Detecting and Mitigating the Dissemination of Fake News: Challenges and Future Research Opportunities

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

14-02-2022 / 22-02-2022

CITATION

Hakak, Saqib; Shahid, Wajiha; Jamshidi, Bahman; Khan, Wazir Zada; Khan, Muhammad Khurram; Isah, Haruna (2022): Detecting and Mitigating the Dissemination of Fake News: Challenges and Future Research Opportunities. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.19169795.v1>

DOI

[10.36227/techrxiv.19169795.v1](https://doi.org/10.36227/techrxiv.19169795.v1)

Detecting and Mitigating the Dissemination of Fake News: Challenges and Future Research Opportunities

Wajiha Shahid, Bahman Jamshidi, Saqib Hakak, Haruna Isah, Wazir Zada Khan, Muhammad Khurram Khan and Kim-Kwang Raymond Choo

Abstract—Fake news is a major threat to democracy (e.g., influencing public opinion), and its impact cannot be understated particularly in our current socially and digitally connected society. The research community from different disciplines (e.g., computer science, political science, information science, and linguistics) have also studied the dissemination, detection and mitigation of fake news, however it remains challenging to detect and prevent the dissemination of fake news in practice. With AI powered systems, its highly crucial to understand the detector's decision of fake news by means of proper user-friendly explanations when it comes to social media. Hence, in this paper, we systematically survey existing state-of-the-art approaches designed to detect and mitigate the dissemination of fake news, and based on the analysis, we discuss several key challenges and present potential future research agenda specially incorporating AI explainable Fake news credibility system.

Index Terms—AI explainability; Fake news, Deep fakes, Misinformation, Deceptive content, Social bots, Social media, news propaganda, block-chain based detection.

1 INTRODUCTION

Numerous solutions [1; 2; 3; 4] have been proposed to address several security and privacy related issues whether it is related to Internet of things (IoT) [5; 6; 7; 8], user-authentication issues [9; 10; 11], improving road traffic safety[12] or other cyberthreats [13; 14; 15]. However, as individuals and organizations become more connected and are used to getting information (e.g., news) in real-time and from different sources (e.g., user-generated content), there is a new risk that such sources and dissemination approaches (e.g., social media platforms) can be abused via fake news [16]. The research in the domain of fake news is still at its infancy. We broadly define fake news to be content that is created to mislead users [17], for example to misguide, cheat, and defame individuals, groups of individuals, organizations, and/or governments.

Fake news can impact our society in different ways [18; 19; 20], such as those presented in Figure 1. For example, the study conducted by CHEQ, a cyber security company, and the University of Baltimore [21], found that the economic impact of fake news at the global level is

Fig. 1. Potential economical impacts

approximately \$78 billion dollars, with a direct economic loss of around \$39 billion a year in the stock market. Recent high profile real-world events that were affected by fake news include COVID-19 vaccine [22; 23; 24; 25; 26; 27; 28; 29] and the 2016 U.S. Presidential Election [30; 31]. These events reinforce the importance of designing techniques and approaches to detect and mitigate the dissemination of fake news [17; 32; 33].

There are, however, a number of challenges in classifying whether the content is genuine or fabricated as a news article may contain certain truth and certain untruth. For example, an article alleging that a particular candidate is corrupt because he/she has committed activities A, B, and C. However, this particular candidate may indeed be corrupt but because he/she has committed activities B, D, and

- W. Shahid, B. Jamshidi, S. Hakak and H. Isah are with Canadian Institute for Cybersecurity, Faculty of Computer Science University of New Brunswick, Fredericton, Canada. E-mail: wajiha.shahid@unb.ca, saqib.hakak@unb.ca, h.isah@unb.ca.
- W. Z. Khan is with University of Wah, Wah Cantt, 47040, Pakistan, 47040, Pakistan. Email: wazir.zada.khan@uow.edu.pk
- M. K. Khan is with Center of Excellence in Information Assurance, King Saud University, Riyadh 11653, Saudi Arabia
- R. Choo is with Department of Information Systems and Cyber Security, University of Texas at San Antonio, San Antonio, TX 78249-0631, USA. E-mail: raymond.choo@fulbrightmail.org

Fig. 2. Fake news life cycle

E, rather activities A, B, and C. Hence, should we classify this news article as authentic or fake?

An example life cycle of fake news is presented in Figure 2, which explains how fake news can be generated and propagated through various platforms. Motivations for such generation and dissemination vary, ranging from financial (e.g., monetary) to political (e.g., left/right wing) to terrorism (e.g., seeking to create societal panic / unrest) to defamation, and so on. Fake news can also exist in different formats, such as text, image, audio, and video.

There have been extensive attempts to design solutions to detect and mitigate the dissemination of fake news, as well as the existence of several popular fact checking websites (e.g., politifact and TruthorFiction). Building on existing literature review and/or survey articles (see Table 1), we will survey more recently published articles in the literature on this topic. Based on our analysis of the articles, we will present a new taxonomy on fake news detection approaches to guide the categorization of existing approaches. Then, we identify and discuss existing and emerging challenges, as well as potential research agendas.

The remaining paper is organized as follows: In the next section, we will explain our survey methodology and discuss the existing approaches. In the third section, the challenges and issues in detecting fake news are discussed. The fourth section discusses potential future directions, prior to us concluding this paper in the last section.

2 EXISTING FAKE NEWS DETECTION APPROACHES

Due to the extensive volume of literature on this topic, we focused only on SCI-indexed technical journal articles from the year 2019 (e.g., we excluded conference articles, and book chapters). Here, we will briefly summarize the most recent works on fake news detection into seven detection categories as shown in Figure 4. The taxonomy comprises

Fig. 3. Different aspects of Features

four layers. On the first layer, the studies are sorted based on the focus of the research. Each color code represents one detection based research focus and based on that, we have divided the work of researchers based on type of fake news content on layer 2, fake news features on layer 3 and data set categories on layer 4, respectively. Based on our study, the researchers have also put much emphasis on feature identification in detecting fake news.

Features have played the most important role in models specifically in fake news detection as real and fake classes have very similar characteristics. Figure 3 aims to represent various aspects of features studied in this survey. We can look at the features from following points of view: *Being imitated*: Some features are difficult to be mimicked by malicious users (topographical features), while most of them can be imitated easily. *Sensitive to time*: For instance, [37] shows that Word n-gram features may provide information which is less relevant at that point of time. *Required level of computational resources*: Some features are readily available, while some others like layer ratio in [38] need processing

TABLE 1
Contribution of Existing Review Papers

References	Theme of the Survey
[34]	This article covers a total of 8 years of research on automatic fake news detection and provides data sets, user and text based features and main methods for the literature.
[32]	The survey paper covers all the state of art methods in detecting fake news until 2019, characterize negative impacts of fake news, and Lastly, study all the data sets used and propose future directions for researchers.
[35]	The paper gives a survey on fake news detection but also provides a model that is able to classify news that is unreliable into real and fake news after giving it a score and will distinguish between real and fake news based on number of parameters obtained from the URL
[36]	This survey evaluates the methods that can detect fake news from four point of views: the false knowledge it is carrying, the source credibility, its pattern of propagation, and the style of writing.

sources to calculate. *Applicability*: Some features cannot be applied to all situations and platforms. For example, domain reputation related features can not be utilized in social network platforms. *Necessary time to emerge*: As an example, reactions of other users to a news article need a time period to be revealed. *Explainable*: Features achieved by deep learning methods act like a black-box and are not interpretable in contrast to hand-crafted ones. *Pre-processing steps*: Unlike features such as n-grams, some characteristics like user profile based ones do not need much pre-processing stages. *Potential to be transferred*: Pre-trained word embedding can be used and updated in new situations. *Source*: Features can be extracted from different sources such as contextual content, images, videos, profiles, etc. *Required size of corpus*: Features gained from LIWC, NELE, different embedding need large corpus. *language-independent*: Features such as word embedding, word-grams, character-grams are dependent on the language, while profile based are independent. *Domain-independent*: Most of the content-based features and features extracted from images cannot apply to all domains. To illustrate, vectors gained from word embedding need a corpus that contains sufficient data for the underlying domain.

In the following subsections, the summaries of the existing works are briefly highlighted. We have described the datasets used in these studies in Tables 2 to 8 respectively.

2.1 Automatic Detection

All studies proposing models that automatically capture discriminatory features from fake news and classify the news based on the hidden layers of the deep learning model comes under the category of Automatic Detection.

Alatas et. al. [16] has used a two-step method for fake news detection. The first step is converting unstructured data into structured dataset by applying pre-processors. TF weighting method and Document-Term Matrix are applied to present dataset in the form of vectors. After this pre-processing, the author implemented twenty-three supervised artificial intelligence algorithms on the structured dataset by text mining methods, which is the second step of this method. Alatas et al evaluated these methods experimentally based on four evaluation metrics using public

Fig. 4. Taxonomy of fake news detection.

datasets. The specified artificial intelligence algorithms used in this paper are BayesNet, Sequential Minimal Optimization (SMO), Multi-Layer Perceptron JRp, Locally weighted learning, Ordinal Learning Model, OneR, Decision Stump, ZeroR, Randomizable filtered classifier, Stochastic gradient descent, CV Parameter Selection, Logistic model tree, Classification via clustering, Weighted instances handler wrapper, Ridor, Simple Cart, Attribute Selected Classifier, J48,, Bagging, Decision tree, IBK and Kernel Logistic Regression. To evaluate intelligent classification algorithms, three datasets have been used which are the BuzzFeed Political News data set, Random Political News data set, ISOT Fake News data set. Different intelligent algorithms show different performance when applied to these three datasets to determine accuracy, F1 score and recall. A mean of their performance is taken and based on that, Decision Tree algorithm is found to have the best accuracy, precision, and F-measure, ZeroR, CVPS, and WIHW algorithms have achieved best recall. The limitation of this work is that it begins with text mining as the stage one but the problem with text mining is that it deals with high dimensional data which results in the

loss of model accuracy. To attain an improved efficiency, the redundant features must be removed to decrease the size of data. The nonavailability of such a large data set is another limitation of the work. Moreover, the textual data must be transformed into the form suitable for particular methodology which is yet another limitation of the work. For text mining, large data must be shortened by removing redundant and such features that are not related to the task, thus becoming a limitation of the model.

