



An intelligent cybersecurity system for detecting fake news in social media websites

Ala Mughaid¹ · Shadi Al-Zu'bi² · Ahmed AL Arjan¹ · Rula AL-Amrat¹ · Rathaa Alajmi¹ · Raed Abu Zitar³ · Laith Abualigah^{4,5}

Accepted: 2 March 2022 / Published online: 21 April 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

People worldwide suffer from fake news in many life aspects, healthcare, transportation, education, economics, and many others. Therefore, many researchers have considered seeking techniques for automatically detecting fake news in the last decade. The most popular news agencies use e-publishing on their websites; even websites can publish any news they want. However, thus before quotation any news from a website, there should be a close look at news resource ranking by using a trusted websites classifier, such as the website world rank, which reflects the repute of these websites. This paper uses the world rank of news websites as the main factor of news accuracy by using two widespread and trusted websites ranking. Moreover, a secondary factor is proposed to compute the news accuracy similarity by comparing the current news with fakes news and getting the possible news accuracy. Experiments results are conducted on several benchmark datasets. The results showed that the proposed method got promising results compared to other comparative methods in defining the news accuracy.

Keywords Artificial intelligence · Cybersecurity · Fake news · Text processing · Cyber privacy · Social media accuracy

1 Introduction

At first, when the Internet was made, it was accessible for public use Abualigah et al. (2021b); Batur Şahin and Abualigah (2021). The primary purpose of the Internet was for accessing information Attiya et al. (2022). However, over the years, the Internet has grown with many untruthful and misleading information, and many previous researchers were

working in Cybercrimes such as in Al-Masalha et al. (2020). Fake news found its way into the Internet, and it becomes difficult for people to find truthful information. Fake news is spread online through social media and fake news websites. Current social media is fertile ground for the spread of fake news Sadiku et al. (2018); Wu et al. (2022). By which content can circulate among users with no third party to arbitrate Otair et al. (2022). Misinformation amplified by new means in the internet age poses a threat to society worldwide.

Fake news is fabricated content deceptively presented as real news Spradling et al. (2021); Goldani et al. (2021); Şahin et al. (2021). It consists of stories designed to increase readership, online sharing, and Internet click revenue Freire et al. (2021). Fake news is published on the Internet to mislead to damage an agency, person, or rival. Fake news spreads faster and more profound than truth. There are many types of Fake as some are Clickbait: exaggerated in reporting false news created for generating clicks to increase ad revenue. E.g., proclaiming that drinking two gallons of water a day is good, chocolate will help lose weight. Propaganda: deceptive story designed to promote the author's agenda. It may be politically motivated. Politicians and governments use propaganda to promote their agenda. Opinion: The author's commentary was invented to influence the reader. Humor: this story is

✉ Ala Mughaid
ala.mughaid@hu.edu.jo
Laith Abualigah
laythyabat@aau.edu.jo

¹ Department of Information Technology, Faculty of prince Al-Hussien bin Abdullah for IT, The Hashemite University, PO Box 330127, Zarqa 13133, Jordan

² Faculty of Science and IT, Al-Zaytoonah University of Jordan, Amman, Jordan

³ Sorbonne Center of Artificial Intelligence, Sorbonne University-Abu Dhabi, Abu Dhabi 38044, United Arab Emirates

⁴ Faculty of Computer Sciences and Informatics, Amman Arab University, Amman 11953, Jordan

⁵ School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang 11800, Malaysia

used for entertainment and satire to discuss public affairs. The authors promote themselves as delivering entertainment and call themselves comedians instead of journalists Choudhary and Arora (2021); Kaliyar et al. (2021).

Several works and approaches have been proposed to uncover fake news Gandomi et al. (2022); Paka et al. (2021); Sivasankari and Vadivu (2021); Fayaz et al. (2022). Saikh et al. in Saikh et al. (2020) presented two models that rely on deep learning to solve the problem of detecting fake news in content using several domains across the Internet. The proposed models are binary classification that aims to classify fake and verified content Abualigah and Diabat (2022). They evaluated these models based on two pre-trained data sets: fake news AMT and celebrity for fake news detection. Then, by testing both models using data sets, they reached an accurate amount of 77.08%, 83.3% when using fake news AMT, and 76.53%, 79% when using celebrity datasets. In another work, Rubin et al. in Saikh et al. (2020); Pérez-Rosas et al. (2017a); Rubin et al. (2016) used a dataset containing satirical news (The Onion and fact-checking sites like Politi-fact and Snopes), but these sites are restricted in a particular domain of news like politics. Perez Rosas et al. (2017) Pérez-Rosas et al. (2017a) created two datasets for a fake news detection mission to cover seven different news domains. They also used exploratory analytics to identify language differences in legitimate and fake news content using the numbering features, followed by N-grams, Complete LIWC, and Syntax features.

Ahmed et al. in Ahmed et al. (2018) presented an approach to detecting pseudo-content that begins with preprocessing the dataset by removing unnecessary characters and words from the data such as stop words, removing punctuation and tokenization of the text, and then do stemming from extracting the n-gram features, and the required document was created. The final step in the classification process is training the classifier. Six different machine learning algorithms were tested, namely (SGD), SVM, (LSVM), K-Nearest Neighbor (KNN), LR, and DT. Veronica Perez-Rosas, et al in Pérez-Rosas et al. (2017b), they search the new ways of fake news classification that published over the Internet such as social media for instantly, in which they proposed mechanism depends on online fake news detection by twofold, Firstly, by the novel two datasets of fake news detection in seven different domains in news, secondly, they conduct a set of learning of experiments to build accurate fake news detectors, in which they used two ways to build a novel datasets, in which used the manual and crowdsourced annotation efforts to build the first dataset and used direct collect from the web to build the second dataset, in which they using these datasets to conduct several exploratory analyses to identify linguistic properties that are predominantly present in fake content and build fake news detectors depends on linguistic features that achieve accuracies of up to 78%, such as a combination of

lexical, syntactic, and semantic information as well features representing text readability properties.

