https://www.dropbox.com/s/7ewzdrbelpmrnxu/rumdetect2017.zip?dl=0
Twitter16 [28]	Propagation trees	1,181 propagation trees, 173,487 users	Unverified, true, false, non-rumor	Social media data	https://www.dropbox.com/s/7ewzdrbelpmrnxu/rumdetect2017.zip?dl=0
LIAR [29]	Text	12.8K	Pants on fire, false, barely true, half-true, mostly true, and true	Political statements	https://paperswithcode.com/dataset/liar
PHEME [30]	Text	5800 tweets	Rumor, Non-rumor	Social media data	https://figshare.com/articles/dataset/PHEME_dataset_of_rumours_and_non-rumours/4010619
FNC-1	Text	75K	Agrees, disagrees, discusses, unrelated	News articles	https://github.com/FakeNewsChallenge/fnc-1
FakeNewsNet [31]	Text	5K	Fake, real	News articles, social media data	https://github.com/KaiDMML/FakeNewsNet
News Aggregator	Text	422,937	Real	News articles	https://www.kaggle.com/uciml/news-aggregator-dataset
Bend the truth [32]	Text	900	Fake, real	News articles	https://github.com/MaazAmjad/Datasets-for-Urdu-news.git
FacebookHoax [33]	Text	15,500	Hoax, non-hoax	scientific news	https://github.com/gabll/some-like-it-hoax/tree/master/dataset
Twitter [34]	Text and Image	992	Rumor, non-rumor	Fact-checked claims	https://github.com/MKLab-ITI/image-verification-corpus/tree/master/mediaeval2015
KaggleFN	Text	13K	Fake	News articles	https://www.kaggle.com/mrisdal/fake-news
FakevsSatire [35]	Text	486	Fake, satire	Political news	https://github.com/jgolbeck/fakenews

were utilized. One of the difficulties in identifying fake news is the shortage of a labeled benchmark dataset with trustworthy ground truth labels and a massive dataset. Based on that, researchers can obtain practical features and construct models [38]. For several usages in DL and ML, such datasets have been collected over the last few years. The datasets are vastly diverse from one another because of different study agendas. For instance, a few datasets are made up entirely of political statements (such as PolitiFact), while others are made up entirely of news articles (FNC-1) or social media posts (Twitter). Datasets can differ based on their modality, labels, and size. Therefore we categorize these datasets in table 3 based on these characteristics. Fake articles are frequently collected from fraudulent websites designed intentionally to disseminate disinformation. These false news stories are eventually shared on social media platforms by their creators. Malicious individuals or bots and inattentive users who do not care to check the source of the story before sharing it assist in spreading fake news through social media. However, most datasets contain only news content. But current language features and writing style are not sufficient enough in developing an efficient detection model.

Fake news, Twitter15, and Liar are the most popular datasets that are publicly available. But some studies trained their model with their created dataset [39]. We defined these datasets as self-collected. Since sufficient information is not provided about their self-collected datasets, we find it difficult to compare with other studies properly. Using the benchmark dataset, a comparative study can be established with current state-of-the-art methods for detecting fake news. Kaliyar *et al.* [40] conducted a comparative study of their suggested model with existing methods using the Kaggle dataset and they reported an accuracy of 93.50% which is the highest, utilizing the same dataset for fake news detection. A pie chart of used benchmark datasets is given in 2.

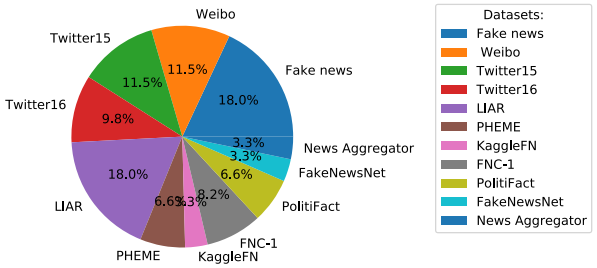


FIGURE 2. A pie chart of the benchmark datasets used in the studies of fake news detection.

IV. NATURAL LANGUAGE PROCESSING

Natural Language Processing (NLP) is an area in machine learning with the capability of a computer to understand, analyze, manipulate, and potentially generate human language. The NLP technique consists of data pre-processing and word embedding. By utilizing deep learning techniques, NLP has seen some colossal advancements in recent years [41]. The natural language must be transformed into a mathematical structure to give machines a sense of natural language. In section IV-A, IV-B, and IV-C, NLP techniques are discussed.

A. DATA PRE-PROCESSING

Data pre-processing is utilized to represent complex structures with attributes, binarize attributes, change discrete attributes, persist, and manage lost and obscure attributes.

During data pre-processing, different visualization procedures are helpful. A cautious pre-processing strategy is required to ingest the data in a neural network for fake news detection because social media data sources are fragmented, unstructured, and noisy. It is a popular fact that amid the learning stage, data pre-processing saves computational time and space. In addition, limiting the impact of artifacts during the learning process, text pre-processing avoids every ingests of noisy data. The data becomes a logical representation after proper text pre-processing. It also included the most representative descriptive words. Umer *et al.* [42] experimented on a fake news detection model in which the accuracy was only 78% when they used the features excluding data cleaning or pre-processing, which is surprisingly poor. After performing the pre-processing steps and removing unnecessary data, the accuracy increases dramatically to 93.0%. Data quality assessment, dimensionality reduction, and splitting of the dataset are the data pre-processing steps used in various studies [39], [41], [43]. The pre-processing steps are elaborated in Sections IV-A1, IV-A2, and IV-A3.

1) DATA QUALITY ASSESSMENT

Data are frequently taken from numerous sources that are ordinarily reliable and are in completely different formats. When working on a machine learning problem, more time is invested in managing data quality issues. It is unreasonable to anticipate that the data would be perfect. There may be some issues due to a human blunder, defects within the data collection process, or restrictions on measuring gadgets. The quality of a dataset is often responsible for the poor performance of fake news detection models. For this reason, the quality of the data used in any machine learning project will have a huge effect on the chances of success. However, only a few studies ensure the quality of their used datasets. S and Chitturi [41] collected the George McIntire dataset from GitHub and dropped the rows that did not have labels in the clarifying process, and the process surely has a huge impact on their success in fake news detection. To ensure the quality of the entire dataset, Wang *et al.* [44] removed duplicate and low-quality images. Alsaeedi and Al-Sarem [45] extended the data cleaning process by URL removal, lowercase and hashtag character (#) removal, mention character (@), and number removal. They also considered words with recurring characters such as ''Likkke'' and handled emoticons by supplanting positive emoticons with a ''positive'' word and with a ''negative'' word for negative emoticons.

2) TRAIN/VALIDATION/TEST SPLIT BASED

The dataset may be divided into train, test, and validation sets. The sample of data that is utilized to adjust the parameters is called the training set. The validation set is a series of examples used to fine-tune the parameters of a model. A set of examples applied only for assessing a fully-specified model's performance is regarded as the test set. Although many studies on fake news detection have divided their dataset into training, validation, and test sets, few studies have used only the training, and test sets [46], [47]. The ratios of data split 60:20:20, 70:30, and 80:20 are very common in fake news detection. The Pareto principle (for many outcomes, roughly 80% of consequences come from 20% of the causes) is used to describe the 80:20 ratio. It is typically a safe bet to use the ratio that all studies applied. Mandical *et al.* [48] applied the ratio of 90:5:5 and 80:10:10 when the number of articles in the dataset was less than 10,000 and greater than 10,000, respectively. However, they did not specify the purpose behind it. Jadhav and Thepade [49] compared their model performance based on the data splitting ratio and showed that 75%–25% data split has more prominent performance than other models possessing diverse splits. The model parameter estimates exhibit more prominent variation with smaller training data. Performance statistics exhibit more prominent variation with smaller testing data. Studies should be careful with splitting data so that neither variation is too large or too small, and it has more to do with the total number of instances in each category rather than the percentage. The optimal split of the test, validation and train sets is determined by hyperparameters, model architecture, data dimension, etc. Table 4 provides an overview of the advantages and disadvantages of the splitting ratios used in most studies:

TABLE 4. The table gives an overview of common dataset partitioning based on training, validation, and testing with advantages and disadvantages. Few studies mentioned their data partitioning, and only those references are given in the table.

Train, Validation, and Test	Advantages	Disadvantages	Reference
80, 0, 20	A limited set of validation data or no validation data is useful when the model has few or no hyperparameters.	Lack of validation data can cause the model to overfit.	[50], [41], [47], [51], [52], [53], [17]
70, 0, 30	An adequate number of the test set is required to assess a model.	With smaller validation data, the assessment measures like accuracy, precision, recall, and F1 score will have a high variation and not lead to adequate model tuning.	[54], [55], [56], [57], [58], [59], [60]
60, 20, 20	With vast hyperparameters to tune, the model requires a more considerable validation set to optimize the model performance.	With a smaller training set , the model would not have enough data to learn.	[4], [61], [62], [63]

3) TOKENIZATION, STEMMING AND LEMMATIZATION

Tokenization is a method of breaking down a text into words. This can be applied to any character. Performing tokenization on a space character is the most common way of tokenization.

Chopping off an end to achieve the base word is called stemming. The removal of derivational affixes is usually included in the stemming. A derivational affix is an affix in which one word is obtained from another. The derived word is usually a distinct class of words from the original.

Lemmatization is a text normalization procedure that morphologically analyzes words, generates the root form of inflated words, and is normally intended to remove inflectional endings [64]. A group of letters applied to the end of a word to modify its meaning is known as an inflectional

ending. Some examples of inflectional endings are s, bat, and bats.

Rusli *et al.* [52] performed two experiments to detect fake news with and without stemming and stop-word removal. They used stemming and stop-word removal for removing all affixes and stop-words. They achieved a 0.82 macro-averaged F1-score by performing the stemming and stop-word removal processes. They also achieved a 0.8 macro-averaged F1-score without performing stemming and stop-word removal. Performing the stemming and stop-word removal processes in the text preprocessing phase was time-consuming, but there was a small difference in the results. Although tokenization, stemming, and lemmatization improve the performance of the classifier, many researchers have not used these techniques [4], [65]. Jain and Kasbe [66] presented simple technique with web scrapping for detecting fake news. They showed that updating the dataset regularly with web scrapping a model's truthfulness can be checked. The authors achieved an accuracy of 91% based on text. The result can be improved greatly with some extra preprocessing, such as stemming and omitting stop words.

B. WORD VECTORIZING

Word vectorizing involves mapping the word/text to a list of vectors. TF-IDF and Bag of Words (BoW) vectorization techniques are commonly used in machine learning strategies to identify fake news [4], [53], [63]. In term frequency inverse document frequency (TF-IDF), the value rises proportionally to the number of times a word emerges in the document but is balanced by the frequency of the word in the body. Although this vectorization is successful, the semantic sense of the words is lost in its attempt to translate to numbers [48]. The BoW technique considers every news article to be a document and computes the frequency count of each word within this document, which is then used to produce a numeric representation of the data. In addition to data loss, this approach also has limitations. The relative location of the words is overlooked, and contextual information is lost. This loss can be costly at times when measured against the benefit in computing convenience with the ease of use [46]. Rusli *et al.* [52] used TF-IDF and Bag of Words feature extraction methods to detect fake news. However, this approach may suffer due to loss of information.

Neural network-based models have accomplished victory on diverse language-related roles as opposed to traditional machine learning-based models such as logistic regression of support vector machine (SVM) by utilizing word embeddings in fake news detection. It maps words or text to a list of vectors. They are low-dimensional, and disseminated feature representations are appropriate for natural languages. The term ''word embedding'' refers to a combination of language modeling and feature learning. Words or expressions from the lexicon are allocated to real-number vectors. Neural network models essentially utilize this method for fake news detection [42], [96]. Word representation was performed using dense vectors in word embedding. These vectors represent the word mapping onto a continuous, high-dimensional vector space. This is considered an improvement over the BoW model, wherein large sparse vectors of vocabulary size were used as word vectors. These large vectors also provided no information about how the two words were interrelated or any other useful information [50]. Recently, fake news detection researchers have used pre-trained word-embedding models such as global vectors for word representation (GloVe) and Word2vec. The primary benefit of using these models is their ability to train with large datasets [40]. Unlike Word2Vec, GloVe supports parallel implementation, making it easier to train the model on huge datasets. Table 5 gives a summary of the NLP techniques and word vector models used in deep learning-based fake news detection papers.

C. FEATURE EXTRACTION

A huge amount of computational power and memory is required to analyze a large number of variables. Classification algorithms may overfit the training samples and induce poorly to new samples. Feature extraction is a process of building combinations of variables to overcome these difficulties while still representing the data with adequate precision. Feature extraction and feature selection are frequently used in text mining [69], [97].

Fake news detection strategies concentrate on applying news content and social context features [98]. News content features highlights depict the meta-information relevant to a chunk of news [5]. Commonly, in news validation, news content (linguistics and visual information) is used as a feature [99], [100]. Textual features comprise the writing style and emotion [101], [102]. Furthermore, hidden textual representations are generated using tensor factorization [103]–[105] and deep neural networks [106]–[108], achieving high performance in detecting false news with news contents. Visual features are retrieved from visual components such as image and video, but only a few studies utilized visual features in fake news detection [109], [110]. In contrast, social context information can also be aggregated for detecting fake news in social media. There are three main perspectives of social content: a) users, b) produced posts and c) networks (connection amidst the users who distributed relevant posts) [5]. User-based features are typically from the user profile in social media [98], [111]. Users' social responses in terms of stances [42], [64], topics [112], or credibility [113]–[115] are represented via post-based features. Recently, several studies have focused on stance features to detect fake news [64]. It can be effective for human fact-checkers to distinguish false claims [113], [114]. To check the authenticity of a claim/report/headline, it is essential to understand what different news agencies are declaring about that particular claim/report/headline. Reference [116]. Features that are network-based are retrieved by creating specialized networks, such as diffusion networks, interaction networks, and propagation networks [117]–[119]. The propagation network contains rich information about user interactions (likes, comments, responses, or shares) that

TABLE 5. The table provides the advantages and disadvantages of Word Vector Models, along with the references.

Method	Advantages	Disadvantages	References
TF-IDF	The TF-IDF model includes information on both the more significant and less important words.	Slow for large vocabularies. Does not capture position in text, semantics, co-occurrences in different documents, etc.	[67, 4, 68, 49, 46, 52, 63, 69, 17, 70, 71, 72]
Bag-of-words	The ease of implementation.	It ignores the ordering of the words in a given document. Ignores the semantic relations among words	[68, 46, 73, 74, 70, 75]
Word2Vec [76]	Maintains the semantic meaning of various words in a text. The context information is preserved. The size of the embedding vector is very small	Inability to deal with unfamiliar words. There are no common representations at the sub-word level.	[40, 42, 41, 74, 77, 46, 78, 79, 73, 60, 80, 55, 51, 81]
Doc2Vec [82]	A numeric representation of a document, regardless of its length. Faster than Word2vec.	The benefit of using doc2vec is diminished for shorter documents	[83, 84]
GloVe [85]	GloVe, unlike Word2vec, does not rely solely on local statistics (Words local context information)	In order to obtain word vectors, global statistics (word co-occurrence) are used.	[40, 50, 41, 86, 60, 51, 87, 80, 88]
BERT [89]	Identify and capture contextual meaning in a sentence or text	Compute-intensive at inference time	[79, 90, 75, 91, 92, 93, 81, 94, 95]

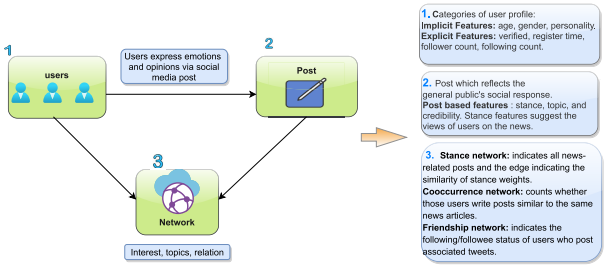


FIGURE 3. The infographic illustrates the social content/context features such as user, post and network elaborately.

show the direction of information flow, timestamp details about interactions, textual information about user interactions, and user profile information about the users who are interacting [120]. We provide Figure 3 depicting important features that were utilized to detect fake news precisely.