To detect fake news, Kaliyar et. al. [39] used deep convolutional neural network (FNDNet). The proposed model is designed to automatically detect the features of fake news and differentiate them from those of real news. For this purpose, the author has used multiple hidden layers in building the deep neural network to classify fake news based on the extracted features. Each layer of the developed deep Convolutional Neural Network (CNN) extracts several features. Based on experimentally attained results, the performance of Kaliyar's model is compared with baseline models and is found to have a great accuracy of 98.36% on the test data. The model is trained through benchmarked data sets and the results are validated through six performance evaluation metrics, which include accuracy, recall, F1 score, precision, true negative rate, and false-positive rate. The methodology adopted for this model takes word embedding as the first step. Pre-trained word embedding models are used for this purpose. GloVe, which is an unsupervised learning algorithm, is implemented in this method to get word-embedding based vectors. FNDNet elements consist of convolutional layers, max-pooling layer, dense layers, dropout, and Rectified Linear Unit (ReLU) as activation function. Using Kaggle news dataset, accuracy of existing models and proposed models is compared. The highest benchmark result has an accuracy of 93.50% while the accuracy obtained through GloVe-enabled deep convolutional-based approach (FNDNet) is 98.36%. Various classifiers are used to test the accuracy of proposed methodology. Multinomial Naive Bayes, an accuracy of 89.97% was achieved. Similarly, the method was applied to various Machine Learning-based models for classification like KNN, Random Forest and Decision Tree. With the increase in data, the performance decreases in machine-learning based models. Therefore, deep learning techniques like CNN, LSTM were used for fake news detection, but the accuracy was still low. The best results for precision, recall and F1 score was obtained through the proposed model i.e., FNDNet with a precision of 99.40%, recall of 96.88% and a F1 score of 98.12%. The method achieved remarkable results but has a drawback that it uses a huge data set to train model. The nonavailability of such large data sets is a major hindrance in the path of implementation of this model since it requires a substantial number of words and vocabulary to formulate a data set for extracting implication of each word in a more efficient way. Moreover, the study was made using binary class data sets instead of using a hybrid approach. Hybrid approach detects the fake or real news based on the combination of the news' content, its context, and temporal level information and can create more impact when using multi-label data sets. In this way, the use of binary class dataset results in limiting the analysis and generalization of results obtained from a broader perspective.

2.2 Language-Specific Detection

Language specific detection defines all those papers dealing in languages other than or more than just English while incorporating multiple language data sets to test their detection model.

Henrique et al. [40] has proposed a text-feature based, language independent methodology for fake news detection. The generated text is independent of the source platform and language. The proposed methodology is experimented on five datasets in three language groups giving satisfactory results in comparison to the benchmark. The results obtained through this study are compared against benchmark, custom set of features and bag of words and word2Vec. The best results were achieved through a bag-of-words approach when implemented on feature set. But this approach generates large matrices when applied on text. Therefore, DCD algorithm is applied in this scenario by reducing the size of matrix without significant loss in performance metric. Pedro trained the model using Natural Language Processing representation as well through a customized set of features to make it robust enough to extract features from the raw news and make the model less dependent on language. He extracted 14 textual features from news content using raw texts, polyglot, Word2Vec model and Hunspell. After the extraction of features, the model is evaluated on five data sets (based on online available content) from three different languages. Four tests are conducted on each data set namely custom feature, Word2Vec, DCDistance and bag-of-words. Four algorithms are applied in each set i.e., KNN, random forest, Gaussian Naïve Bayes and SVM. Out of these four, best results were obtained by applying Random Forest and SVM algorithms. Random Forest gave the highest F1 score of 95% and attained the highest accuracy of 95%. The techniques used in this paper are Machine Learning techniques and involve manual fact-checking making the process slower and less efficient. A more robust methodology like deep learning techniques can do the task much faster but it requires a large dataset of fake news for training. The nonavailability of sufficiently large fake news data set is a hindrance in the path of robustness.

2.3 Proposed data-set based detection

All researches proposing and contributing majorly towards a new corpus while simultaneously testing them on their proposed detection model comes under the category of proposed corpus based detection.

Neves et al. [41] introduce a novel approach (GAN-fingerprint Removal autoencoder - GANprintR) to spoof facial manipulation detection systems. This approach removes the GAN fingerprints without compromising the image quality. Thus, the machines, like humans, will not be able to distinguish fake images from real ones. GANprintR is trained with face images of real persons instead of synthetic face images with GAN-fingerprint. This strategy is based on the premise that by training with real images GANprintR can learn the main structure of real face images, which can be helpful to enhance existing fakes, subsequently. To carry out this study, three state-of-the-art manipulation detection approaches are used: *XceptionNet*[42], *Steganalysis*[43] and *Local Artifacts*[44]. Also, three different scenarios are

TABLE 2
Qualitative Analysis of Data sets in Automatic Detection Schemes

Study	Data set used	Nature of Data set	Limitation	Contribution
[16]	BuzzFeed Political News data set	Hybrid Political news containing persuasive information	Limited data set	Twenty-three supervised artificial intelligence algorithms adapted for the first time
	ISOT Fake News data set	Hybrid Political news containing real and fake news	Needs transformation of textual data into a more suitable representation for analysis	
	Random Political News data set	Hybrid political news containing real and fake news and misreporting	Redundant and unrelated features must be removed from text mining for better accuracy	
[39]	BuzzFeed Political News data set	Hybrid Political news containing persuasive information	large data set for training	Better and competitive results with text-based data set
	Random Political News data set	Hybrid political news containing real and fake news and misreporting	use of binary data set instead of hybrid	
	ISOT Fake News data set	Hybrid Political news containing real and fake news		

TABLE 3
Qualitative Analysis of Data sets in Language-Specific Detection Techniques

Study	Dataset used	Nature of Dataset	Limitation	Contribution
[40]	FakeBrCorpus	Mix news containing fake news and persuasive information	Nonavailability of large data set	Proposed a text-feature based, language independent methodology for fake news detection
	TwitterBR	Social media news containing Citizen Journalism		
	FakeOrRealNews	Hybrid Political news containing real and fake news		
	FakeNewsDataset	Political news containing fake news		
	TwitterBR LTO	Social media news containing Commentary, Feature Writing and Citizen Journalism		

designed: 1) *controlled scenarios* 2) *in-the-wild scenarios* 3) *GAN-fingerprint removal*. In the pre-Processing step all the background information is removed from the images while the facial regions are kept wherever it is possible. In order to gain unbiased results, only the frontal face images are kept as well. After the previous stage, images of constant size (224×224 pixels) are available as input for the systems. According to the results, for controlled scenarios (when the same samples are utilized for both development and evaluation of the detection models) XceptionNet has excellent manipulation detection accuracies with EER values less than 0.5%, then Steganalysis represents good performance, and finally Local Artifacts shows poor accuracy with an average of 35.5% EER. On the contrary, in in-the-wild scenarios, the obtained results for all manipulation detection models experience a high reduction. Similarly, for the third scenario (applying GANprintR) this decrease of performance is even further which proves the success of the GANprintR in building enhanced versions of the original fake images. This study also analyses how different image transformations such as resolution downsizing, applying a low-pass filter and JPEG image compression affect facial manipulation detection systems. The results prove the vulnerability of fake detection systems even to such simple image transformations. From this study it can be concluded that as long as facial manipulation detection systems show a poor performance, it is unfeasible to detect fake news from real ones by comparing their images.

Contents with biased messages, intentional or unintentional, is defined as propaganda [45]. In [37], a framework is proposed for Propaganda Spotting in Online Urdu Lan-

guage (ProSOUL). Due to the lack of data set and LIWC dictionary in Urdu language, authors developed a labeled data set via translating the English data set of QCRI's propaganda (Qprop) and also an English LIWC dictionary into Urdu language. The main contribution of this paper is analyzing several different feature sets, so following pre-processing steps, uni-gram, bi-gram, and tri-gram forms of words and characters are extracted. They also extract NEws Landscape (NELA) features to analyze news articles from stylistic and psycho-linguistic perspectives. Moreover, they use Word2Vec and BERT-Base embeddings. Following extracting feature sets, Several experiments are considered to evaluate the effectiveness of them, so Logistic Regression (LogReg) classifier is trained and evaluated with n-gram (C-gram and W-gram), NELA, and Word2Vec feature sets, and also CNN model is trained with BERT embeddings. In the first experiment, the performance of different combinations of stylistic features of NELA, C-gram and W-gram is evaluated. As far as Urdu text classification is concerned, accuracy, precision, recall, and F-measure values achieved for W-gram feature is in the range of 0.90 - 0.91 which is significantly higher than the results gained for various combinations of NELA and C-gram features. Unlike the Urdu text, accuracy achieved by the classifier for the English text with using top 10,000 n-grams is 0.65 which is significantly lower. Machine translation simplifies language variations and this may be the reason for poor results in English text. In the next analysis which is related to NELA features, it can be observed that the accuracy for TTR is 0.70 which is noticeably better than

other features performance and indicates the importance of lexical diversity to detect the propaganda content in the Urdu language. With regard to word embedding methods, they find that the performance of Word2Vec is better in the Urdu language with 0.87 accuracy, while in the English text classification, BERT performed better than Word2Vec with 0.95 accuracy. The poor result of BERT in the Urdu text classification is mostly because of the restricted vocabulary of Urdu language present in the multilingual BERT. Furthermore, by evaluating ProSOUL on unseen instances, they realized that the combination of NELA, W-gram and C-gram features represent much better performance with the F-measure value of 0.84 compared to others. Finally in another experiment, the real-world data is classified with ProSOUL and according to the results it classify online news better than the general web content.

To address the issue of fake news detection in languages other than English, Silva et al. [46] formulated a semi automated corpus named "Fake.Br" to overcome this barrier of language. This study answers the undiscussed areas of fake news detection like what set of features must be included for automatic fake news detection and the best classification strategies for this purpose. The author used public and real labelled dataset with news in Portuguese. The data set contains a total of 3600 true news and 3600 deceptive news. The author has used machine learning methods for analysis and experimented it using various classification methods and various sets of features. Linguistic based features include pausality, emotiveness, uncertainty over the number of verbs used in the news, non-immediacy, diversity, average size of sentences, average size of words, number of spelling errors. All the features are then normalized. Every text sample was represented in three ways which include bag-of-words (BoW), Word2Vec and FastText techniques. For Bag-of-words, the Term Frequency-Inverse Document Frequency (TF-IDF) technique was used for weight adjustment. Whereas Pre-trained vectors were used for Word2Vec and FastText. 17 datasets based on different domains in Portuguese language were used to train these vectors. The vectors utilized in Word2Vec and FastText training are based on Skip-Gram approach. Fake Br Corpus is formulated by first using the pre-trained vector for each word and then averaged. Z-score normalization is applied during experimentation on linguistic-based features. [46]The classification methods used in this model include logistic regression (LR), support vector machines (SVM), decision trees (DT), random forest (RF), bootstrap aggregating (bagging), and adaptive boosting (Ad- aBoost). Scikit-learn library was used for the implementation of all mentioned methods. Linear kernel is used to evaluate SVM based experiments. The performance of model is evaluated on Legitimate news blocked rate (LBR) or false positive rate, Fake news caught rate (FCR) or recall, Fake news precision rate (FPR) and F-measure. Seven classification methods were studied with linguistic based features and the best results were obtained with Random Forest with LBR of 6%, FCR of 94.1%, FPR of 94.0% and F1 score as high as 94.1%. while DT obtained the worst results. In experiment with full text and truncated text, three feature sets were applied i.e., BoW, Word2Vec and FastText and from these, Bag-of-Word obtained the highest score for both cases. But it is observed

that the performance of model decreased with truncated text as it removed several important features that are used for distinguishing fake news from true news. Since fake news contains misspelled words and slangs, which are regarded as noises, and require a noisy document along with a well written document for training the model. But due to the nonavailability of word embeddings that are trained with corpus based on a combination of well written and noisy documents in Portuguese language, it is not incorporated into this model and is a limitation of this work.