Rafael A. Monteiro et al. in Monteiro et al. (2018), the researchers studied the Brazilian Portuguese language problem in fake news detection for other languages by using analyzing linguistic characteristics and machine learning, in which they created a corpus by collecting samples of fake news and labeled it, which they called "Fake.Br corpus," after they created an own corpus, which is used it in build automatic fake news classifier by using main two of procedures, the first procedure is reducing of text size by truncating the text. The second used Linear SVC technique, in which, after applying all features, machine learning on Fake was used Fake.Br corpus to obtain 89% of the final accuracy. Hanslowski et al. in Hanselowski et al. (2018) conducted a retrospective analysis of the three best participating systems in the fake news challenge to enhance artificial intelligence techniques to combat fake news; through their research paper, they provided a deep analysis of these three best-performing systems, critical evaluation of the experimental setup and also suggested a new measure based on F1 Added to the features used and conducted a detailed analysis of these features.

Reese et al. in Reis et al. (2019) reviewed many previous studies through their research paper and focused on identifying the proposed features in this work, on implementing these features on a recently released data set, consisting of news articles sourced from BuzzFeed related to the US elections, where they explained the use of these features to distinguish between news the truth and the fake, at first they ignored all the stories classified as unreal or false, and the mixture of right and wrong as one group which is fake news, and the rest is accurate. Classify the features into features extracted from news content, news source, and published news environments and then evaluate these features based on some machine learning algorithm. Thorne et al. in Thorne et al. (2018) introduced a new publicly available dataset for checking and verifying textual sources (FEVER: Fact-Extraction and Verification), generated by Wikipedia. In their paper, Vlachos et al. in Vlachos and Riedel (2014), they present the task of verifying information and facts by creating a data set using data validated by fact-checking websites such as PolitiFact, and then explaining the correlation and relationship of fact-checking with natural language processing tasks.

M. Al-Khair et al. Alkhair et al. (2019) presented an Arab group of false news for three Arab celebrities through the information published on YouTube, specifically about the news of their death. They performed a statistical analysis of the data collected and then classified the news using three methods: support vector Machine, Multinomial Naïve Bayes, Decision Tree. To distinguish between rumors and lack of comments, it has been noted that the performance achieved varies depending on the subject of the rumors. A method for analyzing and evaluating the credibility of tweets on Twitter

was proposed by Castillo et al. Castillo et al. (2011), where they divided these tweets as credible or unreliable, based on the features extracted from the tweets, the most important of these features is the behavior of users in the posting process and the number of retweets made by users, where they explained that the publication of reliable news comes from authors who have previously written a large number of topics and have many republishes. Qazvinian and others Qazvinian et al. (2011) relied on several of topics also spread on Twitter and investigated tweets on this news, suggested a classification of support or denial of stories and also relied on features based on content, on the network and some elements of Twitter, and found that the classification on topics increases inaccuracy with the increase in Tweets. Other optimization techniques can be used to solve the Fake news detection Abualigah et al. (2021,2021,2022); Agushaka et al. (2022); Oyelade et al. (2022).

It is essential to be aware of how the above techniques can be used to employ fake news worldly, by giving one example which took place shortly before the most recent USA presidential election in a series of events known as "pizza-gate." Fake news publishers in Macedonia circulated a false political conspiracy theory that former First lady, secretary of state, and presidential candidate Hillary Clinton and other prominent democratic political figures coordinated a child trafficking ring out of a Washington, DC pizzeria the name of Comet Ping Pong. This false news publication was widely shared via Facebook and directed readers to websites to generate advertising revenue. In the bizarre turn of events in December 2016, a man who read the fake publication drove from North Carolina to Washington, DC, and shot open a locked door at the actual Comet Ping Pong pizzeria Klein and Wueller (2018).

In the era of AI, several intelligent applications have been employed with a huge amount of cloud data such as in AlZu'bi et al. (2018, 2019); Elbes et al. (2020); Al-Zu'bi et al. (2021); AlZu'bi and Jararweh (2020); Guo et al. (2021). The most popular news agencies use e-publishing on their websites; even websites can publish any news they want. However, thus before quotation any news from a website, there should be a close look at news resource ranking by using a trusted websites classifier, such as the website world rank, which reflects the repute of these websites. In this paper, we proposed a new method to detect fake news by proposing a novel algorithm of fake news accuracy detection according to the sequences, aggregate, and Live news. The proposed method collects websites ranking of news sources by RankAPI and Alexa website. It collects related news from the Google search engine by live news titles. After executing the news accuracy algorithm, compute the rank of source news website from total news accuracy 50%, and apply to stop words, Tokenizing, Stemming for both text and title of fake news dataset and Text and Title of living news. Then,

implements cosine similarity for last Compute cosine similarity of living news text with related news text from total news accuracy 50% scores. This paper uses the world rank of news websites as the main factor of news accuracy by using two widespread and trusted websites ranking. Moreover, a secondary factor is proposed to compute the news accuracy similarity by comparing the current news with fakes news and getting the possible news accuracy. Experiments results are conducted on several benchmark datasets. The results showed that the proposed method got promising results compared to other comparative methods in defining the news accuracy.

The main sections of this paper are organized as follows: Sect. 2 shows the proposed method for accurate news definition. Section 3 presents the experiments and results. Section 4 shows the discussion and comparisons with state-of-the-art methods. The conclusion and future work directions are given in Sect. 5.

2 Methodology

The proposed system depends on checking the news accuracy published by the news websites, using a different way of the previous techniques that only used machine learning to classify the news as either fake or real. However, this is not enough to classify the news because machine learning depends on words similarity between two documents might be one for real and the other for fake, but they could have high words similarity. The methodology of this work depends on two. Firstly, the news text similarity across fake news datasets such as Kaggle Secondly, ranking of websites for the source text. The ranking system employed two different sources of website ranking, such RankAPI and Alexa, which are specialists for website ranking Aldwairi and Alwahedi (2018a); Aphiwongsophon and Chongstitvatana (2018). The text similarity was given 50%, and the other 50% were given to news websites ranking. Finally, we compute the news accuracy by sum these percentages from 100% as news accuracy percentages.

2.1 Data set

The actual word is done using a fake news dataset generated by the Kaggle website, which is used to train the machine to check the current news similarity with the fake news dataset. In which a full training using attributes such that (id, title, author, text, label) kaggle (2021b).