It is pivotal to choose the correct determination algorithm for decreasing features because feature reduction contains an incredible effect on the text classification results. Some common feature reduction algorithms include Gini Coefficient (GI), Term Frequency-Inverse Document Frequency (TF-IDF), Information Gain (IG), Mutual Information (1v1I), Principal Component Analysis (PCA), and Chi-Square Statistics (CHI). In the process of content classification, the linear classification model works well with the TF-IDF model [121]. PCA and Chi-square were utilized to improve the adaptability of the text classifier combined with deep learning models. A number of studies compared their model accuracy with and without feature extraction and found that with feature extraction, the success rate is higher. Umer *et al.* [42] compared the applications of feature reduction methods (PCA and Chi-square) applied with two deep learning models. When the proposed model is utilized with the reduced feature set, it increases the F1-score and accuracy by 20% and 4%, respectively, compared to the other techniques. However, many studies did not perform feature extraction, although it has a significant impact on the result [16], [122]. Neural networks are considered very powerful machine learning tools due to their ability of complex feature extraction. Instead of relying on manual feature selection and other existing techniques, researchers are currently focusing on neural networks for feature extraction [123]. Yang *et al.* [124] employed a model TI-CNN (Text and Image information based convolutional neural network) to extract latent features from both visual and textual information and achieved promising results [124]. Another study [107] used the deep recurrent neural network model for extraction of a collection of latent features for news producers, posts, and topics.

V. DEEP LEARNING APPROACH FOR FAKE NEWS DETECTION

Deep learning models have seen exceptional growth in recent times owing to their promising success in several fields, including communication and networking [125], [126], computer vision [127], [128], intelligent transportation [129], speech recognition [130], as well as NLP. Deep learning systems have advantages over traditional machine learning methods. Deep learning is a subfield of machine learning strategies, which displays high precision and exactness in fake news detection. Generally, ML methods are based on hand-crafted features. Biased features may appear because feature extraction assignments are challenging and slow. ML approaches failed to achieve prominent results in fake news detection. Because ML approaches produce high-dimensional representations of linguistic information, resulting in the curse of dimensionality. The existing neural network-based models have outperformed the traditional models in terms of their performance owing to their exceptional feature extraction ability [62]. In contrast, DL systems can acquire hidden representations from less complex inputs. The hidden features can be extracted from both the news content and context varieties. A study by Hiramath and Deshpande [78] showed that deep neural networks (DNNs) require less time than other ML-based classification algorithms such as logistic regression, random forest (RF), and SVM, etc. However, DNNs use more memory. Convolutional neural network (CNN) and recurrent neural network (RNN) are two broadly utilized ideal models for deep learning in

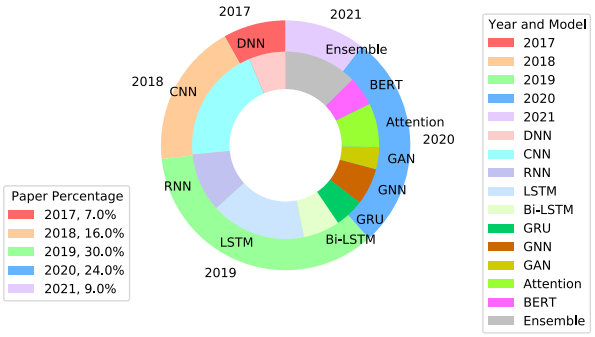


FIGURE 4. A nested pie chart illustrating the percentage of published articles and popular models each year.

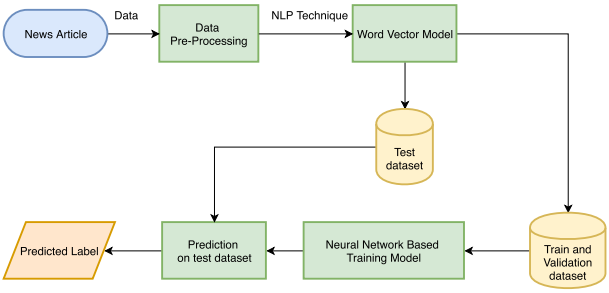


FIGURE 5. The diagram illustrates the general deep learning-based architecture that was used in most studies.

cutting-edge artificial neural networks. Therefore, we provide Figure 4, which shows the percentage of DL-based fake news detection papers with used classifiers in recent years.

After inspecting previous studies, we found a general framework for deep learning-based fake news detection. The first step was to collect a dataset or create one. Most studies have used news articles collected from publicly available datasets. The pre-processing technique was applied after collecting the dataset to feed the data in a neural network [42], [96], [131]. Word2vec and GloVe word embedding methods have mostly been used in previous studies to map words into vectors [41], [78], [80]. We represent an overall process for fake news identification with deep learning in Figure 5 based on various studies [40], [42], [61].

148 DL-based studies were examined to provide a detailed description of these architectures: CNN in section V-A and RNN in Section V-B, Graph Neural Network in Section V-C, Generative Adversarial Network in Section V-D, Attention Mechanism in Section V-E, Bidirectional Encoder Representations for Transformers in Section V-F, and Ensemble Approach in Section V-G.

A. CONVOLUTIONAL NEURAL NETWORK (CNN)

A few deep learning models have been introduced to handle ambiguous detection issues. CNNs and RNNs are the most interesting models [77]. Researchers are trying to boost the performance of the fake news detector with CNN by taking its power of extracting features well and better classification process [132]. However, CNNs are also gaining popularity in the NLP technique too. It is utilized for mapping the features of n-gram patterns. The CNN is similar to a multilayer perceptron (MLP) as it is an unsupervised multilayer feed-forward neural network [45]. The CNN consists of an input layer, an output layer, and a sequence of hidden layers. CNNs are mostly used for picture recognition and classification. Neural networks with 100 or more hidden layers have been reported in recent studies. Backward-propagation and forward-propagation algorithms are utilized in neural networks. These algorithms are used to train neural networks by updating the weights of each layer. The gradient (derivative) of the cost function is utilized to update the weights. When the sigmoid activation function is applied, the value of the gradient decreases per layer. This lengthens the training time. This problem is called the vanishing-gradient problem. A deeper CNN or a direct connection in dense solves this problem. Compared to a normal CNN, a deeper CNN is also less vulnerable to overfitting [67]. Kaliyar *et al.* [40] proposed a model FNDNet (deep CNN), which is designed to learn the discriminatory features for fake news detection using multiple hidden layers. The model is less prone to overfitting but takes a longer time to train. The convolutional layer, pooling layer, and regularization layer are the most utilized layers in CNNs for fake news detection. The input data can be manipulated through pooling and convolution operations. Sections V-A1, V-A2, and V-A3 describe the popular layers used in CNN.

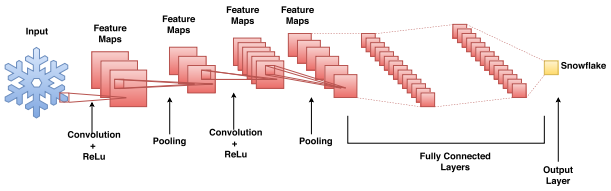


FIGURE 6. The figure shows the architecture of CNN. Here, an input picture of a snowflake is given to the CNN picture classifier. The input goes through a series of convolution layers, pooling layer, fully connected layers, and classifies the object based on learned features.

1) CONVOLUTION LAYER

CNNs work very well with image classification and computer vision because of the convolution operation, and their ability to extract features from inputs for better representation makes them very efficient. These properties make CNNs powerful in sequence processing [131]. Fernández-Reyes and Shinde [77] proposed a CNN architecture called, StackedCNN (2-dimensional convolution layers, rather than 1-dimensional convolutions). It is proven that finding patterns in text data a fusion of pre-trained word embeddings with 2-dimensional convolutional layers helps, but the performance of the StackedCNN is poor compared to state-of-the-art CNN. Another study by Li *et al.* [132] adopted a novel approach with multilevel CNN (MCNN) and Sensitive word's weight calculating method (TFW). MCNN-TFW successfully captured semantic information from the article text content. For this reason, it outperforms the compared methods, including

CNN. Their work did not consider latent-based features. Alsaeedi and Al-Sarem [45] added more convolution layers, and it has an impact on the proposed model performance. According to the results, the model's performance is lowered by about 0.014.

2) POOLING LAYER

A pooling operation that chooses the greatest component from each patch of each feature map covered by the filter is called max pooling. A pooling layer is a new layer attached to the convolutional layer. Its purpose is to continuously diminish the spatial size of the representation in order to decrease the number of parameters and the calculation inside the network. The pooling layer operates autonomously on each feature map. Max pooling or average pooling is the most commonly used function in fake news detection. Alsaeedi and Al-Sarem [45] adjusted the hyperparameter settings in a CNN. They found the best parameter settings that gave an improvement in the model's performance. The recommended CNN model performs best when the number of units in the dense layer is set to 100, the number of filters is set to 100, and the window size is set to 5. The GlobalMaxPooling1D method achieved the highest scores, showing that it works well for fake news detection when compared to other pooling methods [45].

3) REGULARIZATION LAYER

The most crucial problem of classification is to reduce the training and test errors of the classifier. Another common issue is the over-fitting problem (the space between training and testing errors is huge). Overfitting makes it difficult to generalize the model as it becomes more applicable (overfit) to the training set. Regularization is a solution to the overfitting problem. Regularization is applied to the model to lessen the problem of overfitting and decrease the error of generalization, but not the error of training [45]. The dropout regularization method is mostly used for fake news detection [133]. Other methods such as early stopping and weight penalties were not used in previous studies on fake news detection. Dropout avoids overfitting by gradually filtering out neurons. Eventually, all weights are calculated as an average so that the weight is not too high for a single neuron.

B. RECURRENT NEURAL NETWORK (RNN)

The RNN is a type of neural network. In RNN, nodes are sequentially connected to construct a directed graph. The output from the earlier step serves as the input to the current step. RNNs are effective in time and sequence-based predictions. RNN is less compatible with features compared to CNN. RNNs are suitable for studying sequential texts and expressions. However, it cannot process very long sequences when tanh or ReLU is used as an activation function.

The backward-propagation algorithm is utilized in the RNN for training. While training the neural networks, it is required to take tiny steps frequently in the way of the negative error derivative concerning network weights to establish

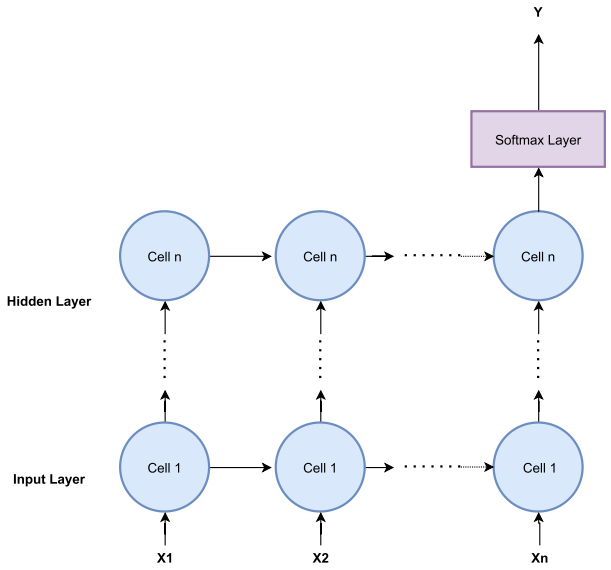


FIGURE 7. The figure shows an architecture of basic RNN with n sequential layers. x represents the inputs and y represents the output generated by the RNN.

a minimum error function. The size of the gradients becomes tiny for each consequent layer. Thus, the RNN suffers from a vanishing gradient issue in the bottom layers of the network. We can deal with the vanishing gradient problem by using three solutions: (1) using rectified linear unit (ReLU) activation function, (2) using RMSProp optimization algorithm, and (3) using diverse network architecture such as long short-term memory networks (LSTM) or gated recurrent unit (GRU). So previous studies focused on LSTM and GRU rather than the state-of-the-art RNN [80], [96], [134]. Bugueño *et al.* [80] proposed a model based on RNN for propagation tree classification. The authors used RNN for sequence analysis. The number of epochs was set as 200, which is relatively high in comparison to their training examples. To predict fake news articles, authors have proposed distinctive RNN models, specifically LSTM, GRU, tanh-RNN, unidirectional LSTM-RNN, and vanilla RNN. RNNs, and in specific LSTM, are especially successful in processing sequential data (human language) and catching significant features out of diverse data sources. Further, in Sections V-B1 and V-B2, we discuss LSTM and GRU.

1) LONG SHORT-TERM MEMORY (LSTM)

LSTM models are front runners in NLP problems. LSTM is an artificial recurrent neural network framework used in deep learning. LSTM is a progressed variation of RNN [41]. RNNs are not capable of learning long-term dependencies because back-propagation in recurrent networks takes a while, particularly for the evolving backflow of blunder. However, LSTM can keep ''Short Term Memories'' for ''Long periods.'' The LSTM is made up of three gates: an input gate, an output gate, a forget gate, and a cell. Through a combination of the three, it calculates the hidden state. The cell can recall values over a large time interval. The word's connection within the

beginning of the content can impact the output of the word afterward within the sentence for this reason [67]. LSTM is an exceptionally viable solution for tending the vanishing gradient issue. Bahad *et al.* [61] proposed an RNN model that suffers from the vanishing gradient issue. To tackle this issue, they implemented an LSTM-RNN. But still, LSTM could not solve the vanishing gradient issue completely. The LSTM-RNN model had a higher precision compared to the initial state-of-the-art CNN. Asghar *et al.* [135] proposed bidirectional LSTM (Bi-LSTM) with CNN for rumor detection. The model preserves the sequence information in both directions. The Bi-LSTM layer is effective in remembering long-term dependency. Even though the BiLSTM-CNN beat the other models, the suggested approach is computationally expensive.

A study by Ruchansky *et al.* [123] suggested a model called CSI, which comprises three modules, Capture, Score, and Integrate. The capture module extracts features from the article, and the score module extracts features from the user. Then by integrating article and user-based features, the CSI model performs the prediction for fake news detection. The CSI model has fewer parameters than other RNN-based models. Another study by Sahoo and Gupta [136] proposed an approach with both user profile and news content features for detecting false news on Facebook. The authors used LSTM to identify fake news, and a set of new features are extracted by Facebook crawling and Facebook API. It requires more time to train and test the suggested model. Liao *et al.* [137] proposed a novel model called fake news detection multi-task learning (FDML). The model explores the influence of topic labels for fake news while also using contextual news information to improve detection performance on short false news. The FDML model, in particular, is made up of representation learning and multi-task learning components that train both the false news detection task and the news topic categorization task at the same time. However, the performance of the model decreases without the author's information.

2) GATED RECURRENT UNIT (GRU)

In terms of structure and capabilities, GRU is comparatively easier and more proficient than LSTM. This is because there are only two gates, to be specific, reset and update. The GRU manages the information flow in the same manner as the LSTM unit does, but without the use of a memory unit. It literally exposes the entire hidden content with no control whatsoever. When it comes to learning long-term dependencies, the quality of GRU is way better than LSTM. Hence, it is a promising candidate for NLP applications [41]. GRUs are more straightforward as well as much more proficient compared to LSTM. GRU is still in its early stages, thus, we are seeing it being used lately to identify false news. GRU is a newer algorithm with a performance comparable to that of LSTM but greater computational efficiency. Li *et al.* [134] used a deep bidirectional GRU neural network (two-layer bidirectional GRU) as rumor detection model. The model suffers from slow convergence. S and Chitturi [41] showed that it is difficult to determine whether one of the gated RNNs (LSTM, GRU) is more successful, and they are usually chosen based on the basis of the available computing resources. Girgis *et al.* [96] experimented with CNN, LSTM, Vanilla, and GRU. Vanilla suffers from a gradient vanishing problem, but GRU solves this issue. Though GRU is said to be the best outcome of their studies, it takes more training time. A bidirectional GRU was utilized by Singhania *et al.* [87] for word-by-word annotation. With preceding and subsequent words, it captures the word's meaning within the sentence. A study by Shu *et al.* [100] proposed a sentence-comment co-attention subnetwork model named dEFEND (Explainable fake news detection) utilizing news content and user comments for fake news detection. The authors considered textual information with bidirectional GRU (Bi-GRU) to achieve better performance. Moreover, the model has a low learning efficiency.