2.4 Early Detection

The early Detection category contains all those researches that focus primarily on the early detection and propagation of fake news and to efficiently learn by means of the classifier when the news is first posted on the internet.

Wang et al, [47] has presented SemSeq4FD which is a graph-based neural network model for the task of fake news detection. The model is designed to detect fake news early and is based on enhanced text representation. The model is developed by presenting pair-wise semantic relations between various sentences and representing them in the form of a graph. Self-attention mechanism is applied using graph convolutional network and global sentence representation is studied. To extract the exact meaning of a sentence, 1D CNN is employed to get the local context of sentence. Enhanced sentence representation is achieved by combining the two representations. LSTM based network is applied afterwards that models the enhanced sentence representation achieved by the combination of two representations. This provides a final document representation which is used for fake news detection. This model is applied on four datasets in Chinese and English including cross-domain and cross-source datasets. The Chinese datasets are Weibo and RCED and the English datasets used in this method are LUN and SLN. SemSeq4FD is a novel model which is a combination of global semantic relationship and local sequential order and thus provides enhanced text representation which is utilized for fake news detection. The framework consists of three modules namely Sentence Encoding, Sentence Representation and Document Representation. In sentence encoding, LSTM encoder is used to bring word embeddings into a fixed-length sentence embedding. In sentence representation, two neural networks are used which are self-attention-based graph convolutional network and Text-CNN i.e., 1D convolutional neural network. In the last module, which is Document representation, LSTM is applied to get an entire document level representation because of its ability to learn contextualized representations. The proposed SemSeq4FD model is compared against three types of baselines models i.e., machine learning models, non-graph deep learning models and graph-based deep learning models. From machine learning based models, SemSeq4FD is compared with SVM and Logistic Regression. From non-graph deep learning network methods, CNN and BERT are compared with the performance of proposed model. [47]Lastly, graph deep learning networks that are compared with SemSeq4FD include GCN, GAT and GAT2H. The performance is evaluated based on five different evaluation metrics. These include Accuracy, Precision, Recall and F1 score and Net Reclassification Improvement (NRI). The proposed model has

TABLE 4
Qualitative Analysis of Datasets in Proposed data-set based detection schemes

Study	Dataset used	Nature of Dataset	Limitation	Contribution
[46]	Fake. Br (Brazilian Portuguese)	Mix news containing fake news and persuasive information	Nonavailability of noisy document to train the model	Comprehensive analysis on best feature and best machine learning method for fake news detection
				Formulated a reference corpus for real and fake news in Portuguese language
[37]	Qprop and its translation into Urdu Language	Mix news containing Commentary	Translating a data set might have a negative impact on the results.	Developing a new dataset in Urdu language by translating Qprop from English
				Developing a LIWC for the Urdu language by translating the original English LIWC
				Analyzing several different feature sets extracted from datasets
[41]	CASIA-WebFace	The real face pictures of models	It is about only face images and neglect other sorts of images.	Developing iFakeFaceDB dataset with the help of GANprintR approach.
	VGGFace2	Real images of actors, athletes and politicians		proposing an approach to spoof facial manipulation detection systems
	TPDNE	Fake and edited face images		
	100K-Faces	Fake photos of male models		
	PGGAN	Fake mix face images		
	iFakeFaceDB	Fake and edited face images		

outperformed the baselines methods in both cross-domain and cross-source datasets. In the cross-source dataset (LUN-test), SemSeq4FD achieved accuracy of 93.78%, precision of 93.90%, recall of 93.78% and F1 score of 93.78%. For the cross-domain dataset (SLN), the proposed achieved accuracy of 93.78%, precision of 93.90%, recall of 93.78% and F1 score of 93.78%. A limitation of this model is that as the length of news increases, the model's performance is affected and therefore, the length of text directly affects the classification results. Another drawback of the proposed approach is that it has slightly less accuracy compared to BERT model.

Zhou et al. [48] leveraged established social and psychological theories and supervised classification for the automatic detection of fake news. The model was developed for early detection of fake news to prevent its propagation on social media platforms. The news articles are represented using hand-crafted features covering both the style across language levels and content structure. Traditional supervised learning classifiers which include SVM, Random Forests and XGBoost were used in this study alongside Word2Vec and Doc2Vec embeddings. BuzzFeed and PolitiFact datasets¹ were used in this study. On both datasets, the proposed method was reported to perform better on all considered performance metrics which include accuracy, precision, recall, and F1 score. However, as in many previous studies, the experimental analysis was limited to only text-based news articles.

Liu et al. [49] leveraged deep learning for early detection of fake news. The major components of the model include a feature extractor, CNN-based news classifier, and a Positive Unlabeled (PU) learning framework. The datasets used in this study described in [50; 51] were generated from Twitter and Weibo. The Weibo dataset consists of 2,351 true news, 2,313 fake news, 2,309 source users and 2,818,002 Retweeters. The Twitter dataset was modified and the final version consists of 353 true news, 327 fake news, 277 source users and 215,691 Retweeters. Experimental results on the two

datasets show a superior performance in comparison with baseline results with an accuracy greater than 90%.

Zubiaga & Jiang [52] proposed a method for hoax detection in social media by using a free collaborative knowledge base and class-specific word embedding. A logistic regression classifier was used to measure the effectiveness of the class-specific word representations. The dataset² utilized in this study was generated from Twitter (using the keyword "RIP", largely associated with death reports) for three years. It contains about 94.2M tweets. The rationale behind the "RIP" keyword is that reports of the death of celebrities are often contaminated with hoaxes. The evaluation metrics used in this study is the F1 score. The experimental results using the proposed method achieved an F1 score of 72% which is better than baseline results based on social features, textual features, Google's word embedding model and textual features using sentence embedding. An obvious limitation of this study is that the data framework is limited to celebrity death report hoaxes making it difficult to generalize to real-world hoax and fake news detection. An interesting research direction suggested in the paper is to explore the proposed method in a more general event scenario.

Zhao et al. [38] applied network analysis methods to understand the distinguishing features between real news and fake news based on their propagation network. The main research goal is to study the evolution and topology of the network propagation in different settings. The experimental analysis focuses on the network with the largest component and topological features such as the shape of a network. The dataset used in this study are user accounts collected from Weibo and Twitter social media platforms. The users account activities and posting relationships were used to generate the networks. The evaluation metrics used in this study include the layer ratio, structural heterogeneity, and characteristic distance, a global feature that measures the distance between pairs of nodes. The experimental results show that all the three metrics, especially characteristic distance produced different results for real news and fake

1. <https://github.com/KaiDMML/FakeNewsNet/tree/old-version/Data>

2. https://figshare.com/articles/dataset/Twitter_Death_Hoaxes_dataset/5688811

news. The three metrics were also utilized as parameters in an SVM classifier which produced a classification accuracy of 79.5%.

2.5 Stance Detection

In Stance Detection, we have categorized all studies which have reflected on the phenomenon of understanding the fake news while checking the stance of all news reporters that are reporting about the same incident in other mediums and utilizing the stance to build a stance detector.

In order to effectively detecting, describing, and modeling fake news spreading on online social media[53], Xu et al. [54] propose a framework that characterizes the Web sites and reputations of the fake and real news publishers, analyzes the similarity between the real and fake news through TF-IDF and Latent Dirichlet Allocation (LDA) topic modeling, and also with the help of Jaccard similarity measures, document similarity between real, fake and hybrid news articles are analyzed. The experiments are conducted on BuzzFeed News dataset[55]. The news articles in this dataset are characterized from two points of view: the reputation of the domains and content understanding. According to domain reputation analysis, the Web sites of the real and fake news producers own different registration behaviors, the timing of registration, domain rankings, and the popularity of domain. Moreover, fake news are liable to be deleted on websites after a certain time period. The characterization of articles based on the content (second perspective) are conducted in three ways: 1) In an experiment designed to identify important topics in each class, tf-idf is used to extract the most frequent tokens. The extracted terms from three corpora (real, fake and hybrid) are very similar, so applying these terms alone to detect fake news is not efficient. 2) Similarly, in another analysis LDA topic modeling is explored to catch the represented topics in real, fake and hybrid news sets separately. After extracting most common terms for the top latent topics belonging to each corpus, it can be seen that both news classes share noticeable similarity in the overall topics, therefore, similar to previous parameters, the LDA model alone is ineffective for detecting fake news in the real world. 3) Finally, the last experiment is analyzing similarity between real and fake news samples. Firstly, 67% of the fake and real news are selected, randomly, as training set and the remaining 33% are used for test. After that, for each news n in the test set, by utilizing Jaccard method the document similarity between n and every news in the fake news training corpus F and the real news training corpus R is measured. The average document similarity between fake news in the test set and the news in F is much higher than the value for those in R . However, this value for the real news in the test set is surprisingly similar for both R and F sets. As a result, document similarity has the potential to detect fake news. In conclusion, some of the proposed features in this study can be efficient for detecting fake news and even for early detection as they are available at early stages of the news dissemination process, but most of them tend to be imitated easily by fake news producers as well.

In this article [56], the authors have proposed a stance detection-based approach to detect fake news. They have

focused on determining the relative relevance between the headline and the news body based on four stance labels i.e., either agree, disagree, unrelated, or discuss. In stance detection, the readers can either agree or disagree with the news headline based on its text body or remain neutral. For the same purpose, the authors have proposed an automated news classifier based on hybrid deep learning models i.e., CNN and LSTM. The classification has been done based on the above-mentioned four stance labels. To carry out the work, they have selected the Fake News Challenge dataset³ which consists of 75,385 labeled samples and 2,587 article bodies, that relate to 2532 headlines. For each claim, 5 to 20 news articles are provided. Necessary data-preprocessing has been carried out on this dataset where noise in the form of stop-words, punctuation marks, etc., have been removed. Some other pre-processing steps including stemming and word-embedding have been explored as well. Finally, four models i.e., CNN-LSTM with pre-processing, CNN-LSTM without pre-processing, CNN-LSTM with PCA, and CNN-LSTM with Chi-Square[57], [58] have been tested. CNN-LSTM with PCA model has achieved a better accuracy of 97.8 percent compared to all tested models.