The fake news detection programmed tool depends on two scenarios:

1. Detect news accuracy by website ranking:

Fig. 1 Fake news dataset features

▲ title	▲ text	▲ subject	📅 date
Donald Trump Sends Out Embarrassing New Year's Eve Message; This is Disturbing	Donald Trump just couldn't wish all Americans a Happy New Year and leave it at that. Instead, he had...	News	December 31, 2017
Drunk Bragging Trump Staffer Started Russian Collusion Investigation	House Intelligence Committee Chairman Devin Nunes is going to have a bad day. He's been under the as...	News	December 31, 2017
Sheriff David Clarke Becomes An Internet Joke For Threatening To Poke People 'In The Eye'	On Friday, it was revealed that former Milwaukee Sheriff David Clarke, who was being considered for ...	News	December 30, 2017
Trump Is So Obsessed He Even Has Obama's Name Coded Into His Website (IMAGES)	On Christmas day, Donald Trump announced that he would be back to work the following day, but he	News	December 29, 2017

In this stage, we gather the rank of the news publisher website by two rank factors, the first from the Alexa website and the second from the Rankapi website. We will illustrate it in section news websites ranking ratio.

2. Detect news accuracy by getting top1 cosine similarity from fake news dataset and compute the true news percentage:

In this paper, the fake news dataset has been considered, and it contains a list of articles considered as "fake" news from the Kaggle website (Fig. 1 illustrates a sample dataset), which contains four columns title, text, subject, and date kaggle (2021a):

This paper depends on lexical similarity for current text news with previous fake news saved in the fake news dataset. We applied cosine similarity for token text with all fake news texts in the dataset. We got one of the most similarities from fake news texts as a similarity factor from 0 to 1. In this paper, we except the id, title, author, and label features, and we used just text feature as a similarity factor.

2.2 Proposed system

In the proposed system, we built a nontraditional algorithm for fake news detection that depends on two main factors. The first factor depends on features collected from worldly news ranking websites, and the other factor is using news text similarity a crossing with previous fake news dataset. Then, using those factors to compute the final news accuracy. However, the proposed system works in two stages, as illustrated as follows:

Stage 1: The system automatically collects the latest news from the news API, and it some-up some features for each news such as news title, news weblink, news source, and getting news text by scraping it from news webpage that has been published this news. Then, the system prepares those features for the next stage to achieve some formulas.

Stage 2: This stage uses the features that are achieved from the previous stage to compute some other formulas, as shown below illustration of the formulas:

– **Compute website news ranking (R):**

In this section, the system computes R-value by summing up R1 and R2, in which R1 and R2 are two ranking sources of news websites as will be illustrated later on, the following formula illustrates how to compute the website ranking ratio:

The formula in equation 1 illustrates how to compute the website ranking ratio:

$$R = R1 + R2 \tag{1}$$

Where:

- R: is the final ratio of news website ranking.
- R1: is Rank API ratio of news website ranking.
- R2: is Alexa ratio of news website ranking.

– **Compute the top-one similarity of news text:(TP)**

In this section, the system computes the more remarkable similarity fake news texts from the fake news dataset, and the following formula illustrates how to compute the cosine similarity ratio with N fake news counts from the fake news dataset:

The formula in equation 2 illustrates how to compute the cosine similarity ratio of the current news text (CN) with N fake news counts from the fake news dataset:

$$TP = Top(Cosine - Similarity(CN, FN_i)) \tag{2}$$

Where:

- TP: is True Positive rate with fake news.
- Top: return the top1 similarity rate of fake news texts (FN_i) with current news text (N).
- CN: is the tested news text.
- FN_i: is Fake News text in the dataset, i=1,...,N.

– **Compute False Positive rate of current news (FP):**

In this section, the system computes FP after collecting the similarity ratio of fake news. The formula in equation 3 illustrates how to compute the False Positive ratio of false news based on the True Positive ratio of fake news.

$$FP = (1 - TP) * 0.5 \tag{3}$$

Where:

- FP: is False Positive ratio of fake news.
- TP: is True Positive ratio of true news.
- Note: We multiply by 0.5 because the news text similarity is 50% of total news accuracy.

Table 1 Ratio Rank API of news website

Rank of 10 (R1)	Ratio %
10	25
9	21
8	19
7	17
6	15
5	13
4	9
3	7
2	5
1	1
0	0

– **Compute the final accuracy of current news (Accuracy):**

In this section, the system computes the final accuracy of current news. The following formula illustrates how to compute the final ratio of news accuracy:

The formula in equation 4 illustrates how to compute the final ratio of news accuracy:

$$Accuracy = R + FP \tag{4}$$

Where:

- Accuracy: is News Accuracy ratio from 100%.
- R: is news website ranking ratio from 50%
- FP: is False Positive of tested news ratio from 50%

2.3 News websites ranking ratio

Tables 1 and 2 represent the ranking ratio of some news collected from news websites. The ranking ration was given 50% out of the total news accuracy. Two websites were employed to achieve this ratio: RankAPI and Alexa, as each of these websites was given 25%, respectively. The ranking ratio of these websites expresses R1 and R2, respectively, from formula 1. The summation of R1 and R2 expresses R, which is the final ratio of website rank.