C. GRAPH NEURAL NETWORK (GNN)

A Graph Neural Network is a form of neural network that operates on the graph structure directly. Node classification is a common application of GNN. Essentially, every node in the network has a label, and the network predicts the labels of the nodes without using the ground truth. The network extends recursive neural networks by processing a broader class of graphs, including directed, undirected graphs, and cyclic, and it can handle node-focused applications except any pre-processing steps [138]. The network extends recursive neural networks by processing a broader class of graphs, including cyclic, directed, and undirected graphs, and it can handle node-focused applications without requiring any pre-processing procedures cite190. GNN captures global structural features from graphs or trees better than the deep-learning models discussed above [139]. GNNs are prone to noise in the datasets. Adding a little amount of noise to the graph via node perturbation or edge deletion and addition has an antagonistic effect on the GNN output. Graph convolutional network (GCN) is considered as one of the basic graph neural networks variants.

A study by Huang *et al.* [140] claimed to be the first that experimented using a rich structure of user behavior for rumor detection. The user encoder uses graph convolutional networks (GCN) to learn a representation of the user from a graph created by user behavioral information. The authors used two recursive neural networks based on tree structure: bottom-up RvNN encoder and top-down RvNN encoder. The tree structure is shown in Figure 8. The proposed model performed worse for the non-rumor class cause user behavior information brings some interference in non-rumor detection.

Another study by Bian *et al.* [139] proposed top-down GCN and bottom-up GCN using a novel method DropEdge [141] for reducing over-fitting of GCNs. In addition, a root feature enhancement operation is utilized to improve the performance of rumor detection. Although it performed well on three datasets (Weibo, Twitter15, Twitter16), the outliers in the dataset affected the models' performance.

On the other hand, GCNs incur a significant memory footprint in storing the complete adjacency matrix. Furthermore, GCNs are transductive, which implies that inferred nodes must be present at the training time. And do not guarantee generalizable representations [142]. Wu *et al.* [143] proposed an algorithm of representation learning with a gated graph neural network named PGNN (propagation graph neural network). The suggested technique can incorporate structural and textual features into high-level representations by propagating information among neighbor nodes throughout the propagation network. In order to obtain considerable performance improvements, they also added an attention mechanism. The propagation graph is built using the who-replies-to-whom structure, but the follower-followee and forward relationships are omitted. Zhang *et al.* [144] presented a simplified aggregation graph neural network (SAGNN) based on efficient aggregation layers. Experiments on publicly accessible Twitter datasets show that the proposed network outperforms state-of-the-art graph convolutional networks while considerably lowering computational costs.

D. GENERATIVE ADVERSARIAL NETWORK (GAN)

Generative Adversarial Networks (GANs) are deep learning-based generative models. The GAN model architecture consists of two sub-models: a generator model for creating new instances and a discriminator model for determining whether the produced examples are genuine or fake, generated by the generator model. Existing adversarial networks are often employed to create images that may be matched to observed samples using a minimax game framework [44]. The generator model produces new images from the features learned from the training data that resemble the original image. The discriminator model predicts whether the generated image is fake or real. GANs are extremely successful in generative modeling and are used to train discriminators in a semi-supervised context to assist in eliminating human participation in data labeling. Furthermore, GANs are useful when the data have imbalanced classes or underrepresented samples. GANs produce synthetic data only if they are based on continuous numbers. But GANs are inapplicable to NLP data because all NLPs are based on discrete values such as words, letters, or bytes [145]. To train GANs for text data, novel techniques are required.

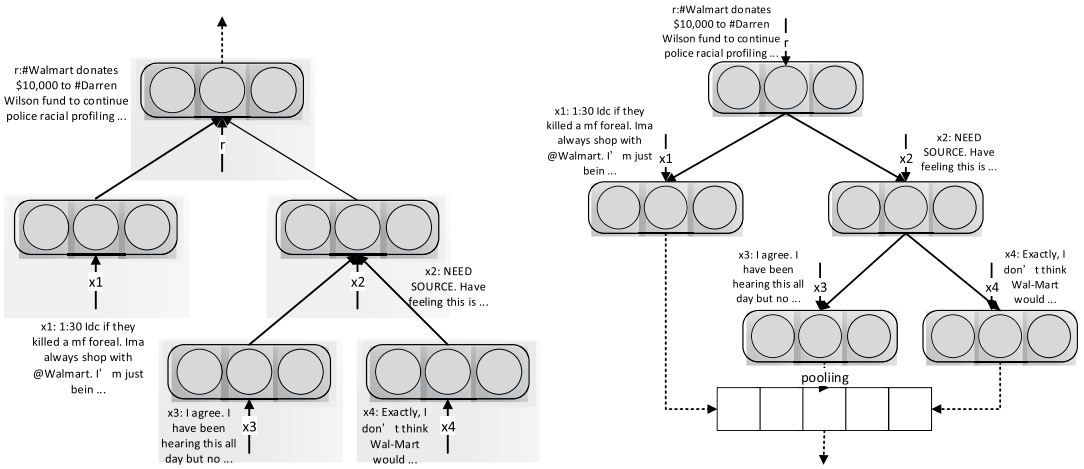
A study by Long [145] proposed sequence GAN (SeqGAN), which is a GAN architecture that overcomes the problem of gradient descent in GANs for discrete outputs by employing reinforcement learning (RL) based approach and Monte Carlo search. The authors provide actual news content to the GAN. Then a classifier based on Google's BERT model was trained to identify the real samples from the samples generated by the GAN. The architecture of SeqGAN is provided in Figure 9.

In generative adversarial networks, the principle of adversarial learning was invented. The adversarial learning concept has produced outstanding results in a wide range of topics, including information retrieval [146], text classification [147], and network embedding [148]. The unique problem for detecting fake news is the recognition of false news on recently emergent events on social media. To solve this problem, Wang *et al.* [44] suggested an end-to-end architecture called event adversarial neural network (EANN). This architecture is used to extract event-invariant characteristics and, therefore, aids in the identification of false news on newly incoming events. It is made up of three major components: a multimodal feature extractor, a fake news detector, and an event discriminator. Another study by Le *et al.* [149] introduced Malcom that generates malicious comments which have fooled five popular fake news detectors (CSI, dEFEND, etc.) to detect fake news as real news with 94% and 90% attack success rates. The authors showed that existing methods are not resilient against potential attacks. Though the model performed well, it is not evaluated using defense mechanisms, namely adversarial learning.

E. ATTENTION MECHANISM BASED

The attention-related approach is another notable advancement. In deep neural networks, the attention mechanism is an effort to implement the same behavior of selectively focusing on a few important items while ignoring others. Attention is a bridge that connects the encoder and decoder, which provides information to the decoder from each encoder's secret state. Using this framework, the model selectively concentrates on the valuable components from the input. Thus the model will be able to discover the associations among them. This allows the model to deal with lengthy input sentences more effectively. Unlike RNNs or CNNs, attention mechanisms maintain word dependencies in a sentence despite the distance between them. The primary downside of the attention mechanism is that it adds additional weight parameters to the model, which might lengthen the training time, especially if the model's input data are long sequences.

A study by Long [150] proposed attention-based LSTM with speaker profile features, and their experimental findings suggest that employing speaker profiles can help enhance fake news identification. Recently, attention techniques have been used to efficiently extract information related to a mini query (article headline) from a long text (news content) [47], [87]. A study by Singhania *et al.* [87] used an automated detector through a three-level hierarchical attention network (3HAN). Three levels exist in 3HAN, one for words, one for sentences, and one for the headline. Because of its three levels of attention, 3HAN assigns different weights to different sections of an article. In contrast to other deep learning models, 3HAN yields understandable results. While 3HAN only uses textual information, a study by Jin *et al.* [47] used image features, including social context and text features, as well as attention on RNN (att-RNN). Another study used RNNs with a soft-attention mechanism to filter out unique linguistic features [151]. However, this method is based on distinct domain and community features without any external evidence. Thus, it provides a restricted context for credibility analysis.



(a) Bottom-up RvNN encoder (b) Top-down RvNN encoder

FIGURE 8. This figure illustrates the propagation tree structure encoder taken from Huang *et al.* [140].

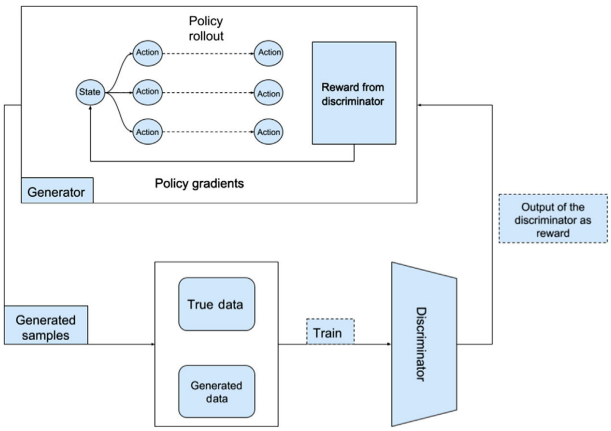


FIGURE 9. A basic SeqGAN architecture. The figure is taken from Hiriyannaiah *et al.* [145].

To overcome the shortcomings of previous works, Aloshban [152] proposed an automatic fake news classification through self-attention (ACT). Their principle is inspired by the fact that claim texts are fairly short and hence cannot be used for classification efficiently. Their suggested framework makes use of mutual interactions between a claim and many supporting responses. The LSTM neural network was applied to the article input. The outcome of the final step of LSTM may not completely reflect the semantics of the article. Connecting all vector representations of words in the text will lead to a massive vector dimension. Therefore, the internal connection between the articles' words can be ignored. As a result, employing the self-attention function on the LSTM model extracts key parts of the article through several feature vectors. Their strategy is heavily reliant on self-attention and an article representation matrix. Graph-aware co-attention networks (GCAN) is an innovative approach for detecting fake news [153]. The authors predict if a source tweet article is false based just on its brief text content and user retweet sequence, as well as user profiles. Given the chronology of its retweeters, GCAN can determine whether a short-text tweet is fraudulent. However, this model is not suitable for long text as it is difficult to find the relationship between a long tweet and retweet propagation.

F. BIDIRECTIONAL ENCODER REPRESENTATIONS FOR TRANSFORMERS (BERT)

BERT is a deep learning model that has shown cutting-edge results across a wide variety of natural language processing applications. BERT incorporates pre-training language representations developed by Google. BERT is a sophisticated pre-trained word-embedding model built on a transformer-encoded architecture [89]. The BERT method is distinctive in its capacity to identify and capture contextual meaning in a sentence or text [90]. The main restriction of conventional language models is that they are unidirectional, which restricts the architectures that could be utilized during pre-training. The BERT model eliminates unidirectional limitations by using a mask language model (MLM). BERT employs the next sentence prediction (NSP) task in addition to the masked language model to jointly pre-train text-pair representations. BERT consists of two stages: pre-training and fine-tuning. During pre-training, the model was trained on unlabeled data using a variety of pre-training tasks. For fine-tuning, the BERT model is first initialized with the pre-trained parameters, and then all of the parameters are fine-tuned using labeled data from the downstream jobs. The architecture of the BERT model is shown in figure 10.

The data utilized in the BERT model are generic data gathered from Wikipedia and the Book Corpus. While these data contain a wide range of information, specific information on individual domains is still lacking. To overcome this problem, a study by Jwa *et al.* [75] incorporated news data in the pre-training phase to boost fake news identification skills. When compared to the state-of-the-art model stackLSTM,

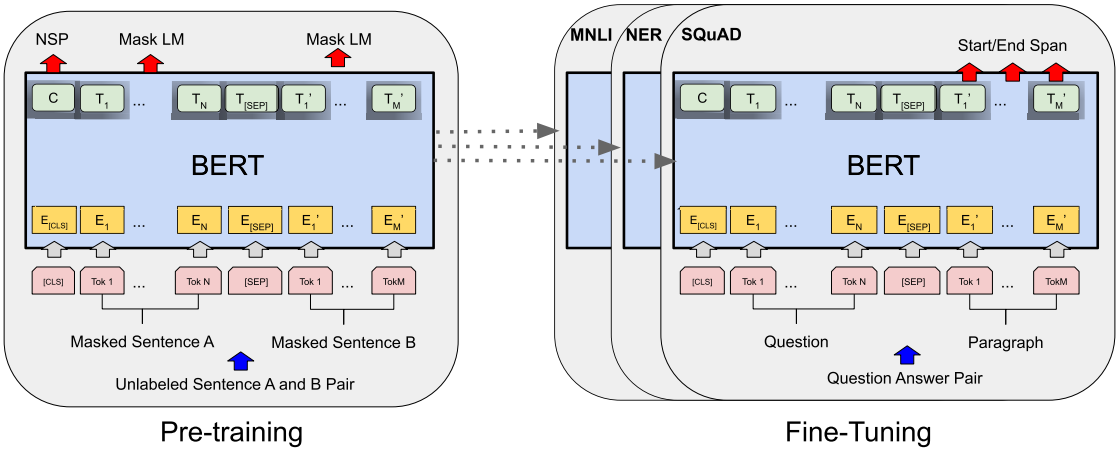


FIGURE 10. The BERT architecture taken from Devlin *et al.* [89].

the proposed model named exBAKE (BERT with extra unlabeled news corpora) outperformed by a 0.137 F1-score. Ding *et al.* [154] discovered that including mental features such as a speaker's credit history at the language level might considerably improve BERT model performance. The history feature helps further the relationship's construction between the event and the person in reality. But these studies did not consider any pre-processing methods.

Zhang *et al.* [91] presented a BERT-based domain-adaption neural network for multimodal false news detection (BDANN). BDANN is made up of three major components: a multimodal feature extractor, a domain classifier, and a false news detector. The pre-trained BERT model was used to extract text features, whereas the pre-trained VGG-19 model was used to extract image features in the multimodal feature extractor. The extracted features are then concatenated and sent to the detector to differentiate between fake and real news. Moreover, the existence of noisy images in the Weibo dataset have affected the BDANN results. Kaliyar *et al.* [92] proposed a BERT-based deep convolutional approach (fakeBERT) for fake news detection. The fakeBERT is a combination of different parallel blocks of a one-dimensional deep convolutional neural network (1d-CNN) with different kernel sizes and filters and the BERT. Different filters can extract convenient information from the training dataset. The combination of BERT with 1d-CNN can deal with both large-scale structure and unstructured text. Therefore, the combination is beneficial in dealing with ambiguity.

G. ENSEMBLE APPROACH

Ensemble approaches are strategies that generate several models and combine them to achieve better results. Ensemble models typically yield more precise solutions than a single model does. An ensemble reduces the distribution or dispersion of predictions and model efficiency. Ensembling can be applied to supervised and unsupervised learning activities [86]. Many researchers have used an ensemble approach to boost their performance [42], [133]. Agarwal and Dixit [63] combined two datasets, namely, Liar and Kaggle, to evaluate the performance of LSTM and achieved an accuracy of 97%. They also used various models like CNN, LSTM, SVM, naive bayes (NB), and k-nearest neighbour (KNN) for building an ensemble model. The authors showed an average accuracy score of their used algorithms but did not show the accuracy of their ensemble model, which is a limitation of their work.

Often the CNN-LSTM ensemble approach has been used in previous DL-based studies. Kaliyar [67] used an ensemble of CNN and LSTM, and the accuracy was slightly lower than that of the state-of-the-art CNN model. However, the precision and recall were effectively improved. Asghar *et al.* [135] obtained an increase in the efficiency of their model by using Bi-LSTM. The Bi-LSTM retains knowledge from both former and upcoming contexts before rendering its input to the CNN model. Even though CNN and RNN typically require huge datasets to function successfully, Ajao *et al.* [133] trained LSTM-CNN with a smaller dataset. The above-mentioned works considered just text-based features for fake news classification, whereas the addition of new features may generate a more significant result. While most studies used CNN with LSTM, a study by Amine *et al.* [131] merged two convolutional neural networks to integrate metadata with text. They illustrate that integrating metadata with text will result in substantial improvements in fine-grained fake news detection. Furthermore, when tested on real-world datasets, this approach shows improvements compared to the text-only deep learning model. Moving further Kumar *et al.* [86] employed the use of an attention layer. It assists the CNN + LSTM model in learning to pay attention to particular regions of input sequences rather than the full series of input sequences. Utilizing the attention mechanism with CNN+LSTM was reported to be efficient by a small margin. Result analysis of DL-based studies is presented in Table 7.