2.6 Feature based Detection

Almost all researches focus on extracting the features of fake news and then utilize them to train their classifier, however, some of these researches have worked on novel and newer features extraction like topological features, semantic features, etc. that has helped these researchers make a robust detection model, these studies have been sorted under the category of Feature-based detection.

Oliveira et al. [59] present a stylistic-computational analysis, which is based on NLP and uses one-class SVM technique for detecting fake news on social media. The dataset used in this study contains 33,000 tweets. After applying NLP methods such as tokenization, removing stop words, etc. TF-IDF is used to represent prepared documents. Following previous steps, 13,716 unique words are achieved from the dataset and by applying LSA approach[60], the number of dimensions is reduced to 2000. Authors develop three different approaches for classifying the news in the final dataset: 1) *The Reduction Methodology With Training* - This model apply One-class (SVM) which is an unsupervised learning method that considers a set of training instances for a single label(True News) and classify any new samples as fake news if they are unable to fit the decision hyperplane derived by the training set [61], [35]. 2) *The Matrix Transformation Methodology* - This method is exactly similar to the former one with the exception of training data dimensions. Following LSA technique, matrix transformation method is used to reduce the dimensions of samples further to 56 from 2000. In this paper, K-means algorithm, through Elbow and Silhouette, is used to determine 56(k). At last on the final matrix, similar to the previous methodology, the one-class SVM classifier is applied. 3) *The Radial Limit Methodology* - In this approach each sample of both classes in training data is converted to R^2 which is the sum of the square of all features, so a set of random variables related to each class of news is achieved. A Statistical analysis of the both

3. <http://www.fakenewschallenge.org/>, Accessed: Oct. 29, 2019.

TABLE 5
Qualitative Analysis of Datasets in Early Detection Techniques

Study	Dataset used	Nature of Dataset	Limitation	Contribution
[38]	Weibo dataset	Mix news containing Commentary, Feature Writing, persuasive information and fake news	Five hours to detect fake news cannot be considered as early detection.	Proposing a technique to design propagation networks of fake and real news
	Twitter dataset	Social media news containing real and fake news	The propagation networks are applicable solely on platforms like Twitter.	studying topological features
[47]	Weibo	Mix news containing Commentary, Feature Writing, persuasive information and fake news	Performance of model decreases with lengthy news	Global semantic relationship and local sequential are included
	RCED	Social media news containing Commentary, Feature Writing, persuasive information, real news and citizen journalism	Less accurate than BERT model	Enhanced sentence representation using self-attention graph convolutional network and 1D cnn
	LUN	Hybrid political news containing Commentary, Feature Writing, persuasive information, real news and citizen journalism		global sequential order document representation
	SLN	Hybrid political news containing Commentary, Feature Writing and persuasive information		using four data sets in two languages
[48]	BuzzFeed	Hybrid Political news containing persuasive information	The features are extracted solely from contextual content of news while images and videos can also be used.	Proposing a theory-driven model for fake news detection
				Investigating news content at lexicon-level, syntax-level, semantic-level, and discourse-level
				Analyzing the relationships among fake news, deception/ disinformation, and click baits
[49]	Weibo	Mix news containing Commentary, Feature Writing, persuasive information and fake news	Representing a news article with k status-sensitive crowd responses is not applicable for websites	Proposing a status-sensitive crowd response feature extractor
	Twitter	Social media news containing real and fake news	the current publicly accessible experimental datasets adopted are small	Using a position-aware attention mechanism
				Applying a multi-region mean-pooling mechanism
[52]	Twitter	Fake and real death reports	Existing datasets pose limitations when one wants to investigate the earliness of veracity classification models	Investigate the application of PU-Learning algorithm
			Proposed model is limited to a specific kind of hoax triggered by death reports	Proposing a semi-automated method to build a large-scale dataset Leveraging class-specific word embeddings to represent instances

sets of random variables represents a considerable vertical shift in the amounts by R^2 and also in the accumulated probabilities depending on the instances. Thus, they developed a probabilistic classifier based on the values of the random variable R^2 . At the end of the paper, in order to evaluate these three models, the ROC curve that reflects the performance of a binary classifier is applied. The ROC curve values achieved by the *matrix transformation methodology*, the *radial limit methodology* and the *reduction methodology with training* are 0.82, 0.63 and 0.54, respectively. Also, the *matrix transformation methodology* represented the best precision at 94%, while the highest accuracy was 86% which belonged to the *reduction methodology*. Methods discussed in this paper can be beneficial for early detection as the required features for the proposed models are available as soon as a news is released.

Schuster et al. [62] developed a benchmark for the detection of fake news produced by language models that label textual content as “real” or “fake” according to its truthfulness score. The study uses Grover ⁴, a model for neural fake news generation and detection. A similar setup described in Zellers et al. [63] was used in this work. The goal of the attacker is to generate a large fake corpus using

a language model so that the verifier can classify it as real. An adaptive and zero-shot verifier settings were used in this study. The dataset used in this study were also generated using various strategies such as prompting a language model to contaminate a news article with an untruthful, but relevant claim or by a complete modification of the news with false information. The evaluation metrics utilized in this study include accuracy, recall, precision, and F1 score. An accuracy of 71% was realized for the true vs false binary classification. An F1 score of 0.9 was achieved with the zero-shot classifier while an F1 score of 0.94 was obtained in the adaptive setting. This result indicates that manipulated or false text produced using a language model might be more difficult to detect than manual or hand-crafted malicious text.

Li et al. [64] leveraged deep learning by proposing a multi-level convolutional neural network (MCNN) and a method of calculating the weight of sensitive words for fake news detection in textual news articles. The method provides deeper semantic analysis and understanding of news article text and its veracity through the relationship between the article text content and the corresponding weight of sensitive words it invokes. The study utilizes the following benchmark fake news datasets (i) LIAR, compiled by Wang et al. [65] (ii) microblog datasets from Twitter

4. <https://github.com/rowanz/grover>

TABLE 6
Qualitative Analysis of Datasets in Stance Detection Studies

Study	Dataset used	Nature of Dataset	Limitation	Contribution
[54]	BuzzFeed News dataset	Hybrid Political news containing persuasive information	It is not suitable for early detection.	Analyzing domain reputations related features
			Experiments are conducted Only in English language.	Analyzing the similarity and dissimilarity of the fake and real news via (TF-IDF) and (LDA) topic modeling
			Experiments are conducted Only in one platform (websites).	document similarity analysis via Jaccard similarity measures
[56]	FNC dataset	Headlines and article bodies containing the element of commentary	Experiments are conducted Only in English language.	Using stance detection to improve fake news detection
			Using only pre-trained word embedding (word2vec) and neglecting updating method during training.	

and Sina Weibo compiled by Ma et al. [50] and Ma et al. [51] respectively (iii) NewsFN⁵ and (iv) KaggleFN⁶. These datasets were combined to form Dataset I (news articles from Weibo, Twitter, and NewsFN) and Datasets II (news articles from LIAR and KaggleFN). The proposed framework was compared with three baseline approaches on the accuracy metric. The comparison results indicate that the proposed technique outperforms the three baseline approaches on both Dataset I (91.67%) and Dataset II (92.08%). One limitation of this study is that both dataset samples were relatively small Dataset I (4180 total, 2100 fake, and 2080 real) and Dataset II (6728 total, 3518 fake, and 3210 real). Using such a small dataset on a deep learning network without data augmentation or the use of pre-trained models may lead to overfitting.

2.7 Ensemble learning based detection

Ensemble Learning based Detection contains all papers that have utilized the approach of ensemble learning model in their detection to achieve better results.

Elhadad et al. [66] developed a model for detecting misleading information related to the COVID-19 pandemic in the English language by applying an ensemble learning approach, which combines different classifiers often machine-based ones. Most of the previous work relies on manually labeled datasets, while in this article, a technique is introduced to collect ground-truth from credible and unbiased information sources. Thus, they get the COVID-19 related ground-truth mostly by scraping the information on the WHO websites and its branches in various regions, and of course the UN and UNICEF related sources. Moreover, they benefit from the Google Fact Check Tools API to search for facts about the COVID-19. Then, the data gathered from various sources are organized in a unique form via some pre-processing such as duplicate removal, data standardization, data integration, as well as the labels are converted to either Misleading or Real. Following that, some other pre-processing operations such as Text Parsing, Data Cleaning, Part of Speech (PoS) Tagging, Stop Words Removal and Stemming are applied. At the same time, they conducted Exploratory Data Analysis (EDA) to gain some insights on data. Next, in the Feature Selection stage by applying some methods like keeping only the terms tagged as

Adjectives, Verbs and Nouns, and removing words shorter than three characters, etc. important features are indicated. The output of former steps is a corpus of stemmed BoW which play the role of input for feature extraction techniques such as TF and TF-IDF with (character level, Unigram, Bigram, Trigram, and N-gram), and also word embedding. Then, they developed detection models using ten classifiers: Decision Tree (DT), kNN, Logistic Regression (LR), Linear SVM, Multinomial Naïve Bayes (MNB), Bernoulli Naïve Bayes (BNB), Perceptron, Neural Network (NN), Ensemble Random Forest (ERF), and XGBoost. These classification algorithms are used with the mentioned feature extraction methods, and their performance are evaluated with twelve measures. From the obtained results, we can see that despite the LR, DT, and NN classifiers giving the highest performance from various points of view, all the results are satisfactory. Thus, the collected ground-truth data are valid and the voting ensemble model is developed to benefit from different classifiers.