Tables 1 and 2 show ranking ratio according to the API ranking source and Alexa, which will be considered as R1 and R2, respectively, in Equation 1:

– **Rank API Factor:**

This website uses an analysis algorithm that depends on Google Page Rank, which is measures how important the web page or websites and computes the quality of content and backlinks. The rank score in this website is between 0

Table 2 Ratio Alexa Rank of news website

Alexa rank (R2)	Ratio%
From >= 1 k To <= 100 k	25
From >= 200 k To < 100 k	21
From >= 300 k To < 200 k	19
From >= 400 k To < 300 k	17
From >= 500 k To < 400 k	15
From >= 600 k To < 700 k	13
From >= 700 k To < 800 k	9
From >= 800 k To < 900 k	7
From >= 900 k To <= 1 m	5
More than 1 m	1
None	0

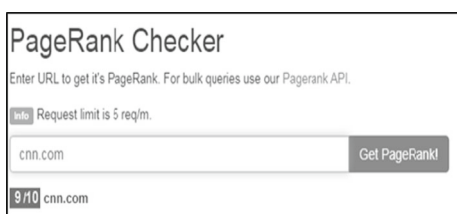


Fig. 2 CNN.com website rank (API)

and 10, where 10 is the highest page rank and 0 lowest page rank. The page rank brings the website closer to a top position in Google search results, and vice versa Rankapi (2011). Figure 2 illustrates an example of how to get CNN website rank from Rank API website:

– **Rank Alexa Factor:**

The Alexa rank ranking system owned by Amazon uses web classification by dynamic data inside the website to classify the top popularity websites. It ranks many websites as of recognition, with the lower site Alexa rank. It also looks at how an internet site is doing relative to other sites, making it a proper benchmarking or competitive analysis. It uses to classification millions of websites, as the rank result number refers to website rank crossing millions of websites Duo (2021). Figure 3 illustrates an example of how to get CNN website rank by Alexa website. Table 3 shows that RankAPI and Alexa ranked websites are using different techniques.

2.3.1 Compute news websites ranking ratio

In this section, we will be getting the first part of the news accuracy ratio by computing R using the formula in equation 1. By getting the news website first-ranking R1 from the Rankapi website, as shown in Table 1, and then getting the

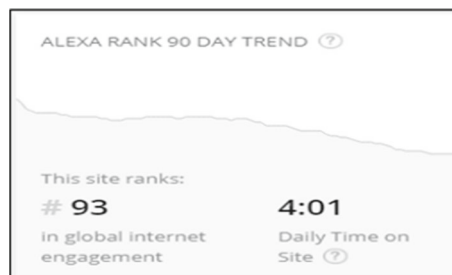


Fig. 3 CNN.com website rank (Alexa)

Table 3 RankAPI and Alexa classification criteria

Rank API factor	Rank alexa factor
Computes the quality of content and backlinks	Using web classification by dynamic data inside the website to classify the foremost popularity websites
Rank score in this website (0–10)	
10 highest page rank	
0 lowest page rank	

news website second-ranking R2 from the Alexa website (2). Finally, we will compute the R-value using Equation 1.

The following example shows how to compute the rank ratio for the CNN website:

- Step 1: Getting first ranking from Rankapi website, it shows in Fig. 1 that CNN has (9/10) which means that R1=21% according to Table 1.
- Step 2: Getting second-ranking from Alexa website, Fig. 2 shows that CNN has (93), which means that R2=25% according to Table 2.
- Step 3: After getting each of R1 and R2, in which we can compute R by Equation 1 as follows:

$$R = R1 + R2 \rightarrow R = 21\% + 25\% = 46\% \quad (5)$$

2.3.2 News text similarity with fake news dataset

In this section, we implement current news similarity with (N) counts of fake news from the fake news dataset, and this matter demands a set of steps to obtain the second part of the news accuracy ratio. To attain that, each current news text and current fake news text must pass in text preprocessing before similarity implementation on current news text and current fake news. The following explanation shows text preprocessing and text similarity:

– **Text Pre-Processing:**

Before texts similarity implementation, we must prepare the two texts with some processing that they become ready for the similarity process as the following steps show text processing:

– **Step 1: The uniform case of letters:**

In this step, we convert all letters in the text to lowercase.

– **Step 2: Remove punctuation and non-ASCII characters:**

This step removes all punctuations from the text and all non-ASCII characters.

– **Step 3: Remove Stop-words:**

In this step, we remove all stop-words from the text, such as 'the,' 'is,' 'are,' 'a,' 'an,' and so on.

– **Step 4: Text tokenization:**

In this step, we are splitting all text into words, by using a space pattern for splitting it.

– **Step 5: Words-stemming:**

In this step, we return all text-words to basic words, i.e., remove any suffix letters at the end of a word, for example, "Playing," "played," and "Plays," all of them refer to one basic word is "Play," this is called word-stemming.

The text will be ready for the next stage, which can be prepared to implement similarities between texts.

– **Create TF matrix for texts:**

The Term frequency (TF) matrix for text terms is a two-dimensional matrix containing two rows and N columns. The rows for text-1 and text-2, the columns for all text terms, and N refer to the count of text terms. The following stage explains how to create a TF matrix.

In this stage, we aggregate all words in text-1 and text-2 as a set without repeated words. These term words are represented by columns headers and the rows labeled as text-1 and text-2, where the crossing of rows and columns is the frequent count in text. Table 4 shows TF matrix tacking in calculations the above details.

- *Text-1: The current news for test.*
- *Text-2: The current fake news for test.*
- *Term1...N: All consecutive words in the text.*

Table 4 TF matrix

	Term1	Term2	...	Term-N
Text-1	A1	A2	...	AN
Text-2	B1	B2	...	BN

- *A_i: The count of frequent the term in text-1, where i<=N.*
- *B_i: The count of frequent the term in text-2, where i>=N.*

– Cosine-Similarity for text-1 and text-2:

After converting all terms in each of text-1 and text-2 to vectors (i.e., numbers as A's and B's in TF matrix) in the previous stage, now we can compute the cosine similarity score between text-1 and text-2, by the formula in Equation 6 Nguyen and Bai (2010):

$$CS(A, B) = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \tag{6}$$

Where:

- *n: The count of terms in the text 1 & 2.*
- *A_i: The term-i frequency in text-1.*
- *B_i: The term-i frequency in text-2.*

– **Compute current news accuracy:**

- Step 1:
From Equation 2, we obtain top1, which is the cosine similarity score of current news with N fake news. Where N is the count of fake news in the dataset, we have to have TP (True Positive rate of current news).
- Step 2:
From Equation 3, we obtain FP (False Positive of current news), by subtraction one from TP, the result multiplication by 0.5 to get the second part of news accuracy (i.e., the second half of 50% of accuracy score).In the website ranking stage, we obtained the first half of the accuracy score from 50%.
- Step 3:
From Equation 4, we obtain the final news accuracy score by adding the R-value to FP-value to get NA (News accuracy score).