TABLE 6. The table contains the strength and limitation of popular existing studies with reference and used classifier.

Reference	Dataset	Classifier	Strength	Limitation
Kaliyar et al. [40]	Fake news	Deep CNN	The model is less prone to overfitting.	The training process takes a longer time.
Sahoo and Gupta [138]	Facebook	LSTM	Crawling and the Facebook API are used to retrieve a set of new features.	The suggested model requires extra time for training and testing.
Liao et al. [139]	LIAR	Bidirectional LSTM	Tackles fake news detection task and news topic classification task together in a unified approach through multi-task learning.	The performance of the model depends on author information
Ruchansky et al. [125]	Weibo & Twitter	RNN	Extracting meaningful latent representations.	Expensive computational cost
Shu et al. [101]	FakeNewsNet	Bidirectional GRU	Provide a novel way to exploit both news content and user comments for misleading news detection and prediction.	Increased training time.
Asghar et al. [137]	Pheme	Bi-LSTM+CNN	The model preserves the sequence information in both directions.	The suggested approach is computationally expensive.
Umer et al. [42]	FNC-1	CNN+LSTM	When compared to pre-trained BERT, the combined CNN+LSTM with PCA and Chi-square performed better.	Because PCA text messages may not have linear connection, some information may be lost, and so the underlying model is dependent on feature extraction.
Albahar [121]	FakeNewsNet	RNN+SVM	The model exploits user comments to improve the detection rate.	Performance depends on feature vector size.
Chen et al. [153]	Twitter and Weibo	attention+RNN	Capable of learning continuous latent representations by capturing long-term dependence and contextual changes in posting series.	Large number of weights parameter.
Wang et al. [44]	Twitter and Weibo	EANN	The model is capable of learning transferrable features for unseen events.	Trained on a imbalanced dataset
Huang et al. [142]	Twitter15 and Twitter16	GNN	Adequately extract user information	User behavior information bring some interference to the detection of non-rumor.
Jwa et al. [75]	FNC-1	BERT (exBAKE)	Incorporating extra knowledge from large news corpora	Absence of data pre-processing.

VI. EVALUATION METRICS

A key step in a predictive modeling pipeline is to evaluate the output of a machine-learning model. Although a model may have a higher classification result once constructed, it must be determined whether it can address the specific problem in different circumstances. Classification accuracy alone is usually insufficient to make this judgment. Other assessment metrics are necessary for proper evaluation. Since a promising method is required to pass the assessment metric's evaluation, it is easy to create a model, but it is more challenging to create a promising strategy. Diverse evaluation metrics are used to evaluate the model's efficiency. The evaluation matrix is an essential device for arranging and organizing an evaluation. The confusion matrix shows an overview of model performance on the testing dataset from the known true values. It provides a review of the model's success and useful results of true positive, true negative, false positive, and false negative. To test their models, researchers considered distinctive sorts of metrics such as accuracy (A), precision (P), and recall (R) [40], [54], [58]. The selection of metrics relies entirely on the model form and its implementation strategy. We provide some evaluation metrics that were widely used in previous studies:

A. ACCURACY

The accuracy score, also known as the classification accuracy rating, is determined as the percentage of accurate predictions in proportion to the total predictions made by the model. The accuracy (A) can be depicted by the given formula in Equation (1).

$$A = \frac{TruePositive + TrueNegative}{TotalNumberofPredictions} \tag{1}$$

B. PRECISION

Precision (P) is defined as the number of actual positive findings divided by the total number of positive results, including incorrectly recognized ones. The precision can be computed using Equation (2).

$$P = \frac{TruePositive}{Positive + FalsePositive} \tag{2}$$

C. RECALL

When the total number of samples that should have been identified as positive is used to divide, the number of true positive results is referred to as recall (R). The recall can be computed using Equation (3).

$$R = \frac{TruePositive}{TruePositive + FalseNegative} \tag{3}$$

D. F1-SCORE

The model's accuracy for each class is defined by the F1-score (F1). If the dataset is not balanced, the F1-score metric is typically used. The F1-score is often used as an assessment matrix in fake news detection [41], [157], [158]. F1-score computation can be performed using Equation (4).

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{4}$$

E. ROC CURVE AND AUC

The Receiver Operating Characteristics (ROC) curve shows the success of a classification model across several classification thresholds. True Positive Rate (Recall) and False Positive Rate (FPR) are used in this curve. AUC is an abbreviation for "Area Under the ROC curve." In other words, AUC tests the

TABLE 7. The table contains the result in accuracy of DL-based studies along with used method and NLP techniques.

Dataset	Method	NLP Techniques	Accuracy	Reference
Fake News	CNN	TF-IDF	98.3%	Kaliyar [67]
	Deep CNN	GloVe	98.36%	Kaliyar et al. [40]
	CNN	Tensorflow embedding layer	96%	Amine et al. [133]
	CNN+LSTM	GloVE	94.71%	K. Shu et al. [50]
	Bi-directional LSTM-RNN	GloVe	98.75%	Bahad et al. [61]
	Passive aggressive	TF-IDF	83.8%	Mandical et al. [48]
	fakeBERT	GloVe, BERT	98.90%	Kaliyar et al. [92]
LIAR	GRU, LSTM, StackedCNN	Word2vec	47.2%, 46.8%, 48.5%	Fernández-Reyes and Shinde [77]
	CNN	-	27%	Girgis et al. [96]
	RCNN	Word2vec	33.7%	Wu et al. [158]
	DSSM-LSTM	TF-IDF	99%	Jadhav and Thepade [49]
	FDML	GloVe	50.8%	Liao et al. [139]
	Passive aggressive	TF-IDF	99%	Mandical et al. [48]
FNC-I	CNN+LSTM	PCA	97.8%	Umer et al. [42]
	Dense Neural Network	TF-IDF	94.31%	Thota et al. [46]
	Bidirectional LSTM concatenated	Word2vec	85.3%	Qawasmeh et al. [62]
	Bidirectional LSTM	Word2vec	94%	Padnekar et al. [64]
PHEME	BiLSTM-CNN	Keras embedding layer	86.12%	Asghar et al. [137]
	CNN	Keras embedding layer	87.1%	Alsaeedi and Al-Sarem [45]
	LSTM Model, LSTM-CNN	- -	82.29%, 80.38%	Ajao et al. [135]
Twitter15	GNN	- -	75.2%	Huang et al. [142]
Twitter16	LSTM, GRU	Word2vec, GloVe	56.98%, 67.03%	Bugueño et al. [80]
	GNN	- -	77.3%	Huang et al. [142]
PolitiFact	3HAN	GloVe	96.77%	Singhania et al. [87]
	DNN	-	92.30%	Choudhary and Arora [159]
The George McIntire dataset	LSTM	Word2vec, GloVe	91.32%	S and Chitturi [41]
Twitter, Weibo	BERT	BERT	83%, 85%	Zhang et al. [91]
	EANN	-	71%, 82%	Wang et al. [44]

whole two-dimensional field under the entire ROC curve. The FPR can be defined as in Equation (5).

$$FPR = \frac{FalsePositive}{FalsePositive + TrueNegative} \tag{5}$$

VII. CHALLENGES AND RESEARCH DIRECTION

Despite the fact that numerous studies have been conducted on the identification of fake news, there is always space for future advancement and investigation. In the sense of recognizing fake news, we highlight challenges and several unique exploration areas for future studies. Although DL-based methods provide higher accuracy compared to the other methods, there is scope to make it more acceptable.

- The feature and classifier selection greatly influences the efficiency of the model. Previous studies did not place a high priority on the selection of features and classifiers. Researchers should focus on determining which classifier is most suitable for particular features. The long textual features require the use of sequence models (RNNs), but limited research works have taken this into account. We believe that studies that concentrate on the selection of features and classifiers might potentially improve performance.
- The feature engineering concept is not common in deep learning-based studies. News content and headline features are the widely used features in fake news detection, but several other features such as user behavior [154], user profile, and social network behavior need to be explored. Political or religious bias in profile features and lexical, syntactic, and statistical-based features can increase the detection rate. A fusion of deeply hidden text features with other statistical features may result in a better outcome.
- Propagation-based studies are scarce in this domain [117]. Network-based patterns of news propagation are a piece of information that has not been comprehensively utilized for fake news detection [159]. Thus, we suggest considering news propagation for fake news identification. Meta-data and additional information can increase the robustness and reduce the noise of a single textual claim, but they must be handled with caution.
- Studies focused only on text data for fake news detection, whereas fake news is generated in sophisticated ways, with text or images that have been purposefully altered [95]. Only a few studies have used image features [109], [110]. Thus, we recommend the use of visual data (videos and images). An examination with video and image features will be an investigation region to build a stronger and more robust system.
- Studies that use a fusion of features are scarce in this domain [160]. Combining information from multiple sources may be extremely beneficial in detecting whether Internet articles are fake [95]. We suggest utilizing multi-model-based approaches with later pretrained word embeddings. Many other hidden features may have a great impact on fake news detection. Hence we encourage researchers to investigate hidden features.
- Fake news detection models that learn from newly emerging web articles in real-time could enhance detection results. Another promising future work is the use of a transfer-learning approach for training a neural network with online data streams.
- More data for a more significant number of fake news should be released since the lack of data is the major problem in fake news classification. We assume that more training data will improve model performance.

Datasets focused on news content are publicly available. On the other hand, datasets based on different textual features are limited. Thus research utilizing additional textual features is scarce.

- Instead of a simple classifier, using an ensemble method produces better results [49]. By constructing an ensemble model with DL and ML algorithms, in which an LSTM can identify the original article while passing auxiliary features through a second model can yield better results [41]. A simpler GRU model performs better than an LSTM [80]. Therefore, we recommend combining GRU and CNNs to urge the leading result.
- Many researchers have achieved high accuracy by using CNN, LSTM, and ensemble models [42], [64]. SeqGAN and Deep Belief Network (DBN) were not explored in this domain. We encourage researchers to experiment with these models.
- Transformers have replaced RNN models such as LSTM as the model of choice for NLP tasks. BERT has been used in the identification of fake news, but Generative Pre-trained Transformer (GPT) has not been used in this domain. We suggest using GPT by fine-tuning fake news detection tasks.
- Existing algorithms make critical decisions without providing precise information about the reasoning that results in specific decisions, predictions, recommendations, or actions [161]. Explainable Artificial Intelligence (XAI) is a study field that tries to make the outcomes of AI systems more understandable to humans [162]. XAI can be a valuable approach to start making progress in this area.

VIII. CONCLUSION

Fake news is escalating as social media is growing. Researchers are also trying their best to find solutions to keep society safe from fake news. This survey covers the overall analysis of fake news classification by discussing major studies. A thorough understanding of recent approaches in fake news detection is essential because advanced frameworks are the front-runners in this domain. Thus, we analyzed fake news identification methods based on NLP and advanced DL strategies. We presented a taxonomy of fake news detection approaches. We explored different NLP techniques and DL architectures and provided their strength and shortcomings. We have explored diverse assessment measurements. We have given a short description of the experimental findings of previous studies. In this field, we briefly outlined possible directions for future research. Fake news identification will remain an active research field for some time with the emergence of novel deep learning network architectures. There are fewer chances of inaccurate results using deep learning-based models. We strongly believe that this review will assist researchers in fake news detection to gain a better, concise perspective of existing problems, solutions, and future directions.

ACKNOWLEDGMENT

The authors would like to thank the Advanced Machine Learning (AML) Lab for resource sharing and precious opinions.

REFERENCES

[1] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *J. Econ. Perspect.*, vol. 31, no. 2, pp. 36–211, 2017.
[2] T. Rasool, W. H. Butt, A. Shaukat, and M. U. Akram, "Multi-label fake news detection using multi-layered supervised learning," in *Proc. 11th Int. Conf. Comput. Autom. Eng.*, 2019, pp. 73–77.
[3] 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. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306457318306794
[4] Abdullah-All-Tanvir, E. M. Mahir, S. Akhter, and M. R. Huq, "Detecting fake news using machine learning and deep learning algorithms," in *Proc. 7th Int. Conf. Smart Comput. Commun. (ICSCC)*, Jun. 2019, pp. 1–5.
[5] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explorations Newslett.*, vol. 19, no. 1, pp. 22–36, 2017.
[6] R. Oshikawa, J. Qian, and W. Y. Wang, "A survey on natural language processing for fake news detection," 2018, *arXiv:1811.00770*.
[7] S. B. Parikh and P. K. Atrey, "Media-rich fake news detection: A survey," in *Proc. IEEE Conf. Multimedia Inf. Process. Retr. (MIPR)*, Apr. 2018, pp. 436–441.
[8] A. Habib, M. Z. Asghar, A. Khan, A. Habib, and A. Khan, "False information detection in online content and its role in decision making: A systematic literature review," *Social Netw. Anal. Mining*, vol. 9, no. 1, pp. 1–20, Dec. 2019.
[9] M. K. Elhadad, K. F. Li, and F. Gebali, "Fake news detection on social media: A systematic survey," in *Proc. IEEE Pacific Rim Conf. Commun., Comput. Signal Process. (PACRIM)*, Aug. 2019, pp. 1–8.
[10] A. Bondielli and F. Marcelloni, "A survey on fake news and rumour detection techniques," *Inf. Sci.*, vol. 497, pp. 38–55, Sep. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025519304372
[11] P. Meel and D. K. Vishwakarma, "Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities," *Expert Syst. Appl.*, vol. 153, Sep. 2020, Art. no. 112986.
[12] K. Sharma, F. Qian, H. Jiang, N. Ruchansky, M. Zhang, and Y. Liu, "Combating fake news: A survey on identification and mitigation techniques," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 3, pp. 1–42, May 2019.
[13] X. Zhou and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," *ACM Comput. Surv.*, vol. 53, no. 5, pp. 1–40, 2020.
[14] B. Collins, D. T. Hoang, N. T. Nguyen, and D. Hwang, "Trends in combating fake news on social media—A survey," *J. Inf. Telecommun.*, vol. 5, no. 2, pp. 247–266, 2021.
[15] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, and R. Procter, "Detection and resolution of rumours in social media: A survey," *ACM Comput. Surveys*, vol. 51, no. 2, pp. 1–36, Jun. 2018.
[16] M. D. Ibrishimova and K. F. Li, "A machine learning approach to fake news detection using knowledge verification and natural language processing," in *Proc. Int. Conf. Intell. Netw. Collaborative Syst.* Cham, Switzerland: Springer, 2019, pp. 223–234.
[17] H. Ahmed, I. Traore, and S. Saad, "Detecting opinion spams and fake news using text classification," *Secur. Privacy*, vol. 1, no. 1, p. e9, Jan. 2018.
[18] H. Ahmed, I. Traore, and S. Saad, "Detection of online fake news using N-gram analysis and machine learning techniques," in *Proc. Int. Conf. Intell., Secure, Dependable Syst. Distrib. Cloud Environ.* Switzerland: Springer, 2017, pp. 127–138.
[19] B. Bhutani, N. Rastogi, P. Sehgal, and A. Purwar, "Fake news detection using sentiment analysis," in *Proc. 12th Int. Conf. Contemp. Comput. (IC)*, Aug. 2019, pp. 1–5.
[20] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on Twitter," in *Proc. 20th Int. Conf. World Wide Web*, Mar. 2011, pp. 675–684, doi: 10.1145/1963405.1963500.

[21] O. Ajao, D. Bhowmik, and S. Zargari, "Sentiment aware fake news detection on online social networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 2507–2511.

[22] B. Ghanem, P. Rosso, and F. Rangel, "An emotional analysis of false information in social media and news articles," *ACM Trans. Internet Technol.*, vol. 20, no. 2, pp. 1–18, May 2020.

[23] A. Giachanou, P. Rosso, and F. Crestani, "Leveraging emotional signals for credibility detection," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2019, pp. 877–880.

[24] D. Khattar, J. S. Goud, M. Gupta, and V. Varma, "MVAE: Multimodal variational autoencoder for fake news detection," in *Proc. World Wide Web Conf.*, May 2019, pp. 2915–2921.

[25] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," in *Proc. 78th ASIST Annu. Meeting, Inf. Sci. Impact, Res. Community*, vol. 52, no. 1, pp. 1–4, 2015.