Huang et al [67] has used a deep learning model for fake news detection. He used a combination of four deep learning techniques namely LSTM, depth LSTM, LIWC CNN, and N-gram CNN and formulated an ensemble learning model. Further he applied Self-Adaptive Harmony Search (SAHS) algorithm to optimize model weights and attained accuracy as high as 99.4%. He also used his developed ensemble learning model to study cross domain intractability issue. For feature analysis, the Fake-or-Real-news dataset was used. For the preprocessing, Word2vec is used for word embedding part, Stanford NLP Parser Tool for Grammar analysis, LIWC for text analysis and classic Ngrams algorithm for extracting more frequently used words. For training the model, LSTM and depth LSTM are used. Three data sets are used to train the model, namely Buzzfeed corpus, Satire data set, and Political dataset. Similarly, three datasets were used to study the cross-domain intractability issue using the developed ensemble learning model. These data sets include Fake or Real news (FOR) data set, Snopes Fake Legit news (SFL) data set, and Fake News Detection (FND) data set. The proposed ensemble learning method is then compared with the already existing methods based on three metrics of accuracy, i.e., performance of four models, Non-Optimized performance (proposed model with equal weights), and Optimized performance (proposed model with optimized weights). Huang et al's[67]Optimized En-

5. https://github.com/joolsa/fake_real_news_dataset

6. <https://www.kaggle.com/mrisdal/fake-news>

TABLE 7
Qualitative Analysis of Datasets in Feature based Detection Techniques

Study	Dataset used	Nature of Dataset	Limitation	Contribution
[59]	33.000 tweets	Real and fake tweets collected from reliable accounts and the profile "Boatos.org"	It is not a real-time method.	Using TF-IDF to represent documents
			Experiments are conducted Only in English language.	Using LSA technique and Matrix Transformation Methodology to reduce dimensions
			Experiments are conducted Only in one platform (Twitter).	Using Radial Limit Methodology to show samples
[62]	newsQA	Mix news containing both real news and commentary.	Other datasets that represent a wide range of LM applications, from whole-article generation to forms of hybrid writing and editing are required.	Proving the inability of stylometry against machine-generated misinformation
	New York Times articles	mix news articles containing real news		Utilizing LMs to generate synthetic fake and true news
	Modified New York Times articles	mix news articles containing fake news and commentary		
[64]	LIAR	Hybrid Political news containing persuasive information	Both datasets were relatively small, so the deep learning network are liable to be overfitted.	Proposing a method for calculating the weight of sensitive words
	Twitter	Social media news containing real and fake news		Designing and evaluating Multi-level convolutional neural network
	Weibo	Mix news containing Commentary, Feature Writing, persuasive information and fake news		

semble Learning Model attained best precision, recall and F1 value irrespective of dataset. The proposed model achieved precision, recall and F1 score of 99.4%. For cross-domain intractability, a combination of SAHS-Train and SAHS-Test is used, and the performance is evaluated. The best accuracy is attained with an optimized ensemble learning model when SAHS-Train and SAHS-Test are the same and had the accuracy of 91.9%. The limitation of this work is that the auxiliary information like pictures is not considered and only text is considered for training the model. This reduces the chance of differentiating between a real and fake news. Another limitation of the study is that the feature analysis to differentiate between real and fake news is not done. The extracted features must be analyzed to train the model for identification of a particular fake news. Through feature analysis, the relevant features can be determined and detected before we train our model. Also, the proposed model is not for an early identification of fake news and it cannot respond to an early propagation of a false news spreading on social media platforms. Similarly, in [68], the authors have proposed another ensemble based method to detect fake news.

Table 9 comprehensively shows machine / deep learning-based methods that achieved a score of above 90% in their evaluation metrics. The highlighted cells show the positive results achieved by the researchers.

Figure 5 shows all the methods used by the studies chosen for this survey categorized into 3 main learning techniques: Deep Learning, Machine Learning and Natural Language Processing. It can be seen that the researchers have spent a significant time in machine learning models and have now started to explore deep learning models to build a robust fake news detector.

3 OPEN RESEARCH CHALLENGES

During the literature review, we found that the researchers encountered many challenges out of which we selected the most common and critical challenges that need to be

addressed in order to make an efficient classifier. Table 10 summarizes the important research challenges found during literature study along with their causes.

3.1 Variable Length of News Content

A recent challenge for fake news detection is regarding the size of news. Generally, the size of true news is larger than that of fake news. To analyze the effect of length of text on the performance of model, this survey studies both full text and truncated text and the results showed that full text achieved comparatively lower results. A major research challenge lies in this regard. Truncated text proves to be an effective way of finding fake news in real-world scenarios as full texts can be easily tricked by increasing the length of fake news. A good performance classifier with full text is the second research challenge.

3.2 Long-distance document dependence

Sentence and document level-based methods perform better for the task of fake news detection, but these methods cannot be applied for long distance document dependence. Sentences that are related to semantics have the possibility of not being positioned close in the document. As the document contains long-distance dependent structures, the model hardly captures the semantic content available in the document. Besides, the algorithms used for this purpose are prone to overfitting and do not have satisfactory generalizing ability. Since the fake news spread on various platforms and through different sources of network with a wide variation of languages and domains. The models must be robust enough to perform efficiently without depending on a particular domain or source. The overfitting problem makes the existing algorithms unfit to be applied on news from diverse sources and domains.

TABLE 8
Qualitative Analysis of Datasets in Ensemble Learning based Detection Techniques

Study	Dataset used	Nature of Dataset	Limitation	Contribution
[66]	7,486 ground-truth data instances collected from reliable sources	ground-truth data related to COVID-19	It covers only one language (English), one topic (COVID-19) and one platform (web-site).	Developing a new Dataset via a semi-automated approach
[67]	Fake_or_Real_news (FOR) dataset	Hybrid Political news containing real and fake news	Auxiliary information not included	Proposing a voting ensemble classifier to detect misinformation related to COVID-19.
	Snopes_Fake_Legit_news (SFL) data set	Political news containing fake news	No feature analysis	Formulated a new model using four techniques
	Fake News Detection (FND) data set	Political news containing fake news	Does not work for early detection of false news propagation	SAHS algorithm for model weights optimization
				Comparison with other methods on different domain-oriented data sets
				Explored the Cross-domain intractability issue

Fig. 5. Methods used in this survey categorized under NLP, DL and ML respectively.

3.3 Manipulation of Deep Fake algorithms in images and videos

Most previous fake news research studies were focused on news in text formats using traditional supervised learning algorithms. However, fake news is widely spread in image and video formats [69]. Deep fake algorithms can create almost real fake images and videos that cannot be comprehended by a human brain as authentic, hence can take the internet by storm by putting deep fakes to serve a propaganda[70]. This is mainly due to the evolution of free and easy-to-use tools for manipulating images and videos which are available in numerous quantities[71]. Therefore,

relying solely on text or features based on user network and activity may not apply to many real-life scenarios. Images or videos embedded in fake news posts can either be maliciously tampered or real images that are misleading. Interestingly, fake news images and videos often have various distinguishing characteristics from real-news images at both physical and semantic levels. A major research challenge in this direction is to develop effective methods for feature representation and extraction in multimedia fake news data sets [72].

TABLE 9
Evaluation metrics for best performing method in aforementioned studies.

Study	Machine Learning Approaches				NLP Technique used to gain better results	Deep Learning Approaches			
	Accuracy	Precision	F1	Others		Accuracy	Precision	F1	Others
	>90%	>90%	>90%	>90%		>90%	>90%	>90%	>90%
Henrique et al. [40]	✓	✗	✓	✗	✗	✗	✗	✗	✗
Huang et al [67]	✗	✗	✗	✗	✗	✓	✓	✓	✗
Alatas et al [16]	✓	✓	✓	✗	✗	✗	✗	✗	✗
Kaliyar et al [39]	✗	✗	✗	✗	✗	✓	✓	✓	✗
Wang et al. [47]	✗	✗	✗	✗	✗	✓	✓	✓	✓
Silva et al. [46]	✓	✓	✓	✓	✓	✗	✗	✗	✗
Oliveira et al. [59]	✗	✓	✗	✗	✓	✗	✗	✗	✗
Xu et al. [54]	✗	✗	✗	✗	✗	✗	✗	✗	✗
Elhadad et al. [66]	✓	✓	✓	✓	✓	✗	✗	✗	✗
Umer et al. [56]	✗	✗	✗	✗	✓	✓	✓	✓	✓
Neves et al. [41]	✗	✗	✗	✗	✗	✗	✗	✗	✗
Kausar et al. [37] (Urdu Language)	✓	✓	✓	✓	✓	✗	✗	✗	✗
Zhou et al. [48]	✗	✗	✗	✓	✓	✗	✗	✗	✗
Liu et al. [49]	✗	✗	✗	✗	✗	✓	✓	✓	✓
Zubiaga & Jiang [52]	✗	✗	✗	✗	✓	✗	✗	✗	✗
Schuster et al. [62]	✗	✗	✗	✗	✗	✗	✗	✗	✗
Li et al. [64]	✗	✗	✗	✗	✓	✓	✗	✓	✓
Zhao et al. [38]	✗	✗	✗	✗	✗	✗	✗	✗	✗

3.4 Encrypted messages

The automatic detection of fake news in encrypted social media platforms such as WhatsApp, Instagram, WeChat and Telegram is very challenging [69]. For instance, WhatsApp is organized in such a way that its contents circulate in specific communication circles making it difficult to be tracked due to the end-to-end encryption. This ensures that only the users involved in the conversation have access to the content shared, shielding abusive content from being detected or removed [73]. While this feature protects data from being read or secretly modified by a third party, it becomes an obstacle to fake news moderation. An interesting research opportunity relevant to this challenge is to develop effective methods for fake news detection while ensuring that encrypted messages are kept secured.

3.5 Training data limitations

A large-scale data set consisting of the ground truth of both fake and real news is vital in understanding relationships among different types of fake news, however, such data sets are so far rarely available. The data sets used in many previous fake news detection models are limited with regards to features and data size. Some models, such as the one in [54] rely on the publisher's domain or user account reputation score. Other models such as the one in [64] were built on relatively small data size that can hardly generalize well to real-world scenarios. The effect of data set size is even more pronounced in deep learning models. Also, with the rise in the use of bots and cloned accounts that closely mimic real users on social media, domain reputation might not be a good feature for detecting fake news, even highly reputable Web sites may publish or share fake news for a variety of reasons.

3.6 The drawbacks of the manual techniques of data set construction

To develop a real time model, the most significant challenge is the scarcity of sufficient and proper data. Previous studies

have shown each domain needs its relevant data set and the same one can not be applied to all domains. This is the case for each language and using machine translation similar to [37] for constructing a new data set could not be a helpful solution as may imply misinterpretations or ambiguities and have negative impacts on the effectiveness of detection processes [74; 75]. On the other hand, the manual approach of creating data sets and verifying by experts is time-consuming, expensive, biased, probably subjective, and tends to be impractical because of the huge volume of available data on the internet [76; 77; 78]. Furthermore, even the best data sets need to be updated regularly to remain effective for training an accurate model.