3 Experimental results and analysis

This section will show the results for ten news from various resources, where the first five news from high popularity sources and the second five from low popularity sources. The proposed system was trained on these samples, often news to get results and then analyzes the results. The following Sects. 4.1 and 4.2, show system results in detail.

Table 5 Aggregated News info

News title	News Source	News URL
Israeli emergency services responding to a synagogue bleacher collapse described as a 'mass casualty event'	CNN	https://www.cnn.com/2021/05/16/middleeast/israel-bleacher-collapse/index.html
Samsung's leaked Galaxy A22 may be its most affordable 5G phone to date	Engadget	https://www.engadget.com/samsung-galaxy-a22-5g-leak-193356168.html
Vaccination centre ponders cutting days to ease nurses' workload	Stuff.co.nz	https://www.stuff.co.nz/auckland/local-news/300309147/vaccination-centre-ponders-cutting-days-to-ease-nurses-workload
Starship SN15 patiently awaits a decision	NASA Spaceflight	https://www.nasaspaceflight.com/2021/05/starship-sn15-reflight-road-orbit/
John Kiely launches attack on GAA over rules and accuses Galway of simulation	The42	https://www.the42.ie/john-kiely-2-5439166-May2021/
Imperialism and national liberation	Socialist worker	https://socialistworker.co.uk/art/51842/Imperialism+and+national+liberation
Zimbabwe says it has enough cash to buy record maize crop	Insiderzim	https://www.insiderzim.com/zimbabwe-says-it-has-enough-cash-to-buy-record-maize-crop/
MC GHIE COFFEE SHOP	Pattaya people	http://pattayapeople.com/2019/03/26/mc-ghie-coffee-shop/
HCMC to further promote caretaking for poor, vulnerable people: City Party Chief	sggpnews	https://sggpnews.org.vn/hochiminhcity/hcmc-to-further-promote-caretaking-for-poor-vulnerable-people-city-party-chief-92396.html
Navi Pillay explains 'human rights' limitations in Geneva on Tamil genocide	Tamil net	https://www.tamilnet.com/art.html?catid=79&artid=39975

3.1 News tables

The system in the initial stage aggregates up-to-date news from the Newsapi server, and then, the system is trained to check for aggregated news by the methodology of the proposed technique, 5 shows some info about aggregated news.

The python program collects live news from the live news API server, as in this paper we are detecting fake direct news, then we trained the system to check for aggregated news using the proposed technique, the following 6 shows some info about Live aggregated news which is (News title, News Source, News URL).

3.2 News websites-ranking tables

Table 6 shows the world ranking of the domain name of news websites, which we employed for differentiating the sources of websites ranking (Alexa and Rankapi) as explained previously in this research.

From Table 6, we have chosen ten news sites that published news, the first five of which are famous and accredited news sites around the world in publishing the news, and the other five are unknown or approved sites to publish news. Then, we

examined the ranking of these news sites using two ranking sites (Alexa and Rankapi) to obtain the results in Table 6.

After getting the values of news websites ranking, we are using Rankapi ratio Table (Table 1) and the Alexa ratio Table (Table 2), to compute website ranking score R using Equation 1. Table 7 shows websites ranking ratios R.

According to the methodology of the proposed technique, after getting the news websites ranking result, we took 25% of each website (Alexa)and (Rankapi), then we added together to get the first 50% of news accuracy as shown in Table 7.

3.3 News text similarity with fake news dataset

The final stage of checking news accuracy is to gather the current news text and all fake news text from the Kaggle dataset. Then, inputs the text to preprocessing for preparation to do the similarity process. The proposed system used cosine-similarity in which the system processes the current news text with 5000 fake news from the dataset to achieve similarity. Then, it gets the top1 similarity score crossing fake news and being named FP. After that, it computes TP from Equation 3. Table 8 shows FP, TP ratios, and final similarity score out of 50% based on the previous explanation.

Table 6 News websites domain names ranking

News#	Domain	Alexa-rank	Other-rank
1	www.cnn.com	96	10/10
2	www.engadget.com	1216	10/10
3	www.stuff.co.nz	4094	10/10
4	www.nasaspaceflight.com	46803	10/10
5	www.the42.ie	70248	10/10
6	https://socialistworker.co.uk	1102118	1/10
7	https://www.insiderzim.com/	8974861	1/10
8	http://pattayapeople.com/	6287109	1/10
9	https://sggpnews.org.vn/hochiminhcity	1261037	1/10
10	https://www.tamilnet.com/	1243830	1/10

Table 7 News websites-ranking R (based on Equation 1)

News#	R1-ratio (Alexa) 25%	R2-ratio (Rankapi) 25%	R=R1+R2 50%
1	25	25	50
2	25	25	50
3	25	25	50
4	25	25	50
5	25	25	50
6	1	1	2
7	1	1	2
8	1	1	2
9	1	1	2
10	1	1	2

Table 8 Cosine-Similarity score

News#	Similarity Count of 5000	TP 100%	FP 100%	FP/2 50%
1	2929	9.3	90.7	45.4
2	5000	8.5	91.5	45.7
3	5000	17.8	82.2	41.1
4	5000	7	93	46.5
5	5000	7.3	92.7	46.3
6	5000	18	82	41.0
7	5000	20	80	40.0
8	5000	24.3	75.7	37.8
9	5000	22.4	77.6	38.8
10	5000	20	80	40

From Table 8, we gather the current news text and all fake news text from the Kaggle dataset. Then, we input the text to preprocessing to do the similarity process. The proposed system used cosine-similarity in which the system processes the current news text with 5000 fake news from the dataset to achieve similarity. Then, it gets the top1 similarity score crossing fake news and being named FP. After that, it computes TP from Equation 3. Figure 4 shows how to distribute each of TP and FP, where a subset of 5000 fake news has been used from the dataset. It can be noticed from Fig. 4 that

the ratio of FP is much higher than the ratio of TP for all the ten news.

3.3.1 Duration time of similarity

Table 9 shows the duration time for each news individually, where a subset of 5000 fake news has been extracted from the dataset for similarity.