[26] A. R. Pathak, A. Mahajan, K. Singh, A. Patil, and A. Nair, "Analysis of techniques for rumor detection in social media," *Proc. Comput. Sci.*, vol. 167, pp. 2286–2296, Jan. 2020.

[27] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, and M. Cha, "Detecting rumors from microblogs with recurrent neural networks," in *Proc. 25th Int. Joint Conf. Artif. Intell. (IJCAI)*. Res. Collection School Comput. Inf. Syst., 2016, pp. 3818–3824.

[28] J. Ma, W. Gao, and K.-F. Wong, "Detect rumors in microblog posts using propagation structure via kernel learning," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics (ACL)*. Vancouver, BC, Canada: Res. Collection School Comput. Inf. Syst., Jul./Aug. 2017, pp. 708–717.

[29] W. Y. Wang, "'Liar, liar pants on fire': A new benchmark dataset for fake news detection," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics*, Vancouver, BC, Canada, Jul. 2017, pp. 422–426. [Online]. Available: https://www.aclweb.org/anthology/P17-2067

[30] A. Zubiaga, M. Liakata, and R. Procter, "Learning reporting dynamics during breaking news for rumour detection in social media," 2016, *arXiv:1610.07363*.

[31] 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.

[32] M. Amjad, G. Sidorov, A. Zhila, H. Gómez-Adorno, I. Voronkov, and A. Gelbukh, "'Bend the truth': Benchmark dataset for fake news detection in Urdu language and its evaluation," *J. Intell. Fuzzy Syst.*, vol. 39, no. 2, pp. 2457–2469, 2020.

[33] E. Tacchini, G. Ballarin, M. L. Della Vedova, S. Moret, and L. de Alfaro, "Some like it hoax: Automated fake news detection in social networks," 2017, *arXiv:1704.07506*.

[34] C. Boididou, S. Papadopoulos, and M. Zampoglou, "Detection and visualization of misleading content," *Int. J. Multimedia Inf. Retr.*, vol. 7, no. 1, pp. 71–86, 2018.

[35] J. Golbeck, M. Mauriello, B. Auxier, K. H. Bhanushali, C. Bonk, M. A. Bouzaghrane, C. Buntain, R. Chanduka, P. Cheakalos, J. B. Everett, and W. Falak, "Fake news vs satire: A dataset and analysis," in *Proc. 10th ACM Conf. Web Sci.*, 2018, pp. 17–21.

[36] P. M. Waszak, W. Kasprzycka-Waszak, and A. Kubanek, "The spread of medical fake news in social media—The pilot quantitative study," *Health Policy Technol.*, vol. 7, no. 2, pp. 115–118, Jun. 2018.

[37] (2020). *The Year of Fake News Covid Related Scams and Ransomware*. Accessed: Mar. 12, 2021. [Online]. Available: https://www.prnewswire.com/news-releases/2020-the-year-of-fake-news-covid-related-scams-and-ransomware-301180568

[38] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "FakeNewsNet: A data repository with news content, social context and spatial-temporal information for studying fake news on social media," 2018, *arXiv:1809.01286*.

[39] Y.-C. Ahn and C.-S. Jeong, "Natural language contents evaluation system for detecting fake news using deep learning," in *Proc. 16th Int. Joint Conf. Comput. Sci. Softw. Eng. (JCSSE)*, Jul. 2019, pp. 289–292.

[40] R. K. Kaliyar, A. Goswami, P. Narang, and S. Sinha, "FNDNet—A deep convolutional neural network for fake news detection," *Cognit. Syst. Res.*, vol. 61, pp. 32–44, Jun. 2020.[Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389041720300085

[41] S. Deepak and B. Chitturi, "Deep neural approach to Fake-News identification," *Proc. Comput. Sci.*, vol. 167, pp. 2236–2243, Jan. 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050920307420

[42] 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.

[43] 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.

[44] 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*, Jul. 2018, pp. 849–857.

[45] A. Alsaeedi and M. Al-Sarem, "Detecting rumors on social media based on a CNN deep learning technique," *Arabian J. Sci. Eng.*, vol. 45, no. 12, pp. 1–32, 2020.

[46] A. Thota, P. Tilak, S. Ahluwalia, and N. Lohia, "Fake news detection: A deep learning approach," *SMU Data Sci. Rev.*, vol. 1, no. 3, p. 10, 2018.

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

[48] R. R. Mandical, N. Mamatha, N. Shivakumar, R. Monica, and A. N. Krishna, "Identification of fake news using machine learning," in *Proc. IEEE Int. Conf. Electron., Comput. Commun. Technol. (CONECCT)*, Jul. 2020, pp. 1–6.

[49] S. S. Jadhav and S. D. Thepade, "Fake news identification and classification using DSSM and improved recurrent neural network classifier," *Appl. Artif. Intell.*, vol. 33, no. 12, pp. 1058–1068, Oct. 2019, doi: 10.1080/08839514.2019.1661579.

[50] A. S. K. Shu, D. M. K. Shu, L. G. M. Mittal, L. G. M. Mittal, and M. M. J. K. Sethi, "Fake news detection using a blend of neural networks: An application of deep learning," *Social Netw. Comput. Sci.*, vol. 1, no. 3, pp. 1–9, Jan. 1970. [Online]. Available: https://link.springer.com/article/10.1007/s42979-020-00165-4

[51] A. P. S. Bali, M. Fernandes, S. Choubey, and M. Goel, "Comparative performance of machine learning algorithms for fake news detection," in *Proc. Int. Conf. Adv. Comput. Data Sci.* Switzerland: Springer, 2019, pp. 420–430.

[52] A. Rusli, J. C. Young, and N. M. S. Iswari, "Identifying fake news in Indonesian via supervised binary text classification," in *Proc. IEEE Int. Conf. Ind. 4.0, Artif. Intell., Commun. Technol. (IAICT)*, Jul. 2020, pp. 86–90.

[53] V. Tiwari, R. G. Lennon, and T. Dowling, "Not everything you read is true! Fake news detection using machine learning algorithms," in *Proc. 31st Irish Signals Syst. Conf. (ISSC)*, Jun. 2020, pp. 1–4.

[54] A. Verma, V. Mittal, and S. Dawn, "FIND: Fake information and news detections using deep learning," in *Proc. 12th Int. Conf. Contemp. Comput. (IC)*, Aug. 2019, pp. 1–7.

[55] M. Z. Hossain, M. A. Rahman, M. S. Islam, and S. Kar, "BanFakeNews: A dataset for detecting fake news in Bangla," in *Proc. 12th Lang. Resour. Eval. Conf.* Marseille, France: European Language Resources Association, May 2020, pp. 2862–2871. [Online]. Available: https://www.aclweb.org/anthology/2020.lrec-1.349

[56] P. Savyan and S. M. S. Bhanu, "UbCadet: Detection of compromised accounts in Twitter based on user behavioural profiling," *Multimedia Tools Appl.*, vol. 79, pp. 1–37, Jul. 2020.

[57] J. Kapusta and J. Obonya, "Improvement of misleading and fake news classification for flective languages by morphological group analysis," in *Informatics*, vol. 7, no. 1. Switzerland: Multidisciplinary Digital Publishing Institute, 2020, p. 4.

[58] 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. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X20330466

[59] M. G. Hussain, M. Rashidul Hasan, M. Rahman, J. Protim, and S. A. Hasan, "Detection of Bangla fake news using MNB and SVM classifier," in *Proc. Int. Conf. Comput., Electron. Commun. Eng. (iCCECE)*, Aug. 2020, pp. 81–85.

[60] G. Gravanis, A. Vakali, K. Diamantaras, and P. Karadais, "Behind the cues: A benchmarking study for fake news detection," *Expert Syst. Appl.*, vol. 128, pp. 201–213, Aug. 2019.

[61] 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. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050920300806

[62] E. Qawasmeh, M. Tawalbeh, and M. Abdullah, "Automatic identification of fake news using deep learning," in *Proc. 6th Int. Conf. Social Netw. Anal., Manage. Secur. (SNAMS)*, Oct. 2019, pp. 383–388.

[63] A. Agarwal and A. Dixit, "Fake news detection: An ensemble learning approach," in *Proc. 4th Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, May 2020, pp. 1178–1183.

[64] S. M. Padnekar, G. S. Kumar, and P. Deepak, "BiLSTM-autoencoder architecture for stance prediction," in *Proc. Int. Conf. Data Sci. Eng. (ICDSE)*, Dec. 2020, pp. 1–5.

[65] M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," in *Proc. IEEE 1st Ukraine Conf. Electr. Comput. Eng. (UKRCON)*, May 2017, pp. 900–903.

[66] A. Jain and A. Kasbe, "Fake news detection," in *Proc. IEEE Int. Students' Conf. Electr., Electron. Comput. Sci. (SCEECS)*, 2018, pp. 1–5.

[67] R. K. Kaliyar, "Fake news detection using a deep neural network," in *Proc. 4th Int. Conf. Comput. Commun. Autom. (ICCCA)*, Dec. 2018, pp. 1–7.

[68] G. Bhatt, A. Sharma, S. Sharma, A. Nagpal, B. Raman, and A. Mittal, "Combining neural, statistical and external features for fake news stance identification," in *Proc. Companion The Web Conf. Web Conf. (WWW)*, 2018, pp. 1353–1357, doi: 10.1145/3184558.3191577.

[69] F. A. Ozbay and B. Alatas, "Fake news detection within online social media using supervised artificial intelligence algorithms," *Phys. A, Stat. Mech. Appl.*, vol. 540, Feb. 2020, Art. no. 123174.

[70] B. Al-Ahmad, A. M. Al-Zoubi, R. A. Khurma, and I. Aljarah, "An evolutionary fake news detection method for COVID-19 pandemic information," *Symmetry*, vol. 13, no. 6, p. 1091, Jun. 2021.

[71] S. Shabani and M. Sokhn, "Hybrid machine-crowd approach for fake news detection," in *Proc. IEEE 4th Int. Conf. Collaboration Internet Comput. (CIC)*, Oct. 2018, pp. 299–306.

[72] C. M. M. Kotteti, X. Dong, N. Li, and L. Qian, "Fake news detection enhancement with data imputation," in *Proc. IEEE 16th Int. Conf. Dependable, Autonomic Secure Comput., 16th Int. Conf. Pervasive Intell. Comput., 4th Int. Conf. Big Data Intell. Comput. Cyber Sci. Technol. Congr. (DASC/PiCom/DataCom/CyberSciTech)*, Aug. 2018, pp. 187–192.

[73] X. Zhou, A. Jain, V. V. Phoha, and R. Zafarani, "Fake news early detection: A theory-driven model," *Digit. Threats, Res. Pract.*, vol. 1, no. 2, pp. 1–25, Jul. 2020.

[74] P. H. A. Faustini and T. F. Covões, "Fake news detection in multiple platforms and languages," *Expert Syst. Appl.*, vol. 158, Nov. 2020, Art. no. 113503.

[75] H. Jwa, D. Oh, K. Park, J. Kang, and H. Lim, "ExBAKE: Automatic fake news detection model based on bidirectional encoder representations from transformers (BERT)," *Appl. Sci.*, vol. 9, no. 19, p. 4062, Sep. 2019.

[76] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 3111–3119.

[77] F. C. Fernández-Reyes and S. Shinde, "Evaluating deep neural networks for automatic fake news detection in political domain," in *Proc. Ibero-Amer. Conf. Artif. Intell.*, Nov. 2018, pp. 206–216. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-03928-8_17

[78] C. K. Hiramath and G. C. Deshpande, "Fake news detection using deep learning techniques," in *Proc. 1st Int. Conf. Adv. Inf. Technol. (ICAIT)*, Jul. 2019, pp. 411–415.

[79] A. P. B. Veyseh, M. T. Thai, T. H. Nguyen, and D. Dou, "Rumor detection in social networks via deep contextual modeling," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Aug. 2019, pp. 113–120.

[80] M. Bugueño, G. Sepulveda, and M. Mendoza, "An empirical analysis of rumor detection on microblogs with recurrent neural networks," in *Proc. Int. Conf. Hum.-Comput. Interact.*, Jul. 2019, pp. 293–310. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-21902-4_21

[81] E. Providel and M. Mendoza, "Using deep learning to detect rumors in Twitter," in *Proc. Int. Conf. Hum.-Comput. Interact.* Switzerland: Springer, 2020, pp. 321–334.

[82] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 1188–1196.

[83] S. Sangamnerkar, R. Srinivasan, M. R. Christhuraj, and R. Sukumaran, "An ensemble technique to detect fabricated news article using machine learning and natural language processing techniques," in *Proc. Int. Conf. Emerg. Technol. (INCET)*, Jun. 2020, pp. 1–7.

[84] S. Helmstetter and H. Paulheim, "Weakly supervised learning for fake news detection on Twitter," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2018, pp. 274–277.

[85] J. Pennington, R. Socher, and C. Manning, "GloVe: Global vectors for word representation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1532–1543.

[86] S. Kumar, R. Asthana, S. Upadhyay, N. Upreti, and M. Akbar, "Fake news detection using deep learning models: A novel approach," *Trans. Emerg. Telecommun. Technol.*, vol. 31, no. 2, p. e3767, Feb. 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/ett.3767

[87] S. Singhania, N. Fernandez, and S. Rao, "3HAN: A deep neural network for fake news detection," in *Proc. Int. Conf. Neural Inf. Process.* Switzerland: Springer, 2017, pp. 572–581.

[88] J. A. Nasir, O. S. Khan, and I. Varlamis, "Fake news detection: A hybrid CNN-RNN based deep learning approach," *Int. J. Inf. Manage. Data Insights*, vol. 1, no. 1, Apr. 2021, Art. no. 100007.

[89] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.

[90] S. Kula, M. Choraś, and R. Kozik, "Application of the bert-based architecture in fake news detection," in *Proc. Comput. Intell. Secur. Inf. Syst. Conf.* Switzerland: Springer, 2019, pp. 239–249.

[91] T. Zhang, D. Wang, H. Chen, Z. Zeng, W. Guo, C. Miao, and L. Cui, "BDANN: BERT-based domain adaptation neural network for multimodal fake news detection," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–8.

[92] R. K. Kaliyar, A. Goswami, and P. Narang, "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach," *Multimedia Tools Appl.*, vol. 80, no. 8, pp. 11765–11788, Mar. 2021.

[93] W. Shishah, "Fake news detection using BERT model with joint learning," *Arabian J. Sci. Eng.*, vol. 46, pp. 1–13, Jun. 2021.

[94] H. Yuan, J. Zheng, Q. Ye, Y. Qian, and Y. Zhang, "Improving fake news detection with domain-adversarial and graph-attention neural network," *Decis. Support Syst.*, vol. 151, Dec. 2021, Art. no. 113633.

[95] A. Giachanou, G. Zhang, and P. Rosso, "Multimodal multi-image fake news detection," in *Proc. IEEE 7th Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2020, pp. 647–654.

[96] S. Girgis, E. Amer, and M. Gadallah, "Deep learning algorithms for detecting fake news in online text," in *Proc. 13th Int. Conf. Comput. Eng. Syst. (ICCES)*, Dec. 2018, pp. 93–97.

[97] H. Reddy, N. Raj, M. Gala, and A. Basava, "Text-mining-based fake news detection using ensemble methods," *Int. J. Autom. Comput.*, vol. 17, pp. 1–12, Apr. 2020.

[98] K. Shu, S. Wang, and H. Liu, "Understanding user profiles on social media for fake news detection," in *Proc. IEEE Conf. Multimedia Inf. Process. Retr. (MIPR)*, Apr. 2018, pp. 430–435.

[99] M. L. Della Vedova, E. Tacchini, S. Moret, G. Ballarin, M. DiPierro, and L. de Alfaro, "Automatic online fake news detection combining content and social signals," in *Proc. 22nd Conf. Open Innov. Assoc. (FRUCT)*, May 2018, pp. 272–279.

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

[101] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein, "A stylometric inquiry into hyperpartisan and fake news," 2017, *arXiv:1702.05638*.

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

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

[104] R. K. Kaliyar, A. Goswami, and P. Narang, "DeepFakE: Improving fake news detection using tensor decomposition-based deep neural network," *J. Supercomput.*, vol. 77, no. 2, pp. 1015–1037, Feb. 2021.