3.7 The shortage of discriminating features

Another challenge in the area of fake news detection is the shortage of discriminatory features. As we know, fake news producers can imitate real news writers and this makes it difficult to extract enough suitable features to distinguish both classes. In addition, most of the efficient characteristics to differentiate fake news from real ones are not available at the early stages of releasing the news hence usually early detection fails because of the lack of such types of features that can conveniently classify a fake news from real news. Also, some features are available solely in some platforms and can not be utilized in others like user based features are specifically utilized in Facebook and other social media.

3.8 Identification of News Bots (Cyborgs)

There are several types of automated accounts that have slightly different characteristics and are applied for various applications. While there is a possibility that some users who are real will be sharing somethings only occasionally whereas the bots will obviously be storming the internet with hundreds and thousands of tweets that are often just linked to only one subject. The content is often liked to be reposted instead of being fresh and authentic. However, a kind of hybrid account that is a combination of a

bot's tirelessness with human subtlety is known as cyborgs. Cyborg accounts are practically those in which a human occasionally and periodically takes over the bot account to respond to queries and comments of other users and to post some new authentic content. They are not only more expensive and time consuming to work, but also they are better hidiers as robots⁷. Operators of such accounts also are often called "cyborgs" because of this mix of automated and "human" posts. The activators often use the social media management platform Hootsuite, to control many accounts simultaneously⁸. The challenge applies across all sorts of platforms, including and just not limited to dating apps, twitter, online therapy services as well as Facebook⁹. Many studies have investigated fake accounts detection problems. For instance, [79] has proposed a simple and effective approach in which they extract digital DNA sequences from users online actions to model their behaviour. Also, [80] have used a similar technique and has improved it to cover many more features of a tweet by a character label. In these studies, According to the lifestyle of user accounts, their fingerprints are produced. With the help of this simple method, they have managed to show that automated accounts have less diverse behaviour than genuine user accounts. Even though Cyborgs are a sophisticated kind of bots and have a very similar behaviour to real users, utilizing such methods along with other feature sets can help in detecting them.

4 FUTURE RESEARCH DIRECTIONS

From the survey, we conclude that fake news detection can be further improved by working on the following areas.

4.1 Blockchain fake news detection

Blockchain technology, with its transparent and trace ability can now be a game turner in the world of fake news detection based on the fact that it permanently records any transaction being held at any date and cannot be manipulated to portray any wrong transaction over time [81; 82; 83]. Using the multiple advantages of Blockchain's peer-to-peer network concepts, a way to find and detect fake news on social media can be helpful. It is now possible to not only verify that the information is genuine but also check its sources and easily build the most needed trust in the news that are being displayed on Internet [84]. The Blockchain enabled platform can also provide the online readers with not just a reliable way of checking the content and the source with which its originating, but also the profiles that register on the platform will be checked to see if they have any authentic connection with genuine news agencies or journalism.

4.2 Deep fake detection

The detection of deepfakes is an interesting area for future research [85; 86]. Visual forensics features are normally used

for the manipulation detection of image and video detection. However, most forensics features are created manually just for detecting specified manipulated traces and are not applicable for real images attached in fake news. In addition, these hand-crafted features are labor-intensive and limited in terms of learning complicated patterns leading to poor generalization performance on the task of fake news detection [72]. Deep learning methods for computer vision is a promising approach for detecting manipulated and misleading contents in multimedia data because of the recent successes and growing popularity of transfer learning and the use of open-source pre-trained models for boosting detection performance even on a small dataset [69].

4.3 Synthetic training data generation for fake news detection

When working with machine learning, especially deep learning, it is important to have a high-quality data set to train the algorithm. This means that the data should not only be sufficiently large, to cover as many cases as possible, but also be a good representation of the reality. Synthetic data sets can also be generated to augment existing data sets and improve the prediction results. The creation of synthetic data can be useful for several reasons including the goal of avoiding using the original data for privacy reasons and oversampling minority classes when learning from imbalanced data, a common problem in cyber security. Synthetic data sets can also be generated to augment existing data sets and improve the prediction results. An open research area is to develop easy-to-use techniques for generating synthetic data sets of various formats including images and videos that closely resemble the target data. Generative Adversarial Networks (GAN) ¹⁰ is an innovative technology that can be leveraged to generate synthetic data in various formats for fake news detection.

4.4 User profile based feature

One of the most important features to be considered in detection of fake news is the user profile-based feature [87]. This will give a prior knowledge about the profiles spreading fake news around the medium[88](e.g. number of posts, account age, the followers number, etc.) and any of their posts will be taken as suspicious. It is obvious that while considering the user based features, approach-ability is the most important factor to be taken into account. Due to privacy constraints, data of users are not readily available and user interactions are protected [89; 90]. Previous studies have shown that bots have been seen in propagation of fake news over the internet so a bot detection technique that will distinguish a bot account from that of a normal user can be really helpful in better exploitation of user profile features in fake news detection [91].

4.5 Multi-model approach

A very serious shortcoming of previous researches for identification of fake news is that they are unable to learn a feature representation of information like multi modal

7. <https://www.voanews.com/silicon-valley-technology/cyborgs-trolls-and-bots-guide-online-misinformation>

8. <https://www.bbc.com/news/world-latin-america-42322064>

9. <https://medium.com/@DFRLab/human-bot-or-cyborg-41273cdb1e17>

10. <https://developers.google.com/machine-learning/gan>

TABLE 10
Open challenges and their causes

Challenges	Causes
Variable Length of news content	Full texts can be easily tricked by increasing the length of fake news.
Long-distance document dependence	Document contains long-distance dependent structures, the model hardly captures the semantic content available in the document
Manipulated images and videos	The evolution of free and easy-to-use tools for manipulating images and videos
Encrypted messages	Automatic detection of fake news in encrypted social media platforms such as WhatsApp, Instagram, WeChat and Telegram is very challenging
Training data limitations	The datasets used in many previous fake news detection models are limited with regards to features and data size.
Manual construction of datasets	The manual approach of creating datasets and verifying by experts is slow, expensive, highly subjective and biased
Shortage of Discriminatory features	(a) Difficulty in obtaining suitable features to distinguish fake from real (b) Efficient characteristics not available at early detection (c) Not all features are available on every platform

(textual+visual)[92]. There have been methods focusing on multi-modal detection but they employ an additional sub-task to try to discriminate events and find correlations across these modalities. The outcomes are not only heavily dependent on these sub-tasks but also in their absence can reduce the performance of the detector by up to 10%[92]. So for future research can be the inclusion of multi-modal approach using a combination of different learning techniques utilizing the strength of each to combine to form an efficient and robust fake news detector for textual plus visual based information. Furthermore, this approach will aid in solving the problem of multi-class fake news detection.

4.6 New and efficient methods for data set construction

Developing semi-automated techniques for constructing data sets and gathering data from reliable sources and credible knowledge bases can be considered as a productive and exciting field of study. For instance in [66], by using ground-truth obtained from credible and unbiased information sources a model was developed for (COVID-19) which can be applied to other topics, domains and platforms as well as in other languages as suggested by the authors to evaluate its efficiency thoroughly.

4.7 Extracting new discriminatory features

As we know for developing accurate models, in addition to clean and exact data and strong classifiers we need discriminating features and as discussed in [46], there is not yet a consensus on what combination of feature sets is optimal for early detection of fake news or at least for specific conditions. Even though one of the benefits of Deep Learning techniques is to eliminate the need of manual feature extraction, training such models with proper features

as opposed to all unprocessed data might have a positive impact on the performance as can be seen in [56]. Thus, extracting and analyzing new features that are capable to show different aspects of news can be considered as a field that needs to be studied further. For instance, in [67] only the sentence depths of news articles is considered as a new discriminating feature, while more helpful characteristics can be extracted in the patterns of sentences of an article through analyzing grammar and also all 130 NELA features.

4.8 Detection of Online Religious Content

Social media platforms are often bombarded with fake religious content[93; 94; 95] to create panic among the people. There have been numerous attempts to alter the religious content[96; 97; 98; 99] and spread the fake information [100] by sharing tampered verses through different media outlets. Hence, detecting the fake religious content which can be in different languages such as Arabic[101], Portuguese [102], Persian, Chinese etc, is one of the future research direction that needs to be addressed.

4.9 AI explainable multi-modal credibility analysis systems

With the recent advancement in social media systems, the line between fake news and that of facts have blurred. A very good research direction would be in the field of AI explainable credibility analysis system for fake news on social media. A manual censor board is not efficient enough nor capable of handling news on large scale to detect and explain why its Fake news [103]. One such automated framework exists for health blogs that gives credibility score for author, text and image of the blog and also provide user-friendly explanation. These framework systems can be further extended towards other domains like political

agenda, education, health misinformation etc., and should take into perspective the credibility of author's claims as well to detect misinformation. [104]

5 CONCLUSION

With the development in digital world, the online fake content is increasing drastically. Being in common people's access, such fake widespread content can cause potential set-back to journalism and democracy by misleading people. The decisions and opinions of the public are greatly influenced by false content as they spread much quicker and leave a great impact. Such contents gain popularity because most of the users on social media are unaware about certain topics and are easily deluded by fake content. Other reasons may include people's reliance on online media platforms and the catchy draft of fake news. With ever increasing fake content, the research on the fake news detection is also making progress and various approaches from vast domains have been implemented ranging from artificial intelligence to linguistic and knowledge engineering but no ideal methodology has been devised yet that can accurately classify real news and fake news. The major challenge in this task is the growing social media content which is increasing exponentially daily. The social sites and apps like Facebook, Twitter, Instagram, and other blogs have enabled people to put any random unchecked content including their personal opinions and thoughts that has become quite challenging in devising a fake news detection algorithm. In this paper, a thorough study on existing fake news detection techniques along-with a new taxonomy is included and major challenges of fake news detection have been discussed with future recommendations to further improve this area.