Table 9 shows the duration time for each news individually, where the system used 5000 fake news for similarity, as we note the time taken for each news was less than a minute,

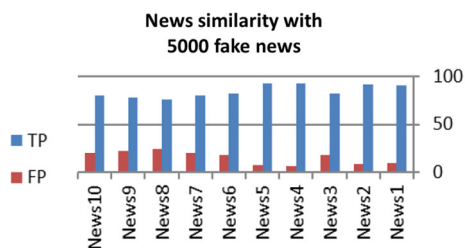


Fig. 4 Distribute TP and FP

Table 9 Time of similarity

News#	Duration time/Seconds
1	24 Seconds
2	26 Seconds
3	30 Seconds
4	39 Seconds
5	31 Seconds
6	35 Seconds
7	30 Seconds
8	33 Seconds
9	38 Seconds
10	32 Seconds

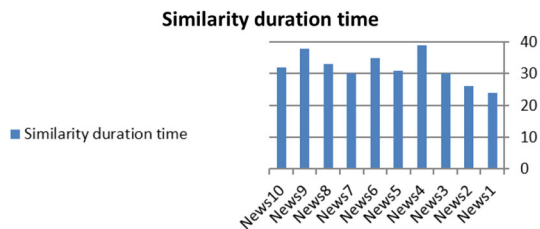


Fig. 5 Similarity duration time

which indicates the speed of the system, this helps to detect big news in record time. We noticed that the duration time increases similarity if the text is long and vice versa. Figure

Table 10 The final news accuracy score

News#	R-Score 50%	CS-Score 50%	Accuracy Score= R+CS 100%	Accuracy Label
1	50	45.4	95.4	Accurate
2	50	45.7	95.7	Accurate
3	50	41.1	91.1	Accurate
4	50	46.5	96.5	Accurate
5	50	46.3	96.3	Accurate
6	2	41.0	43	Inaccurate
7	2	40.0	42	Inaccurate
8	2	37.8	39.8	Inaccurate
9	2	38.8	40.8	Inaccurate
10	2	40	42	Inaccurate

5 illustrates how to distribute the duration time of each news for achieving similarity.

We noticed from Fig. 5 that the duration time increases in similarity if the text is long, and vice versa, if the similarity is less the time duration is short.

3.4 Computing the final news accuracy score

In this section, we are reaching the final stage of computing a news accuracy score out of 100%, based on Equation 4. Table 10 shows the final results of news accuracy for the 10 news shown in Table 5.

The results in Table 10 are considered as an acceptable achievement as the final accuracy ratio depends on news website ranking classification, and also it depends on news text similarity with a set of fake news. Now, if we consider websites that repeated fake news, thus the similarity ratio against a set of fake news is high even if the website rank has a high ratio, resulting in a low ratio in news accuracy. Furthermore, if the similarity against the set of fake news is a low ratio, and news website ranking ratio was low, thus appears the news accuracy low ratio as well, and vice versa is true. This illustrates the big difference in the news accuracy results in Table 10. In another case, if the news accuracy ratio is between 50% to 80%, that means the news similarity against the set of fake news was high, and the news website has a high ranking classification. Hence, the website popularity affects the news accuracy according to the developed system. Figure 6 illustrates the final score for the ten news out of 100% based on Ranking and Cosine.

4 Discussion

In this research, we have used two techniques for achieving news accuracy: a mix between news text processing using cosine similarity against fake news dataset and search about news website ranking by RankAPI and Alexa website. Com-

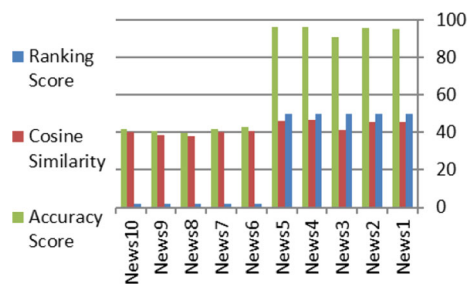


Fig. 6 The final news accuracy score

paring the proposed technique to machine learning, we took ten real news using linear model regression to train machine learning. The accuracy score of the train and test is over 95%. However, after inputting this news to the machine learning for prediction. The ML classified six news as real and four as fake, but the results were better after inputting the same news into the proposed system. Since the FP and TP of checked news referred to reality more than fake news, and the final accuracy of checked news is over 90%, this is nearing reality. The following Table 11 shows the results of comparison between the proposed technique and machine learning techniques:

Researchers have adopted different methods in detecting fake news, Table 12 presents some of them, and what are their results compared to the method adopted in this research.

In Ahmed et al. (2017), the researchers proposed a fake news detection model, using two different methods to extract the feature: Term Frequency (TF) and Term Frequency-Inverted Document Frequency (TF-IDF).

For text analysis, they used n-gram features and six machine learning classification techniques, namely: K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Logistic Regression (LR), Linear Support Vector Machine (LSVM), Decision Tree (DT), and random regression (SGD). The dataset used a dataset gathered from real news sites (Reuters.com) and fake news from the fake news dataset on kaggle.com. They used 12,600 honest news articles and 12,600 fake news articles from kaggle.com. The focus was on political news. Preprocessing of the data set was performed by removing the stop words, tokenization, lower casing, sentence segmentation, stemming (use Porter stemming), and removing punctuation marks which help in reducing the size of the data. The final step is to train the classifier by running the six machine learning algorithms and studying the effect of size n in n-gram on performance. The dataset was split into 80% for training and 20% for testing. The highest accuracy was achieved with unigram and SVM linear classifier by 92, and increasing the size of n in n-gram reduces the accuracy of the algorithms.

Finally, they ran an additional experiment by running the proposed model on a public dataset (Adali and Horne) and got 87% accuracy using n-gram and Linear SVM algorithm.