[105] R. K. Kaliyar, A. Goswami, and P. Narang, "EchoFakeD: Improving fake news detection in social media with an efficient deep neural network," *Neural Comput. Appl.*, vol. 33, pp. 1–17, Jan. 2021.

[106] M. Dong, L. Yao, X. Wang, B. Benatallah, Q. Z. Sheng, and H. Huang, "DUAL: A deep unified attention model with latent relation representations for fake news detection," in *Proc. Int. Conf. Web Inf. Syst. Eng.* Switzerland: Springer, 2018, pp. 199–209.

[107] J. Zhang, B. Dong, and P. S. Yu, "FakeDetector: Effective fake news detection with deep diffusive neural network," in *Proc. IEEE 36th Int. Conf. Data Eng. (ICDE)*, Apr. 2020, pp. 1826–1829.

[108] 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.

[109] D. Mangal and D. K. Sharma, "Fake news detection with integration of embedded text cues and image features," in *Proc. 8th Int. Conf. Rel., INFOCOM Technol. Optim., Trends Future Directions (ICRITO)*, Jun. 2020, pp. 68–72.

[110] P. Qi, J. Cao, T. Yang, J. Guo, and J. Li, "Exploiting multi-domain visual information for fake news detection," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2019, pp. 518–527.

[111] K. Shu, X. Zhou, S. Wang, R. Zafarani, and H. Liu, "The role of user profiles for fake news detection," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Aug. 2019, pp. 436–439.

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

[113] J. C. S. Reis, A. Correia, F. Murai, A. Veloso, and F. Benevenuto, "Explainable machine learning for fake news detection," in *Proc. 10th ACM Conf. Web Sci. (WebSci)*, New York, NY, USA, 2019, pp. 17–26, doi: 10.1145/3292522.3326027.

[114] J. Kim, B. Tabibian, A. Oh, B. Schölkopf, and M. Gomez-Rodriguez, "Leveraging the crowd to detect and reduce the spread of fake news and misinformation," in *Proc. 11th ACM Int. Conf. Web Search Data Mining*, Feb. 2018, pp. 324–332.

[115] K. Popat, S. Mukherjee, A. Yates, and G. Weikum, "DeClarE: Debunking fake news and false claims using evidence-aware deep learning," 2018, *arXiv:1809.06416*.

[116] T. Saikh, A. Anand, A. Ekbal, and P. Bhattacharyya, "A novel approach towards fake news detection: Deep learning augmented with textual entailment features," in *Proc. Int. Conf. Appl. Natural Lang. Inf. Syst.* Switzerland: Springer, 2019, pp. 345–358.

[117] L. Wu and H. Liu, "Tracing fake-news footprints: Characterizing social media messages by how they propagate," in *Proc. 11th ACM Int. Conf. Web Search Data Mining*, Feb. 2018, pp. 637–645.

[118] K. Shu, S. Wang, and H. Liu, "Beyond news contents: The role of social context for fake news detection," in *Proc. 12th ACM Int. Conf. Web Search Data Mining*, Jan. 2019, pp. 312–320.

[119] F. Monti, F. Frasca, D. Eynard, D. Mannion, and M. M. Bronstein, "Fake news detection on social media using geometric deep learning," 2019, *arXiv:1902.06673*.

[120] M. Albahar, "A hybrid model for fake news detection: Leveraging news content and user comments in fake news," *IET Inf. Secur.*, vol. 15, no. 2, pp. 169–177, Mar. 2021.

[121] B. Al Asaad and M. Erascu, "A tool for fake news detection," in *Proc. 20th Int. Symp. Symbolic Numeric Algorithms Sci. Comput. (SYNASC)*, Sep. 2018, pp. 379–386.

[122] S. Aphiwongsophon and P. Chongstitvatana, "Detecting fake news with machine learning method," in *Proc. 15th Int. Conf. Electr. Eng., Electron., Comput., Telecommun. Inf. Technol. (ECTI-CON)*, Jul. 2018, pp. 528–531.

[123] N. Ruchansky, S. Seo, and Y. Liu, "CSI: A hybrid deep model for fake news detection," in *Proc. ACM Conf. Inf. Knowl. Manage.*, New York, NY, USA, Nov. 2017, pp. 797–806, doi: 10.1145/3132847.3132877.

[124] Y. Yang, L. Zheng, J. Zhang, Q. Cui, Z. Li, and P. S. Yu, "TI-CNN: Convolutional neural networks for fake news detection," *CoRR*, vol. abs/1806.00749, pp. 1–11, Jun. 2018.

[125] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[126] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapé, "Mobile encrypted traffic classification using deep learning: Experimental evaluation, lessons learned, and challenges," *IEEE Trans. Netw. Service Manag.*, vol. 16, no. 2, pp. 445–458, Feb. 2019.

[127] P. Yildirim and D. Birant, "The relative performance of deep learning and ensemble learning for textile object classification," in *Proc. 3rd Int. Conf. Comput. Sci. Eng. (UBMK)*, Sep. 2018, pp. 22–26.

[128] D. Shen, G. Wu, and H. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jun. 2017.

[129] M. Veres and M. Moussa, "Deep learning for intelligent transportation systems: A survey of emerging trends," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 8, pp. 3152–3168, Aug. 2020.

[130] U. Kamath, J. Liu, and J. Whitaker, *Deep Learning for NLP and Speech Recognition*, vol. 84. Switzerland: Springer, 2019.

[131] B. M. Amine, A. Drif, and S. Giordano, "Merging deep learning model for fake news detection," in *Proc. Int. Conf. Adv. Electr. Eng. (ICAEE)*, Nov. 2019, pp. 1–4.

[132] Q. Li, Q. Hu, Y. Lu, Y. Yang, and J. Cheng, "Multi-level word features based on CNN for fake news detection in cultural communication," *Pers. Ubiquitous Comput.*, vol. 24, no. 2, pp. 1–14, 2019.

[133] O. Ajao, D. Bhowmik, and S. Zargari, "Fake news identification on Twitter with hybrid CNN and RNN models," in *Proc. 9th Int. Conf. Social Media Soc.*, New York, NY, USA, Jul. 2018, pp. 226–230, doi: 10.1145/3217804.3217917.

[134] L. Li, G. Cai, and N. Chen, "A rumor events detection method based on deep bidirectional GRU neural network," in *Proc. IEEE 3rd Int. Conf. Image, Vis. Comput.*, Jun. 2018, pp. 755–759.

[135] M. Z. Asghar, A. Habib, A. Habib, A. Khan, R. Ali, and A. Khattak, "Exploring deep neural networks for rumor detection," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 4, pp. 1–19, 2019.

[136] S. R. Sahoo and B. B. Gupta, "Multiple features based approach for automatic fake news detection on social networks using deep learning," *Appl. Soft Comput.*, vol. 100, Mar. 2021, Art. no. 106983.

[137] Q. Liao, H. Chai, H. Han, X. Zhang, X. Wang, W. Xia, and Y. Ding, "An integrated multi-task model for fake news detection," *IEEE Trans. Knowl. Data Eng.*, early access, Jan. 28, 2021, doi: 10.1109/TKDE.2021.3054993.

[138] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, Jan. 2008.

[139] T. Bian, X. Xiao, T. Xu, P. Zhao, W. Huang, Y. Rong, and A. Huang, "Rumor detection on social media with bi-directional graph convolutional networks," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 1, pp. 549–556.

[140] Q. Huang, C. Zhou, J. Wu, M. Wang, and B. Wang, "Deep structure learning for rumor detection on Twitter," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 1–8.

[141] Y. Rong, W. Huang, T. Xu, and J. Huang, "DropEdge: Towards deep graph convolutional networks on node classification," 2019, *arXiv:1907.10903*.

[142] Y. Ren, B. Wang, J. Zhang, and Y. Chang, "Adversarial active learning based heterogeneous graph neural network for fake news detection," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2020, pp. 452–461.

[143] Z. Wu, D. Pi, J. Chen, M. Xie, and J. Cao, "Rumor detection based on propagation graph neural network with attention mechanism," *Expert Syst. Appl.*, vol. 158, Nov. 2020, Art. no. 113595. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S095741742030419X

[144] L. Zhang, J. Li, B. Zhou, and Y. Jia, "Rumor detection based on SAGNN: Simplified aggregation graph neural networks," *Mach. Learn. Knowl. Extraction*, vol. 3, no. 1, pp. 84–94, Jan. 2021. [Online]. Available: https://www.mdpi.com/2504-4990/3/1/5

[145] S. Hiriyannaiah, A. Srinivas, G. K. Shetty, G. Siddesh, and K. Srinivasa, "A computationally intelligent agent for detecting fake news using generative adversarial networks," in *Hybrid Computational Intelligence: Challenges and Applications*. Amsterdam, The Netherlands: Elsevier, 2020, p. 69.

[146] J. Wang, L. Yu, W. Zhang, Y. Gong, Y. Xu, B. Wang, P. Zhang, and D. Zhang, "IRGAN: A minimax game for unifying generative and discriminative information retrieval models," in *Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Aug. 2017, pp. 515–524.

[147] Y. Li and J. Ye, "Learning adversarial networks for semi-supervised text classification via policy gradient," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2018, pp. 1715–1723.

[148] B. Hu, Y. Fang, and C. Shi, "Adversarial learning on heterogeneous information networks," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 120–129.

[149] T. Le, S. Wang, and D. Lee, "MALCOM: Generating malicious comments to attack neural fake news detection models," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2020, pp. 282–291.

[150] Y. Long, Q. Lu, R. Xiang, M. Li, and C.-R. Huang, "Fake news detection through multi-perspective speaker profiles," in *Proc. 8th Int. Joint Conf. Natural Lang. Process.*, vol. 2. Taipei, Taiwan: Asian Fed. Natural Lang. Process., Nov. 2017, pp. 252–256. [Online]. Available: https://aclanthology.org/I17-2043/

[151] T. Chen, X. Li, H. Yin, and J. Zhang, "Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection," in *Proc. Pacific–Asia Conf. Knowl. Discovery Data Mining*. Switzerland: Springer, 2018, pp. 40–52.

[152] N. Aloshban, "ACT: Automatic fake news classification through self-attention," in *Proc. 12th ACM Conf. Web Sci.*, Jul. 2020, pp. 115–124.
[153] Y.-J. Lu and C.-T. Li, "GCAN: Graph-aware co-attention networks for explainable fake news detection on social media," 2020, *arXiv:2004.11648*.
[154] J. Ding, Y. Hu, and H. Chang, "BERT-based mental model, a better fake news detector," in *Proc. 6th Int. Conf. Comput. Artif. Intell.*, New York, NY, USA, Apr. 2020, pp. 396–400, doi: 10.1145/3404555.3404607.
[155] L. Wu, Y. Rao, H. Yu, Y. Wang, and A. Nazir, "False information detection on social media via a hybrid deep model," in *Proc. Int. Conf. Social Inform.*, Sep. 2018, pp. 323–333, doi: 10.1007/978-3-030-01159-8_31.
[156] A. Choudhary and A. Arora, "Linguistic feature based learning model for fake news detection and classification," *Expert Syst. Appl.*, vol. 169, May 2021, Art. no. 114171.
[157] D. K. Vishwakarma, D. Varshney, and A. Yadav, "Detection and veracity analysis of fake news via scrapping and authenticating the web search," *Cognit. Syst. Res.*, vol. 58, pp. 217–229, Dec. 2019.
[158] Z. Jin, J. Cao, Y. Zhang, and J. Luo, "News verification by exploiting conflicting social viewpoints in microblogs," in *Proc. 13th AAAI Conf. Artif. Intell. (AAAI)*, 2016, pp. 2972–2978.
[159] X. Zhou and R. Zafarani, "Fake news detection: An interdisciplinary research," in *Proc. Companion World Wide Web Conf.*, May 2019, p. 1292.
[160] R. Kumari and A. Ekbal, "AMFB: Attention based multimodal factorized bilinear pooling for multimodal fake news detection," *Expert Syst. Appl.*, vol. 184, Dec. 2021, Art. no. 115412.
[161] A. Nascita, A. Montieri, G. Aceto, D. Ciuonzo, V. Persico, and A. Pescape, "XAI meets mobile traffic classification: Understanding and improving multimodal deep learning architectures," *IEEE Trans. Netw. Service Manage.*, early access, Jul. 19, 2021, doi: 10.1109/TNSM.2021.3098157.
[162] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

M. F. MRIDHA (Senior Member, IEEE) received the Ph.D. degree in AI/ML from Jahangirnagar University, in 2017. He joined as a Lecturer at the Department of Computer Science and Engineering, Stamford University Bangladesh, in June 2007. He was promoted as a Senior Lecturer at the Department of Computer Science and Engineering, in October 2010, and promoted as an Assistant Professor at the Department of Computer Science and Engineering, in October 2011. Then, he joined as an Assistant Professor at UAP, in May 2012. He worked as a CSE Department Faculty Member at the University of Asia Pacific and a Graduate Coordinator, from 2012 to 2019. He is currently working as an Associate Professor with the Department of Computer Science and Engineering, Bangladesh University of Business and Technology. His research experience, within both academia and industry, results in over 80 journals and conference publications. For more than ten years, he has been with the masters and undergraduate students as a supervisor of their thesis work. His research interests include artificial intelligence (AI), machine learning, deep learning, natural language processing (NLP), and big data analysis. He has served as a program committee member for several international conferences/workshops. He served as an associate editor for several journals.

ASHFIA JANNAT KEYA was born in Dhaka, Bangladesh. She received the B.Sc. degree in computer science and engineering from the Bangladesh University of Business and Technology (BUBT), in 2021. She is currently working as a Research Assistant with the Department of CSE, BUBT. She also works as a Researcher with the Advanced Machine Learning Lab. Her research interests include deep learning, natural language processing (NLP), and computer vision. She has experienced working in C++, Python, Keras, TensorFlow, Sklearn, NumPy, Pandas, and Matplotlib.

MD. ABDUL HAMID was born in Sonatola, Pabna, Bangladesh. He received the Bachelor of Engineering degree in computer and information engineering from the International Islamic University Malaysia (IIUM), in 2001, and the combined master's and Ph.D. degree from the Computer Engineering Department, Kyung Hee University, South Korea, in August 2009, majoring in information communication. His education life spans over different countries in the world. From 1989 to 1995, his high school and college graduation at the Rajshahi Cadet College, Bangladesh. He has been in the teaching profession throughout his life, which also spans over different parts of the globe. From 2002 to 2004, he was a Lecturer with the Computer Science and Engineering Department, Asian University of Bangladesh, Dhaka, Bangladesh. From 2009 to 2012, he was an Assistant Professor with the Department of Information and Communications Engineering, Hankuk University of Foreign Studies (HUFS), South Korea. From 2012 to 2013, he was an Assistant Professor with the Department of Computer Science and Engineering, Green University of Bangladesh. From 2013 to 2016, he was an Assistant Professor with the Department of Computer Engineering, Taibah University, Madinah, Saudi Arabia. From 2016 to 2017, he was an Associate Professor with the Department of Computer Science, Faculty of Science and Information Technology, American International University-Bangladesh, Dhaka. From 2017 to 2019, he was an Associate Professor and a Professor with the Department of Computer Science and Engineering, University of Asia Pacific, Dhaka. Since 2019, he has been a Professor with the Department of Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia. His research interests include network/cyber-security, natural language processing, machine learning, wireless communications, and networking protocols.

MUHAMMAD MOSTAFA MONOWAR received the B.Sc. degree in computer science and information technology from the Islamic University of Technology (IUT), Bangladesh, in 2003, and the Ph.D. degree in computer engineering from Kyung Hee University, South Korea, in 2011. He worked as a Faculty Member at the Department of Computer Science and Engineering, University of Chittagong, Bangladesh. He is currently working as an Associate Professor at the Department of Information Technology, King Abdulaziz University, Saudi Arabia. His research interests include wireless networks, mostly *ad-hoc*, sensor, and mesh networks, including routing protocols, MAC mechanisms, IP and transport layer issues, cross-layer design, and QoS provisioning, security and privacy issues, and natural language processing. He has served as a program committee member for several international conferences/workshops. He served as an editor for a couple of books published by CRC Press and Taylor & Francis Group. He also served as a guest editor for several journals.