REFERENCES

- [1] A. Waqar, A. Raza, H. Abbas, and M. K. Khan, "A framework for preservation of cloud users' data privacy using dynamic reconstruction of metadata," *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 235–248, 2013.
- [2] S.-J. Horng, S.-F. Tzeng, Y. Pan, P. Fan, X. Wang, T. Li, and M. K. Khan, "b-specs+: Batch verification for secure pseudonymous authentication in vanet," *IEEE transactions on information forensics and security*, vol. 8, no. 11, pp. 1860–1875, 2013.
- [3] H. Ghayvat, S. N. Pandya, P. Bhattacharya, M. Zuhair, M. Rashid, S. Hakak, and K. Dev, "Cp-bdhca: Blockchain-based confidentiality-privacy preserving big data scheme for healthcare clouds and applications," *IEEE Journal of Biomedical and Health Informatics*, 2021.
- [4] S. Hakak, S. Ray, W. Z. Khan, and E. Scheme, "A framework for edge-assisted healthcare data analytics using federated learning," in *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 2020, pp. 3423–3427.
- [5] F. Wang, D. Jiang, H. Wen, and S. Qi, "Security level protection for intelligent terminals based on differential privacy," *Telecommunication Systems*, vol. 74, no. 4, pp. 425–435, 2020.
- [6] Z. Liu, X. Huang, Z. Hu, M. K. Khan, H. Seo, and L. Zhou, "On emerging family of elliptic curves to secure internet of things: Ecc comes of age," *IEEE Transactions on Dependable and Secure Computing*, vol. 14, no. 3, pp. 237–248, 2016.
- [7] C. Higgins, L. McDonald, M. Ijaz Ul Haq, and S. Hakak, "Iot hardware-based security: A generalized review of threats and countermeasures," *Security and Privacy in the Internet of Things: Architectures, Techniques, and Applications*, pp. 261–296, 2021.
- [8] T. R. Gadekallu, M. Manoj, N. Kumar, S. Hakak, S. Bhattacharya *et al.*, "Blockchain-based attack detection on machine learning algorithms for iot-based e-health applications," *IEEE Internet of Things Magazine*, vol. 4, no. 3, pp. 30–33, 2021.
- [9] S. Kumari, M. K. Khan, and R. Kumar, "Cryptanalysis and improvement of 'a privacy enhanced scheme for telecare medical information systems'," *Journal of medical systems*, vol. 37, no. 4, pp. 1–11, 2013.
- [10] M. K. Khan, "Fingerprint biometric-based self-authentication and deniable authentication schemes for the electronic world," *IETE Technical Review*, vol. 26, no. 3, pp. 191–195, 2009.
- [11] M. K. Khan and K. Alghathbar, "Cryptanalysis and security improvements of 'two-factor user authentication in wireless sensor networks'," *Sensors*, vol. 10, no. 3, pp. 2450–2459, 2010.
- [12] S.-F. Tzeng, S.-J. Horng, T. Li, X. Wang, P.-H. Huang, and M. K. Khan, "Enhancing security and privacy for identity-based batch verification scheme in vanets," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3235–3248, 2015.
- [13] F. K. Parast, C. Sindhav, S. Nikam, H. I. Yekta, K. B. Kent, and S. Hakak, "Cloud computing security: A survey of service-based models," *Computers & Security*, vol. 114, p. 102580, 2022.
- [14] C. Beaman, A. Barkworth, T. D. Akande, S. Hakak, and M. K. Khan, "Ransomware: Recent advances, analysis, challenges and future research directions," *Computers & Security*, vol. 111, p. 102490, 2021.
- [15] N. Mashtalyar, U. N. Ntaganzwa, T. Santos, S. Hakak, and S. Ray, "Social engineering attacks: Recent advances and challenges," in *International Conference on Human-Computer Interaction*. Springer, 2021, pp. 417–431.
- [16] F. A. Ozbay and B. Alatas, "Fake news detection within online social media using supervised artificial intelligence algorithms," *Physica A: Statistical Mechanics and its Applications*, vol. 540, p. 123174, 2020.
- [17] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22–36, 2017.
- [18] T. Diehl and S. Lee, "Testing the cognitive involvement hypothesis on social media: news finds me' perceptions, partisanship, and fake news credibility," *Computers in Human Behavior*, vol. 128, p. 107121, 2022.
- [19] A. Gupta, H. Li, A. Farnoush, and W. Jiang, "Understanding patterns of covid infodemic: A systematic and pragmatic approach to curb fake news," *Journal*

- of business research*, vol. 140, pp. 670–683, 2022.
- [20] A. Bodaghi and J. Oliveira, “The theater of fake news spreading, who plays which role? a study on real graphs of spreading on twitter,” *Expert Systems with Applications*, vol. 189, p. 116110, 2022.
 - [21] Fake news and its impact on the economy. [Online]. Available: <https://priorityconsultants.com/blog/fake-news-and-its-impact-on-the-economy/>
 - [22] J. Ashton, “Covid-19 and the anti-vaxxers,” *Journal of the Royal Society of Medicine*, vol. 114, no. 1, pp. 42–43, 2021.
 - [23] T. Burki, “The online anti-vaccine movement in the age of covid-19,” *The Lancet Digital Health*, vol. 2, no. 10, pp. e504–e505, 2020.
 - [24] Y. H. Khan, T. H. Mallhi, N. H. Alotaibi, A. I. Alzarea, A. S. Alanazi, N. Tanveer, and F. K. Hashmi, “Threat of covid-19 vaccine hesitancy in pakistan: the need for measures to neutralize misleading narratives,” *The American journal of tropical medicine and hygiene*, vol. 103, no. 2, pp. 603–604, 2020.
 - [25] G. Vaidyanathan, “News feature: Finding a vaccine for misinformation,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 32, pp. 18 902–18 905, 2020.
 - [26] M. Scerri and V. Grech, “Covid-19, its novel vaccination and fake news—what a brew,” *Early Human Development*, p. 105256, 2020.
 - [27] S. Hakak, W. Z. Khan, M. Imran, K.-K. R. Choo, and M. Shoaib, “Have you been a victim of covid-19-related cyber incidents? survey, taxonomy, and mitigation strategies,” *IEEE Access*, vol. 8, pp. 124 134–124 144, 2020.
 - [28] S. Khan, S. Hakak, N. Deepa, B. Prabadevi, K. Dev, and S. Trelova, “Detecting covid-19-related fake news using feature extraction,” *Frontiers in Public Health*, vol. 9, 2021.
 - [29] L. Zheng, J. D. Elhai, M. Miao, Y. Wang, Y. Wang, and Y. Gan, “Health-related fake news during the covid-19 pandemic: perceived trust and information search,” *Internet Research*, 2022.
 - [30] A. Bovet and H. A. Makse, “Influence of fake news in twitter during the 2016 us presidential election,” *Nature communications*, vol. 10, no. 1, pp. 1–14, 2019.
 - [31] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer, “Fake news on twitter during the 2016 us presidential election,” *Science*, vol. 363, no. 6425, pp. 374–378, 2019.
 - [32] X. Zhang and A. A. Ghorbani, “An overview of online fake news: Characterization, detection, and discussion,” *Information Processing & Management*, vol. 57, no. 2, p. 102025, 2020.
 - [33] S. Hakak, W. Z. Khan, S. Bhattacharya, G. T. Reddy, and K.-K. R. Choo, “Propagation of fake news on social media: Challenges and opportunities,” in *International Conference on Computational Data and Social Networks*. Springer, 2020, pp. 345–353.
 - [34] J. V. de Souza, J. Gomes Jr, F. M. de Souza Filho, A. M. de Oliveira Julio, and J. F. de Souza, “A systematic mapping on automatic classification of fake news in social media,” *Social Network Analysis and Mining*, vol. 10, no. 1, pp. 1–21, 2020.
 - [35] S. Gaonkar, S. Itagi, R. Chalippatt, A. Gaonkar, S. Aswale, and P. Shetgaonkar, “Detection of online fake news : A survey,” in *2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)*, 2019, pp. 1–6.
 - [36] X. Zhou and R. Zafarani, “A survey of fake news: Fundamental theories, detection methods, and opportunities,” *ACM Comput. Surv.*, vol. 53, no. 5, Sep. 2020. [Online]. Available: <https://doi.org/10.1145/3395046>
 - [37] S. Kausar, B. Tahir, and M. A. Mehmood, “Prosoul: A framework to identify propaganda from online urdu content,” *IEEE Access*, vol. 8, pp. 186 039–186 054, 2020.
 - [38] Z. Zhao, J. Zhao, Y. Sano, O. Levy, H. Takayasu, M. Takayasu, D. Li, J. Wu, and S. Havlin, “Fake news propagates differently from real news even at early stages of spreading,” *EPJ Data Science*, vol. 9, no. 1, p. 7, 2020.
 - [39] R. K. Kaliyar, A. Goswami, P. Narang, and S. Sinha, “Fndnet—a deep convolutional neural network for fake news detection,” *Cognitive Systems Research*, vol. 61, pp. 32–44, 2020.
 - [40] P. H. A. Faustini and T. F. Covões, “Fake news detection in multiple platforms and languages,” *Expert Systems with Applications*, p. 113503, 2020.
 - [41] J. Neves, R. Tolosana, R. Vera-Rodriguez, V. Lopes, H. Proenca, and J. Fierrez, “Ganpintr: Improved fakes and evaluation of the state of the art in face manipulation detection,” pp. 1038–1047, 04 2020.
 - [42] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” 07 2017, pp. 1800–1807.
 - [43] L. Nataraj, T. M. Mohammed, B. Manjunath, S. Chandrasekaran, A. Flenner, M. J. Bappy, and A. Roy-Chowdhury, “Detecting gan generated fake images using co-occurrence matrices,” *Electronic Imaging*, vol. 2019, pp. 532–1, 01 2019.
 - [44] F. Matern, C. Riess, and M. Stamminger, “Exploiting visual artifacts to expose deepfakes and face manipulations,” 01 2019, pp. 83–92.
 - [45] J. Ellul and K. Kellen, *Propaganda: The Formation of Men’s Attitudes*. Vintage, 1973.
 - [46] R. M. Silva, R. L. Santos, T. A. Almeida, and T. A. Pardo, “Towards automatically filtering fake news in portuguese,” *Expert Systems with Applications*, vol. 146, p. 113199, 2020.
 - [47] Y. Wang, L. Wang, Y. Yang, and T. Lian, “Semseq4fd: Integrating global semantic relationship and local sequential order to enhance text representation for fake news detection,” *Expert Systems with Applications*, vol. 166, p. 114090, 2020.
 - [48] X. Zhou, A. Jain, V. V. Phoha, and R. Zafarani, “Fake news early detection: A theory-driven model,” *Digital Threats: Research and Practice*, vol. 1, no. 2, pp. 1–25, 2020.
 - [49] Y. Liu and Y.-F. B. Wu, “Fned: A deep network for fake news early detection on social media,” *ACM Transactions on Information Systems (TOIS)*, vol. 38, no. 3, pp. 1–33, 2020.
 - [50] 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,” 2016.
 - [51] J. Ma, W. Gao, and K.-F. Wong, “Detect rumors in mi-