Table 11 Comparison between Machine Learning and Proposed system

News title	News source	Machine-learning prediction	Similarity with fake news	TP	News Accuracy
1- Asteroids could be approaching Earth undetected thanks to quirk of the planet's rotation	Pharmacy times.com	Fake News	0.107605	0.892395	95%
2- AT&T Stadium design plays role with sun, Jumbotron affecting multiple plays	Philippine Star	Real News	0.169867	0.830133	93%
3- Brother confirms death of animal lover in Tonga tsunami	Independent	Real News	0.162772	0.837228	96%
4- Building the off-Earth economy	The Indian Express	Real News	0.093613	0.906387	93%
5- COVID-19 Daily Bulletin	The Globe And Mail	Fake News	0.129035	0.870965	94%
6- Dutch TV suspends show over sexual misconduct claims	Reuters	Real News	0.142656	0.857344	92%
7- Everton want Belgium manager and contact Belgian Football Association	BBC News	Real News	0.120857	0.879143	93%
8- India Fighting Another Wave While Maintaining Economic Growth, Says PM At World Economic Forum	Associated Press	Real News	0.142416	0.857584	93%
9- Nightclub helping police over missing woman	The Punch	Fake News	0.047605	0.952395	93%
10- Police did not receive any rally application	Bring Me the News	Fake News	0.057764	0.942236	93%

Table 12 Comparing previous some works with the method used in the proposed search

Source number	Use cosine similarity	Use the news site rating	Compare the news with fake news	Dealing with live news	Determine the source of news
[11]	No	No	Yes	No	Yes
[12]	Yes	No	Yes	No	Yes
[13]	Yes	No	Yes	No	Yes
[14]	No	No	Yes	No	Yes
[15]	No	No	Yes	No	Yes
[16]	Yes	No	Yes	No	Yes

In the proposed model, we compared the results of only the textual content similarity criterion (which is the part similar to what the researchers did), the highest result we reached through the tested sample is 46.5 (which is 50%), we multiplied it by 2 to get the value from (100%), so the accuracy becomes (93%), and we notice that we have surpassed them in one criterion without taking into account the second criterion.

In Thota et al. (2018a), the researchers presented a proposed model based on the structure of the neural network to accurately predict the stance between a pair of the title and the article, depending on the extent of the similarity between the content of the article text and the title (classifying stance to agree, disagree, discuss, unrelated). Using the Fake News Challenge dataset (FNC-1), they split it into a 67% training set, 33% validation, or test set. Then, they preprocess the data by removing the stop words, which is the first step by using the natural language toolkit (NLTK) library, punctuation removal, and stemming. Then, they prepared the text, converted it into raw text, extracted the features by using two methods of word bag and TF-IDF, then trained the model on the structure of the dense neural network using TF-IDF and the word bag and found the perfect similarity between the title and article pair. Their proposed model obtained the results TF-IDF on unigram and bigram with cosine similarity 94.31%, a bag of words with cosine similarity 89.23%.

In this research, we aggregated live news and related news and compared live news with the highest similarity to related news (two datasets were adopted), and what we have done is more accurate, as the news was compared with different sources and is similar live news.

In Bali et al. (2019) during this research, the researchers proposed a model for detecting fake news from the perspective of natural language processing and machine learning. An assessment known as cross-validation was performed on three data sets and compared the performance of seven machine learning algorithms in terms of accuracy and F1 scores.

The model was tested on three datasets: the open-source dataset, the Kaggle dataset for fake news, and the GitHub

repository for fake and real news. Next, previous datasets were preprocessed, removing symbols, links, and information unimportant in analyzing text features. To extract the features, it was used: n-gram: (use unigram, bigram, and trigram to count the number of occurrences), use TF-IDF: frequency of the term–frequency of the inverse document, frequency of the term is the number of times a word appears in a given document that can be calculated Equation 7:

$$tf(t, d) = (0.5 + 0.5)ft, d/\max ft', d : t' \in d \quad (7)$$

Equation 8 is used to calculate the frequency of the reverse document

$$idf(t, D) = \frac{\log N}{|d \in D : t \in d|} \quad (8)$$

Then, it calculates the cosine similarity of the standard vectors tf-idf for headers and texts Contents.

1. Include words: replace each word with a vector of real value.
2. Emotional Score: used to analyze sentiment in tone followed by different articles, using the Open-Source Library Natural Language Toolkit (NLTK); Emotional intensity was analyzed for positive, negative, neutral, and compound emotions.
3. Linguistic: The readability criteria are defined for news articles. The lexical diversity of articles is calculated and used as features. In this paper, 163 total features were used.

To evaluate the model, cross-validation was used on the data sets, and the model was evaluated on seven machine learning algorithms which are: Random Forest (RF), Support Vector Classifier (SVC), Gaussian Naïve Bayes (GNB), AdaBoost (AB), K-Nearest Neighbor (KNN), Multilayer Perceptron (MLP) and Gradient Enhancement (XGB). By comparing the accuracy with cross-validation and three datasets, the XGB algorithm was observed to outperform the XGB algorithm with an accuracy of 87.2% in the open-source data set

and 92.0% in the Kaggle, and 87.3% in the George McIntire dataset.

In this research, if we consider the similarity of textual content only (which is similar to what the researchers did), the highest similarity we got was 93% (46.5% * 2). The similarity values distinguish us, and what distinguishes us is the second criterion, which is the ranking of news websites sources globally, which gave additional accuracy to reveal the accuracy of the news.

In Al Asaad and Erascu (2018), the researcher uses the machine learning technique. Based on article content, an algorithm uses data mining to detect fake news. The algorithm extracts the text features (characteristics of content and publisher). Performs linguistic and visual studies on the extracted features (source, headlines, body text, image, video). The construct machine learning model will use external sources (e.g., readers of Facebook) by using the static dataset, not feedback, to train the model. Take an article from a news website, not social media site. They detect fake news based on the clickbait title system studies the relationship between the article title and the body. The used data set of fake and real news. <http://www.fakenewschallenge.org>. Use Scikit-learn library in python. Extract features from the dataset (using text representation models). Test two classification approaches on the title content. The researcher experts the result with the combined tools, and he gets a result greater than 0.8 for content and title classification. He observes that the linear classification model works best with TF-TDF the result was .094. Probabilistic classification gives low accuracy score when combined with TF-TDF. Both classifications give the same result for title 0.95. Bi-gram frequency model gives low accuracy for title classifications compared with Bag-of-words TF-TDF.