MD. SAIFUR RAHMAN is currently working as an Assistant Professor at the Department of Computer Science and Engineering, Bangladesh University of Business and Technology. He has expertise in software development and has developed numerous management systems. He has been a successful Director of the International Collegiate Programming Contest (ICPC), Dhaka Regional Contest, in 2014. Apart from the collaboration and development domain, his skills cover theoretical background in computer engineering sectors. His research interests include system design and artificial intelligence-based systems. He received coach awards in ICPC Dhaka Regional Contests.

Characterization, Classification and Detection of Fake News in Online Social Media Networks

Xavier Jose*, S.D Madhu Kumar† and Priya Chandran‡
Department of Computer Science and Engineering
National Institute of Technology Calicut
Calicut, India
Email: *xavier.jose55@gmail.com, †madhu@nitc.ac.in, ‡priya@nitc.ac.in

2021 IEEE Mysore Sub Section International Conference (MysuruCon) | 978-1-6654-3888-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/MYSURUCON52639.2021.9641517

***Abstract*—Due to its increasing popularity, low cost, and easy-to-access nature, Online Social Media (OSM) networks have evolved as a powerful platform for people to access, consume, and share news. However, this has led to the large-scale distribution of fake news, i.e., deliberate, false, or misleading information. Fake news is a pressing dilemma, as it has serious negative implications for individual users and for society as a whole. The news contents in the OSM networks are distributed rapidly, so the identification systems should predict news items as soon as possible to avoid spreading false news. Therefore, it is extremely crucial and technically challenging to detect fake news in social media networks. In this paper, we have discussed different characteristics and types of fake news and also propose an effective solution to detect fake news in OSM networks. The stance detection model and the fabricated content classifier are the main two components of the solution. The stance detection model achieved an accuracy of 90.37% with Logistic Regression, and the fabricated content classifier achieved an accuracy of 93.46% with Bi-directional LSTM.**

***Index Terms*—Online fake news, Social media analytics, Natural Language Processing, Machine learning, Fake news detection.**

I. Introduction

In all sectors, including journalism, Online Social Media (OSM) networks have emerged as an essential medium of communication. As social media platforms are becoming more popular, more users prefer to search for news in social media instead of conventional news sources. Although consuming news on social media offers many benefits, the absence of control and convenient access has led to the wide dissemination of fake news or misinformation in OSM networks. Fake news in online social media networks poses many challenges in multiple dimensions. The widespread distribution of fake news can have a significant adverse effect on users and society as a whole. Propagandists usually use fake news to transmit or propagate misinformation for economic or political benefits. Also, malicious user accounts like social bots can be used to publish and propagate fake news in online social media networks.

Fake news propagates quickly in online social media networks and reaches a wider audience in very little time. Considering the dynamic nature of social media networks, there is a severe demand for effective automatic real-time fake news detection mechanisms. Identifying relevant features that can distinguish fake news from real news makes fake news detection a challenging problem. Fake news identification on traditional news media depends predominantly on the features extracted from the news content only. But in the case of fake news in OSM networks, social context features can also be utilized along with news content features in recognizing fake news. The features extracted from the user's social engagements in news consumption and sharing on social media networks are termed as social context features [1].

The rest of this paper is organized as follows. Section II is a review of related works, and Section III presents the classification of fake news in OSM networks. The design of the proposed solution is described in Section IV, and Section V explains the implementation. Results and discussion are presented in Section VI, and Section VII makes concluding remarks and discusses future work.

II. Related Works

Over the past few years, fake news detection on social media has generated a lot of attention among researchers, and several attempts were made to address the problem. Most approaches proposed for fake news detection have been based on text classification using machine learning models. Various machine learning models like Logistic regression, Naive Bayes, Random forests, Decision Trees and Support Vector Machines have been built and trained to classify a particular news item as reliable or not [2]. Among the supervised machine learning algorithms, SVMs are one of the most widely used methods for classification in many research works concerning fake news detection. As per the experimental results of [3], SVMs have shown better accuracy than other supervised machine learning approaches for fake news prediction. Classification based on deep learning techniques are also found to be very useful in identifying fake news [4].

The research work of [5] shows that there are considerable differences between fake news and real news, particularly in the title of the news articles. The authors find that the length of fake news titles is more compared to real news titles and fake news titles used fewer nouns overall but more proper nouns and fewer stopwords. Also, the length of body text of fake news articles is less compared to real news articles and uses simple words, more occasional quotes, fewer punctuations, etc. They tried various machine learning algorithms, and the best performing model could achieve an accuracy of 91%. They

978-0-7381-4662-1/21/$31.00 ©2021 IEEE

conclude that real news articles persuade users through sound arguments while fake news convinces users through heuristics.

The authors in [6] try to come up with a solution that can identify sites containing misinformation based on *clickbaits*. They suggested that most fake news contains what is called *clickbaits*. *Clickbait* is an article, mostly sensational or having a catchy headline, targeted at clicking. The authors in [7] propose a prototype for identifying fake news items from twitter posts by using machine learning classifiers trained on datasets like the FNC-I [8]. The solution compares the main tweet object with its child objects if they agree on the topic or not. This might help to detect a fake tweet if one of the child objects discusses a different topic. On comparing different learning algorithms, SVM and Naive Bayes classifier gave better accuracy.

An approach based on n-gram analysis and machine learning techniques has been implemented in [9]. The paper presented a machine learning model for fake news using n-gram analysis with the help of different feature extraction techniques. Six different machine learning techniques were compared, employing two different feature extraction techniques. The best performing model was found to be the linear SVM classifier using unigram features.

The authors in [10] proposes a novel hybrid deep learning model that combines convolutional and recurrent neural networks for fake news classification. The classification accuracy of the model was significantly better than other non-hybrid baseline methods. A similarity-aware multi-modal method is proposed to predict fake news in [11]. In this method both textual and visual features of news content are extracted and investigated. The experimental results show that both multi-modal features and cross-modal similarity are important in fake news detection.

A novel automatic fake news detection model based on geometric deep learning is proposed in [12]. The model was trained and tested on news articles from Twitter. Experimental results show that social network structure and propagation are key features enabling highly accurate fake news detection. The results imply that propagation-based approaches can be used as an alternative to content-based approaches for fake news detection.

The following section presents the classification of fake news in online social media networks.

III. Classification of Fake News in OSM Networks

In general, "fake news refers to all kinds of false stories, rumors or news that are mainly published and distributed on the Internet, in order to mislead, befool or lure readers for financial, political or other gains" [13]. First Draft News project has identified seven types of fake news in its preliminary findings [14]. Following are the different types of fake news in OSM networks:

1) ***Satire/Parody****:* These are entertainment-oriented articles presented in conventional news media format mostly created without any intention to spread misinformation. But these articles can mislead readers when shared out of context.

2) ***Image/Video Manipulation****:* There can be mainly two types of image/video manipulation in news articles. One uses real images/videos that are not associated with the news content and tries to create a false association to mislead users purposefully. Other uses edited images/videos inside the news article to establish the claims made by the news.

3) ***Imposter content****:* Impersonation of legitimate/credible news sources, for example, by using an established news agency's branding.

4) ***Sponsored content****:* These are news article that claims to be unbiased media content when, in fact, it is public relations or advertising campaigns.

5) ***False connection****:* News articles in which the headline/title doesn't support/agree with the news content. This includes clickbait which contains a catchy/sensational headline but the content will be unrelated.

6) ***Misleading content****:* Selective reporting of real information by a news article to develop an issue or create a false narrative is called misleading content. For example, reporting only partially chosen real facts related to an incident to frame issues.

7) ***False context****:* These are news articles in which real facts are shared with incorrect background information.

8) ***Fabricated content****:* These are news stories that are completely made up but appear as legitimate news articles. It tries to mimic legitimate news articles but differs from legitimate journalism in writing styles/patterns, structural and other ways. These are intentionally produced for some benefits and not to report facts, so they frequently contain opinionated and biased language.

The design of the proposed solution is described in the following section.

IV. Design of Solution

From the types of fake news listed in Section III, our solution focuses on detecting false connection and fabricated content. The input to the system is the tweet containing the news article. The solution consists of two main phases. In the first phase, relevant words, phrases, and sentences are extracted from the news article using NLP techniques and query them in reputable/credible sources using Google News API. A credibility score is assigned based on whether the news content in the tweet agrees, disagrees, or unrelated to the news content from credible/reputable sources. In the second phase, the tweet is given to the machine learning models for classification. Figure 1 shows the design of the solution.

- *Input:* URL/Tweet Id of the tweet containing the news article
- *Output:* There will be four output parameters
  - List of news articles from credible sources related to the news content
  - Credibility score (Between 0 and 1)
  - False connection : Yes or No
  - Fabricated content : Yes or No

The steps involved are :

1) Fetching the tweet and social context information associated with the given tweet using Twitter Search API.
2) Using the links present in the tweet, extract the content from the pages using web scraping techniques, and append it to the news text.
3) The news text is converted to lowercase and tokenized into n-grams using the NLTK module. Stop-words are removed from the list of n-grams.
4) The remaining n-grams are reconstructed into the original strings and query them in reputable/credible sources using Google News API. This will return a list of news articles from credible sources related to our news content.
5) For each news content in the list, check whether the given news content in the tweet agrees, disagrees, or is unrelated to it using the stance detection model. Based on this a credibility score is assigned to the tweet.

   Let a be the number of news articles from credible sources that agree and d be the number of news articles from credible sources that disagree, then credibility score,

$$C = \begin{cases} 0 \text{ if } a + d = 0 \\ \frac{a}{a+d} \text{ otherwise} \end{cases}$$

6) Pass the headline/title and content of the news article to the trained stance detection model exposed as a REST web service. This will predict whether given headline-content pairs agree, disagree, discuss the same topic, or are not related at all.
   a) If the headline-content pairs disagree or are not related at all, then it is classified as false connection.
   b) Not false connection otherwise.
7) Classify the tweet containing the news article as fabricated content or not using the trained machine learning fabricated content classifier exposed as a REST web service. This is a binary classifier that outputs yes or no.

A. *Stance Detection Model*

Stance Detection is a natural language processing problem to detect the text's stance towards another target text. Stance detection is used in numerous applications such as textual entailment, opinion summarization, emotion detection, etc. Our stance detection model classifies body text stance towards a target into one of the four stance categories.

- *Input:* The news headline and the news body text (news content).
- *Output:* The stance of the news body relative to the claim asserted in the news headline:
  - *Agree:* The news body agrees with the news headline.
  - *Disagree:* The news body disagrees with the news headline.

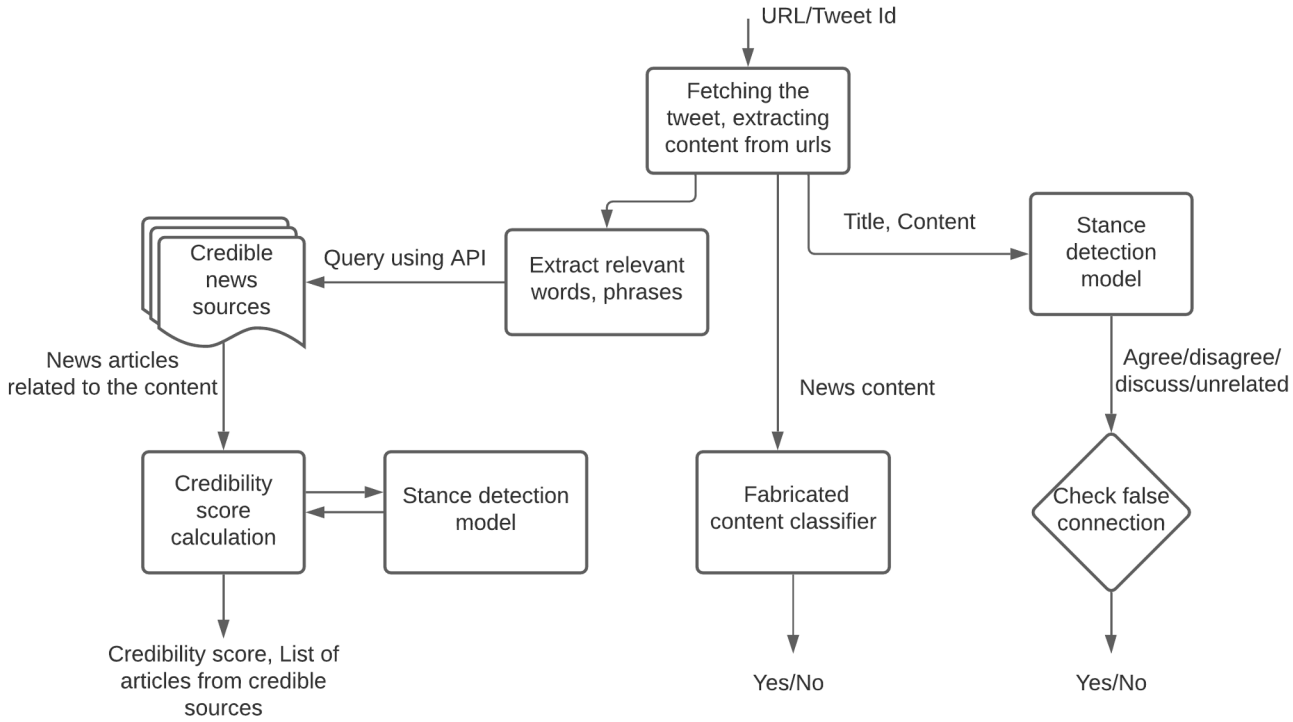


Fig. 1: Design of the solution

- *Discusses:* The news body discuss the same topic as the news headline.
- *Unrelated:* The news body text is not related to the news headline.

B. Fabricated Content Classifier

Fabricated content includes news stories that are completely made up but appear as legitimate news articles. It tries to mimic legitimate news articles but differs from legitimate journalism in writing styles/patterns, structural and other ways. As these are purposely produced for some benefits and not to report facts, they frequently contain biased or offensive language. Thus, to detect fabricated content, it is sensible to utilize linguistic features that extract various writing patterns and sensational captions.

- *Input:* The news text.
- *Output:* Classify the text of the news into one of the two classes:
  - *1:* The news text is fabricated content.
  - *0:* The news text is not fabricated content.

Following section explains the implementation of stance detection model and fabricated content classifier.

V. Implementation

A. Stance Detection Model

1) ***Dataset****:* The Fake News Challenge dataset [15], [8] have been used to train the model. The dataset contains 50,000 stance tuples. Every data item includes the news headline, body text, and the annotated stance that tells whether the headline's claim agrees, disagrees, discusses, or is unrelated to the body text. We divided the dataset into the train (90%), and test (10%) sets for constructing the model.

2) ***Preprocessing****:* The input text is first converted to lower case and removed punctuations, stop words, and non-alphanumeric characters. Natural Language Toolkit (NLTK) library of python is used for these preprocessing tasks.

3) ***Feature Extraction****:* We used the scikit-learn function *TfidfVectorizer* for converting the text data to a matrix of TF-IDF features.

4) ***Training****:*

a) Logistic Regression: The Logistic Regression model is trained with the training data set. The scikit-learn implementation of Logistic Regression is used with parameter *class_weight='balanced'*.

b) Decision Tree Classifier: The scikit-learn implementation of decision tree is used with parameters *criterion='gini'* and *min_samples_split=2*. The model is trained with the training data set.

c) Random Forest Classifier: The scikit-learn implementation of random forest is used with parameters *n_estimators=100*, *criterion='gini'* and *min_samples_split=2*. The model is trained with the training data set.

d) Multinomial Naive Bayes Classifier: The scikit-learn implementation of multinomial Naive Bayes classifier is used with parameters *alpha=0.3* and *fit_prior='True'*. The model is trained with the training data set.

e) Support Vector Machine: The scikit-learn implementation of SVM is used with parameter *kernel='linear'*. The model is trained with the training data set.

B. Fabricated Content Classifier

1) ***Dataset****:* The model is trained using the Fake News dataset [16]. The Fake News dataset is a collection of fake news and truthful articles, collected from different legitimate news sites and sites flagged as unreliable by fact checking websites. These fake news articles include news contents that are completely made up but appear as legitimate news articles. We used the *text* (text of the news article) field in the CSV file for training the model.

2) ***Preprocessing****:* The input text is first converted to lower case and removed punctuations and non-alphanumeric characters. Words are converted to their base form using *WordNetLemmatizer*. Natural Language Toolkit (NLTK) library of python is used for these preprocessing tasks.