- croblog posts using propagation structure via kernel learning." Association for Computational Linguistics, Conference Proceedings.
- [52] A. Zubiaga and A. Jiang, "Early detection of social media hoaxes at scale," *ACM Transactions on the Web (TWEB)*, vol. 14, no. 4, pp. 1–23, 2020.
- [53] N. Ruchansky, S. Seo, and Y. Liu, "Csi: A hybrid deep model for fake news detection," 11 2017, pp. 797–806.
- [54] K. Xu, F. Wang, H. Wang, and B. Yang, "Detecting fake news over online social media via domain reputations and content understanding," *Tsinghua Science and Technology*, vol. 25, no. 1, pp. 20–27, 2020.
- [55] This analysis shows how viral fake election news stories outperformed real news on facebook. [Online]. Available: <https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook>
- [56] 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. 156 695–156 706, 2020.
- [57] M. Ahmad, D. Ihsan, and D. Ulhaq, "Linear unmixing and target detection of hyperspectral imagery using osp," 01 2011.
- [58] M. Ahmad, M. Alqarni, A. Khan, R. Hussain, M. Mazzara, and S. Distefano, "Segmented and non-segmented stacked denoising autoencoder for hyperspectral band reduction," *Optik*, vol. 180, 11 2018.
- [59] N. R. de Oliveira, D. S. V. Medeiros, and D. M. F. Mattos, "A sensitive stylistic approach to identify fake news on social networking," *IEEE Signal Processing Letters*, vol. 27, pp. 1250–1254, 2020.
- [60] N. Halko, P.-G. Martinsson, and J. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM Review*, vol. 53, pp. 217–288, 01 2011.
- [61] R. Perdisci, G. Gu, and W. Lee, "Using an ensemble of one-class svm classifiers to harden payload-based anomaly detection systems," in *Sixth International Conference on Data Mining (ICDM'06)*, 2006, pp. 488–498.
- [62] T. Schuster, R. Schuster, D. J. Shah, and R. Barzilay, "The limitations of stylometry for detecting machine-generated fake news," *Computational Linguistics*, pp. 1–12, 2020.
- [63] R. Zellers, A. Holtzman, H. Rashkin, Y. Bisk, A. Farhadi, F. Roesner, and Y. Choi, "Defending against neural fake news," in *Advances in neural information processing systems*, Conference Proceedings, pp. 9054–9065.
- [64] 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," *Personal and Ubiquitous Computing*, pp. 1–14, 2019.
- [65] W. Y. Wang, "liar, liar pants on fire": A new benchmark dataset for fake news detection," *arXiv preprint arXiv:1705.00648*, 2017.
- [66] M. K. Elhadad, K. F. Li, and F. Gebali, "Detecting misleading information on covid-19," *IEEE Access*, vol. 8, pp. 165 201–165 215, 2020.
- [67] Y.-F. Huang and P.-H. Chen, "Fake news detection using an ensemble learning model based on self-adaptive harmony search algorithms," *Expert Systems with Applications*, p. 113584, 2020.
- [68] 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 Generation Computer Systems*, vol. 117, pp. 47–58, 2021.
- [69] F. Chollet, *Deep Learning with Python, Second Edition*. Manning Publications, 2021. [Online]. Available: <https://books.google.ca/books?id=2-D8zQEACAAJ>
- [70] T. Nguyen, C. M. Nguyen, T. Nguyen, T. Duc, and S. Nahavandi, "Deep learning for deepfakes creation and detection: A survey," 09 2019.
- [71] S. Zerdoumi, A. Q. M. Sabri, A. Kamsin, I. A. T. Hashem, A. Gani, S. Hakak, M. A. Al-Garadi, and V. Chang, "Image pattern recognition in big data: taxonomy and open challenges: survey," *Multimedia Tools and Applications*, vol. 77, no. 8, pp. 10 091–10 121, 2018.
- [72] P. Qi, J. Cao, T. Yang, J. Guo, and J. Li, "Exploiting multi-domain visual information for fake news detection," in *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2019, pp. 518–527.
- [73] J. C. Reis, P. Melo, K. Garimella, J. M. Almeida, D. Eckles, and F. Benevenuto, "A dataset of fact-checked images shared on whatsapp during the brazilian and indian elections," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 14, 2020, pp. 903–908.
- [74] V. Rubin, Y. Chen, and N. Conroy, "Deception detection for news: Three types of fakes," 11 2015.
- [75] V. Rubin, "Pragmatic and cultural considerations for deception detection in asian languages," *ACM Transactions on Asian Language Information Processing*, vol. 13, 06 2014.
- [76] R. Kumar and R. Verma, "Kdd techniques : A survey," 2012.
- [77] A. Shastri and M. Deshpande, *A Review of Big Data and Its Applications in Healthcare and Public Sector*, 01 2020, pp. 55–66.
- [78] S. Changdar, "A survey of data mining applications and techniques," *International Journal of Computer Science and Information Technologies*, 12 2015.
- [79] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "Dna-inspired online behavioral modeling and its application to spambot detection," *IEEE Intelligent Systems*, vol. 31, no. 5, pp. 58–64, 2016.
- [80] D. Kosmajac and V. Keselj, "Twitter user profiling: Bot and gender identification," in *CLEF*, 2019.
- [81] S. Paul, J. I. Joy, S. Sarker, A. . A. . H. Shakib, S. Ahmed, and A. K. Das, "Fake news detection in social media using blockchain," in *2019 7th International Conference on Smart Computing Communications (ICSCC)*, 2019, pp. 1–5.
- [82] S. Hakak, W. Z. Khan, G. A. Gilkar, N. Haider, M. Imran, and M. S. Alkathiri, "Industrial wastewater management using blockchain technology: Architecture, requirements, and future directions," *IEEE Internet of Things Magazine*, vol. 3, no. 2, pp. 38–43, 2020.
- [83] S. Hakak, W. Z. Khan, G. A. Gilkar, M. Imran, and N. Guizani, "Securing smart cities through blockchain

- technology: Architecture, requirements, and challenges," *IEEE Network*, vol. 34, no. 1, pp. 8–14, 2020.
- [84] "How can blockchain solve the fake news problem." Jan 2021. [Online]. Available: <https://www.leewayhertz.com/blockchain-fake-news/>
- [85] A. Zervopoulos, A. G. Alvanou, K. Bezas, A. Papamichail, M. Maragoudakis, and K. Kermanidis, "Deep learning for fake news detection on twitter regarding the 2019 hong kong protests," *Neural Computing and Applications*, vol. 34, no. 2, pp. 969–982, 2022.
- [86] A. M. Almars, "Deepfakes detection techniques using deep learning: a survey," *Journal of Computer and Communications*, vol. 9, no. 5, pp. 20–35, 2021.
- [87] S. Agarwal and A. Samavedhi, "Profiling fake news: Learning the semantics and characterisation of misinformation," in *International Conference on Advanced Data Mining and Applications*. Springer, 2022, pp. 203–216.
- [88] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, and R. Procter, "Detection and resolution of rumours in social media: A survey," *ACM Computing Surveys (CSUR)*, vol. 51, no. 2, pp. 1–36, 2018.
- [89] W. Z. Khan, M. Y. Aalsalem, M. K. Khan, and Q. Arshad, "Data and privacy: Getting consumers to trust products enabled by the internet of things," *IEEE Consumer Electronics Magazine*, vol. 8, no. 2, pp. 35–38, 2019.
- [90] W. Z. Khan, M. Y. Aalsalem, and M. K. Khan, "Communal acts of iot consumers: A potential threat to security and privacy," *IEEE Transactions on Consumer Electronics*, vol. 65, no. 1, pp. 64–72, 2018.
- [91] K. Shu, S. Wang, and H. Liu, "Understanding user profiles on social media for fake news detection," in *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*. IEEE, 2018, pp. 430–435.
- [92] P. Shah, "Multimodal fake news detection using a cultural algorithm with situational and normative knowledge," 06 2020.
- [93] S. Hakak, A. Kamsin, O. Tayan, M. Y. I. Idris, A. Gani, and S. Zerdoumi, "Preserving content integrity of digital holy quran: Survey and open challenges," *Ieee Access*, vol. 5, pp. 7305–7325, 2017.
- [94] S. Hakak, A. Kamsin, W. Zada Khan, A. Zakari, M. Imran, K. bin Ahmad, and G. Amin Gilkar, "Digital hadith authentication: Recent advances, open challenges, and future directions," *Transactions on Emerging Telecommunications Technologies*, p. e3977, 2020.
- [95] K. Nanath, S. Kaitheri, S. Malik, and S. Mustafa, "Examination of fake news from a viral perspective: an interplay of emotions, resonance, and sentiments," *Journal of Systems and Information Technology*, 2022.
- [96] S. Hakak, A. Kamsin, S. Palaiahnakote, O. Tayan, M. Y. Idna Idris, and K. Z. Abukhir, "Residual-based approach for authenticating pattern of multi-style diacritical arabic texts," *PloS one*, vol. 13, no. 6, p. e0198284, 2018.
- [97] S. Hakak, A. Kamsin, O. Tayan, M. Y. I. Idris, and G. A. Gilkar, "Approaches for preserving content integrity of sensitive online arabic content: A survey and research challenges," *Information Processing & Management*, vol. 56, no. 2, pp. 367–380, 2019.
- [98] S. Hakak, A. Kamsin, J. Veri, R. Ritonga, and T. Herawan, "A framework for authentication of digital quran," in *Information Systems Design and Intelligent Applications*. Springer, 2018, pp. 752–764.
- [99] G. A. Gilkar, S. Hakak, W. Z. Khan, and H. H. Alshamrani, "Content integrity techniques for digital quran," in *2020 International Conference on Computing and Information Technology (ICCIT-1441)*. IEEE, 2020, pp. 1–4.
- [100] S. Senbel, C. Seigel, and E. Bryan, "Word frequency analysis of community reaction to religious violence on social media," in *Intelligent Computing*. Springer, 2022, pp. 626–633.
- [101] F. Harrag and M. K. Djahli, "Arabic fake news detection: A fact checking based deep learning approach," *Transactions on Asian and Low-Resource Language Information Processing*, vol. 21, no. 4, pp. 1–34, 2022.
- [102] Á. F. L. d. Sousa, G. Schneider, H. E. F. d. Carvalho, L. B. d. Oliveira, S. V. M. A. Lima, A. R. d. Sousa, T. M. E. d. Araújo, E. L. S. Camargo, M. O. B. Oriá, C. V. Ramos *et al.*, "Covid-19 misinformation in portuguese-speaking countries: Agreement with content and associated factors," *Sustainability*, vol. 14, no. 1, p. 235, 2022.
- [103] A. Gupta, A. Anjum, S. Gupta, and R. Katarya, "Recent trends of fake news detection: A review," *Machine Learning, Advances in Computing, Renewable Energy and Communication*, pp. 483–492, 2022.
- [104] V. Wagle, K. Kaur, P. Kamat, S. Patil, and K. Kotecha, "Explainable ai for multimodal credibility analysis: Case study of online beauty health (mis)-information," *IEEE Access*, vol. 9, pp. 127 985–128 022, 2021.