3) ***Feature Extraction****:* We map each news text into a high-dimensional vector space using the technique called word embedding. Each word in the corpus is mapped to a real-valued vector in an n-dimensional vector space. *Keras Embedding layer* provides a suitable way to convert words into word embeddings. One-hot encoding has to be done before embedding text data. *Keras* provides the *one_hot()* function that creates efficient integer encoding of each word in the document. The vocabulary size used is 30000. We are using word embeddings as the input to the LSTM layer. It aims to map semantic meaning into a real vector domain and is an improvement over traditional approaches such as *bag-of-word* encoding schemes. We used the *Keras Embedding layer* as the first layer of the network with *input_dim=vocabulary size, output_dim=100,* and *input_length=1000*.

4) ***Training****:*

a) Long Short-Term Memory - Recurrent Neural Network:

LSTMs are very effective solution for sequential input like text input. We used the *Tensorflow Keras* implementation of the LSTM RNN for building the model. We split the dataset into the train (90%) and test (10%) sets.

The first layer is the *Keras Embedding layer* that uses 100-dimensional vectors to encode each word into the real vector domain. The second layer is the LSTM layer with 100 units. The next layer is the dense output layer with one neuron. Since this is a binary classification, we used the *sigmoid activation function* in the output layer to make 0 or 1 predictions for the two classes. For binary classification, *binary_crossentropy* is used as loss function along with *ADAM optimization algorithm*. Figure 2 shows the model architecture plot. Training is done iteratively for ten epochs with batch_size=64 to minimize loss function and to improve accuracy.

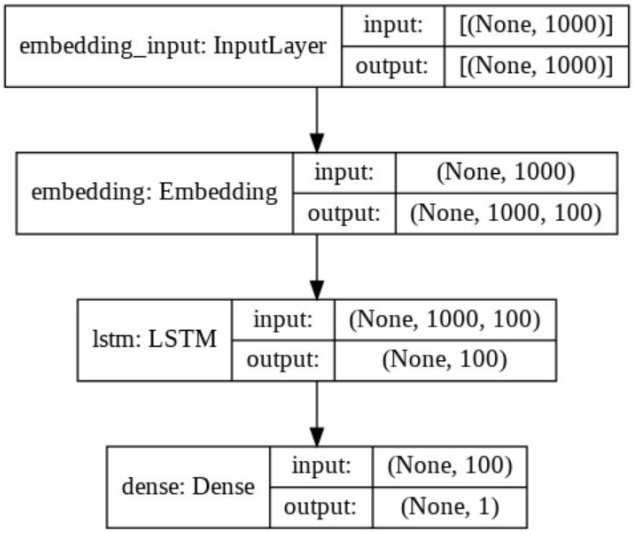


Fig. 2: LSTM Model Architecture Plot

b) Bi-directional Long Short-Term Memory - Recurrent Neural Network:

The first layer is the *Keras Embedding layer* that uses 100-dimensional vectors to encode each word into the real vector domain.

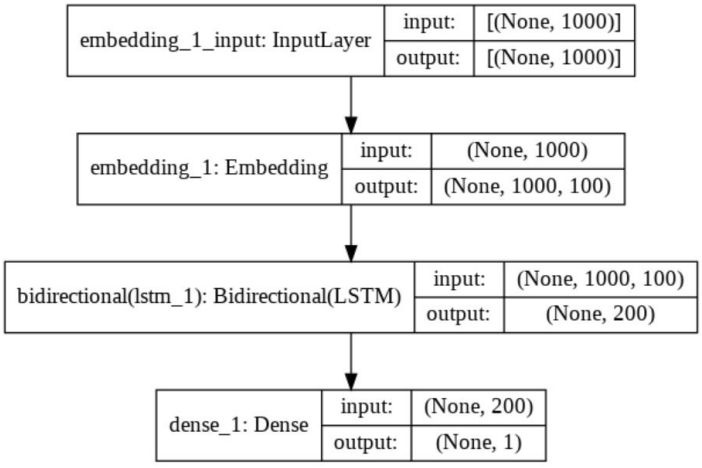


Fig. 3: Bi-directional LSTM Model Architecture Plot

The second layer is obtained by wrapping the LSTM layer (100 smart neurons) with a Bi-directional layer. The next layer is the dense output layer with one neuron. Since this is a binary classification, we used the *sigmoid activation function* in the output layer to make 0 or 1 predictions. For binary classification, *binary_crossentropy* is used as loss function along with the *ADAM optimization algorithm*. Figure 3 shows the model architecture plot. Training is done iteratively for 20 epochs with batch_size=64.

VI. Results and Discussion

A. *Stance Detection Model*

As discussed in Section V-A, we trained models with five different machine learning algorithms using the Fake News Challenge dataset [15], [8]. The testing is done using the test dataset. Below figures shows the confusion matrix of Logistic Regression, Decision Tree, Random Forest, Multinomial Naive Bayes and SVM Classifier.

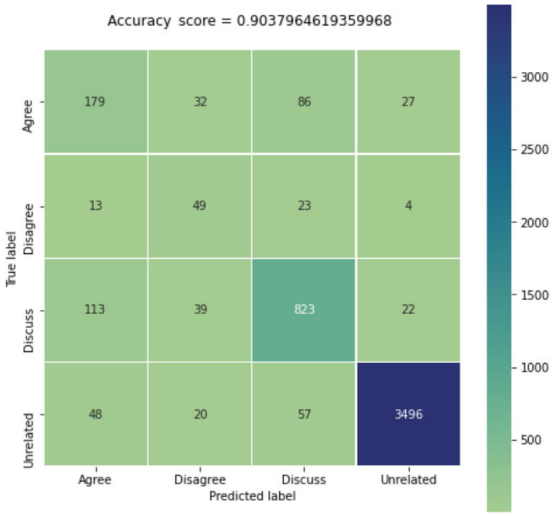


Fig. 4: Confusion Matrix of Logistic Regression

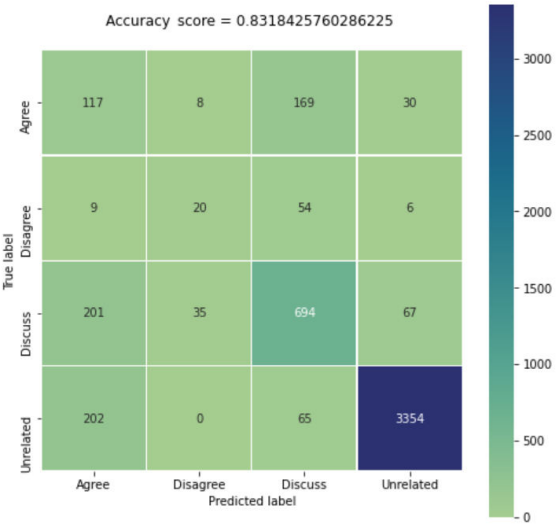


Fig. 5: Confusion Matrix of Decision Tree Classifier

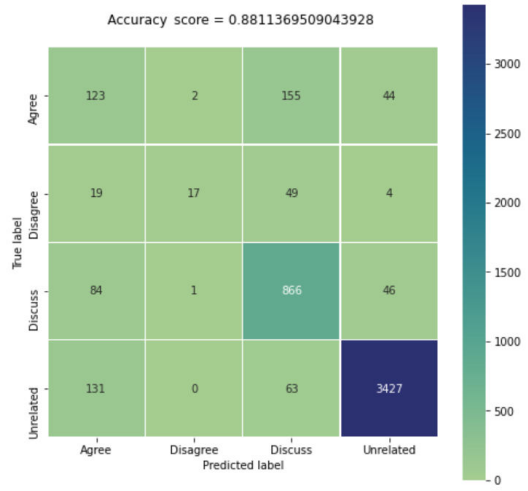


Fig. 6: Confusion Matrix of Random Forest Classifier

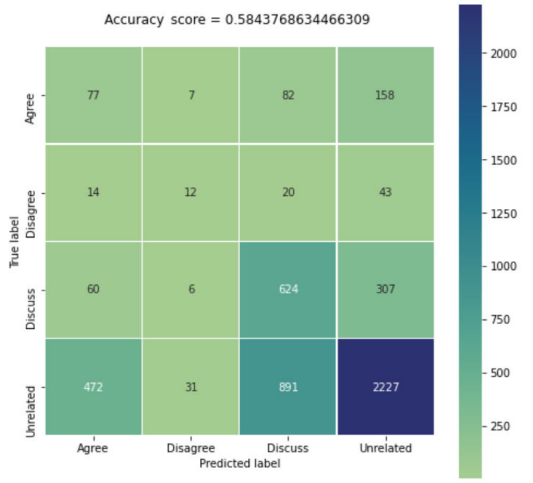


Fig. 7: Confusion Matrix of Multinomial Naive Bayes Classifier

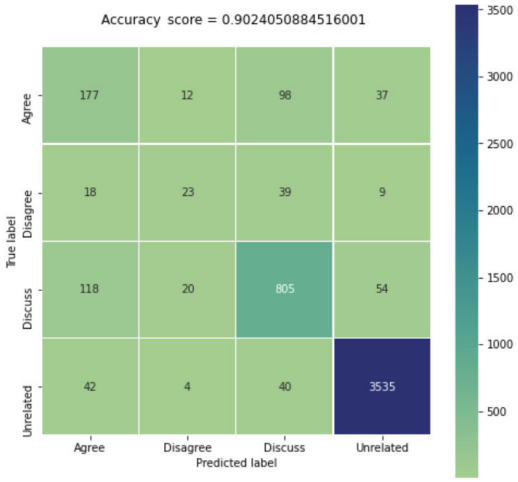


Fig. 8: Confusion Matrix of SVM Classifier

Out of the algorithms tried for stance detection, logistic regression and Support Vector Machine(SVM) show the best performance with accuracy above 90%. Random forest algorithm also performs well, with an accuracy of around 88%. Table I shows the comparison of the algorithms.

Algorithm	Accuracy
Logistic Regression	0.903
Decision Tree	0.831
Random Forest	0.881
Multinomial Naive Bayes	0.584
SVM	0.902

Table I. Comparison of Performance of Different Algorithms

B. Fabricated Content Classifier

As discussed in Section V-B, Long Short-Term Memory Recurrent Neural Network and Bi-directional Long Short-Term Memory Recurrent Neural Network are trained using Fake News dataset.

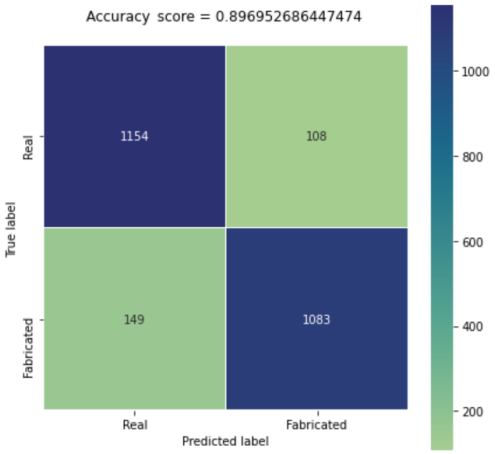


Fig. 9: Confusion Matrix of LSTM

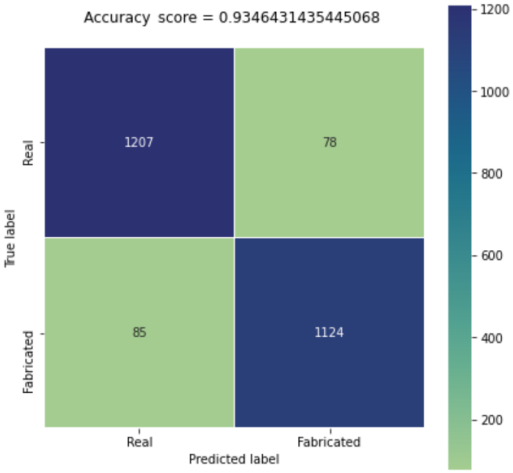


Fig. 10: Confusion Matrix of Bi-directional LSTM

Figure 9 shows the confusion matrix of LSTM. Figure 10 shows the confusion matrix of Bi-directional LSTM. Bi-directional LSTM shows the best performance with an accuracy of 93.46%. Table II shows the comparison of the models.

Model	Accuracy
LSTM RNN	0.896
Bi-directional LSTM RNN	0.934

Table II. Comparison of Performance of Two Models

VII. Conclusion and Future Work

Over the past few years, the use of social media to propagate fake news has increased tremendously. This has numerous negative consequences, and hence there is an urgent need for powerful automatic fake news identification mechanisms. This paper has discussed different characteristics and types of fake news in OSM networks. Also, we have proposed an effective solution to detect fake news, particularly for false connection and fabricated content identification in OSM networks. The stance detection model and the fabricated content classifier are the main two components of the solution. We tried different machine learning algorithms for implementing both models, and the best-performing ones are chosen for building the solution. The stance detection model achieved an accuracy of 90.37% with Logistic Regression, and the fabricated content classifier achieved an accuracy of 93.46% with Bi-directional LSTM.

The proposed solution focuses only on detecting false connection and fabricated content out of the eight types of fake news found in OSM networks. However, as future work, the solution can be improved by developing more machine learning models or techniques for detecting other types of fake news. It might be technically challenging to detect some types while others, like image/video manipulation, can be accurately identified. Also, we can extend the solution to support multiple languages. Here the major challenge is preparing the datasets and building the models as there are no publicly available fake news datasets for most regional languages.

References

[1] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 19(1):22–36, 2017.

[2] Alessandro Bondielli and Francesco Marcelloni. A survey on fake news and rumour detection techniques. *Information Sciences*, 497:38–55, 2019.

[3] Hu Zhang, Zhuohua Fan, Jiaheng Zheng, and Quanming Liu. An improving deception detection method in computer-mediated communication. *Journal of Networks*, 7(11):1811–1816, 2012.

[4] Natali Ruchansky, Sungyong Seo, and Yan Liu. Csi: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, Singapore*, pages 797–806, 2017.

[5] Benjamin Horne and Sibel Adali. This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In *Proceedings of the International AAAI Conference on Web and Social Media, Montreal, Canada*, volume 11, 2017.

[6] Monther Aldwairi and Ali Alwahedi. Detecting fake news in social media networks. *Procedia Computer Science*, 141:215–222, 01 2018.

[7] Ehesas Mia Mahir, Saima Akhter, Mohammad Rezwanul Huq, et al. Detecting fake news using machine learning and deep learning algorithms. In *2019 7th International Conference on Smart Computing & Communications (ICSCC), Sarawak, Malaysia*, pages 1–5. IEEE, 2019.

[8] Fake News Challenge Stage 1 (FNC-I): Stance Detection. http://www.fakenewschallenge.org/. Online; accessed: 2021-06-12.

[9] Hadeer Ahmed, Issa Traore, and Sherif Saad. Detection of online fake news using n-gram analysis and machine learning techniques. In *International conference on intelligent, secure, and dependable systems in distributed and cloud environments, Vancouver, Canada*, pages 127–138. Springer, 2017.

[10] Jamal Abdul Nasir, Osama Subhani Khan, and Iraklis Varlamis. Fake news detection: A hybrid cnn-rnn based deep learning approach. *International Journal of Information Management Data Insights*, 1(1):100007, 2021.

[11] Xinyi Zhou, Jindi Wu, and Reza Zafarani. SAFE : Similarity-aware multi-modal fake news detection. *Advances in Knowledge Discovery and Data Mining*, 12085:354, 2020.

[12] Federico Monti, Fabrizio Frasca, Davide Eynard, Damon Mannion, and Michael M Bronstein. Fake news detection on social media using geometric deep learning. *arXiv preprint arXiv:1902.06673*, 2019.

[13] Xichen Zhang and Ali A Ghorbani. An overview of online fake news: Characterization, detection, and discussion. *Information Processing & Management*, 57(2):102025, 2020.

[14] Fake news. It's complicated. https://firstdraftnews.org/articles/fake-news-complicated/. Online; accessed: 2021-07-25.

[15] FakeNewsChallenge/fnc-1. https://github.com/FakeNewsChallenge/fnc-1. Online; accessed: 2021-06-12.

[16] Fake News Dataset. https://www.kaggle.com/c/fake-news/data. Online; accessed: 2021-06-12.