In Aldwairi and Alwahedi (2018b), the researcher locates the credible clickbait database. Then, compute the attributes and produce the data file for WEKA. They collect URLs for the clickbait from the web by focusing on the social media sites that have more fake news or clickbait ads or articles such as Facebook, Forex, and Reddit. After gathering URLs in a file, the researcher uses a python script to compute the attributes from the title and the content of the web pages. Then, they extracted the features from the web pages. They use cross-validation 10-fold in all experiments. After reading the website attributes file into WEKA. The researcher ranks the attributes based on several algorithms. To choose the most relevant and increase accuracy, decrease time. The researcher obtains classification results as Metrics and Classifier. The classification of the researcher is based on: Precision, Recall, f-measure, and ROC. The logistic classifier has the highest Precision with a result= of 99.4%. Logistic and random Tree classifiers had the best recall(sensitivity)=99.3%. BayesNet and Naivebytes are the best areas under the ROC curve.

In Thota et al. (2018b), the researcher uses: FNC-1 dataset (2 models) to detect fake news. The first model, Riedel et with 88% accuracy, with (lexical and similarity), passed through a multi-layer perception (MLP) with one hidden layer. The second model is called Davis et al. with 93% accuracy. Glove representation (GloVe)Word2Vec predictive model. These techniques are applied to both the headlines and website articles. He allocates 67% of FNC data to the train set and the remaining 33% to the test set. Training data are divided into validation set (80/20) splits. All experiments are conducted on training and validation setup. The researcher trained models on Dense neural networks (DNN).

Compared with the previous research, we detect fake news with the proposed fake news accuracy detection algorithm according to these sequences. Aggregate Live news, Collect websites ranking of news sources by(RankAPI and Alexa website), Collect related news from Google search engine by live news titles, execute news accuracy algorithm, Compute rank of source news website from total news accuracy 50%, apply(stop words, Tokenizing, Stemming) for both Text and Title of fake news dataset and Text and Title of living news, then implements cosine similarity for last Compute cosine similarity of living news text with related news text from total news accuracy 50% scores.

5 Conclusion and future works

With advanced technology and communication methods, the information spread among people without verification. The problem is that fake news plays a vital role in our life, that's why researchers started looking for a solution to stop fake news and disinformation from spreading widely. It is hard to control the flow of information online. We attempted to verify the news by implementing an algorithm by combining several methods to get satisfactory results. In this paper, we took a different approach to detect fake news as the first step. We collect live news from the Google search engine by live news title, then we rank news websites according to the results generated from RnkAPI and Alexa websites. Then, we execute algorithm programming language using python; the programmed tool in this paper depends on two primary sections. Detect news accuracy by ranking news and rating news accuracy by getting top1 cosine similarity from fake news dataset. After that, we applied News text similarity with fake news dataset, texts similarity implementation we applied text processing (uniform case of letters, remove punctuation and non-ASCII characters, Remove Stop-words, Text tokenization, Words-stemming) then we Created TF matrix for texts to compute current news accuracy (Compute rank of source news website from total news accuracy 50% + cosine similarity of living news 50% scores).

Funding This research received no external funding.

Data availability Data are available from the authors upon reasonable request.

Declarations

Conflict of Interest The authors declare that there is no conflict of interest regarding the publication of this paper.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Abualigah L, Diabat A (2022) Chaotic binary group search optimizer for feature selection. *Expert Syst Appl* 192:116368
- Abualigah L, Diabat A, Mirjalili S, Abd Elaziz M, Gandomi AH (2021) The arithmetic optimization algorithm. *Comput Methods Appl Mech Eng* 376:113609
- Abualigah L, Diabat A, Sumari P, Gandomi AH (2021b) Applications, deployments, and integration of internet of drones (iod): a review. *IEEE Sens J*
- Abualigah L, Yousri D, Abd Elaziz M, Ewees AA, Al-Qaness MA, Gandomi AH (2021) Aquila optimizer: a novel meta-heuristic optimization algorithm. *Comput Ind Eng* 157:107250
- Abualigah L, Abd Elaziz M, Sumari P, Geem ZW, Gandomi AH (2022) Reptile search algorithm (rsa): a nature-inspired meta-heuristic optimizer. *Expert Syst Appl* 191:116158
- Agushaka JO, Ezugwu AE, Abualigah L (2022) Dwarf mongoose optimization algorithm. *Comput Methods Appl Mech Eng* 391:114570
- Ahmed H, Traore I, Saad S (2017) Detection of online fake news using n-gram analysis and machine learning techniques. In: *International conference on intelligent, secure, and dependable systems in distributed and cloud environments*, Springer, pp 127–138
- Ahmed H, Traore I, Saad S (2018) Detecting opinion spams and fake news using text classification. *Secur Priv* 1(1):e9
- Al Asaad B, Erascu M (2018) A tool for fake news detection. In: *2018 20th international symposium on symbolic and numeric algorithms for scientific computing (SYNASC)*, IEEE, pp 379–386
- Al-Masalha H, Hnaif AA, Kanan T (2020) Cyber-crime effect on jordanian society. *Int J Advance Soft Compu Appl* 12(3)
- Al-Zu'bi S, Hawashin B, Mughaid A, Baker T (2021) Efficient 3d medical image segmentation algorithm over a secured multimedia network. *Multimed Tools Appl* 80(11):16887–16905
- Aldwairi M, Alwahedi A (2018) Detecting fake news in social media networks. *Procedia Comput Sci* 141:215–222
- Aldwairi M, Alwahedi A (2018) Detecting fake news in social media networks. *Procedia Comput Sci* 141:215–222
- Alkhair M, Meftouh K, Smaïli K, Othman N (2019) An arabic corpus of fake news: Collection, analysis and classification. In: *international conference on arabic language processing*, Springer, pp 292–302
- AlZu'bi S, Jararweh Y (2020) Data fusion in autonomous vehicles research, literature tracing from imaginary idea to smart surrounding community. In: *2020 Fifth international conference on fog and mobile edge computing (FMEC)*, IEEE, pp 306–311
- AlZu'bi S, Jararweh Y, Al-Zoubi H, Elbes M, Kanan T, Gupta B (2018) Multi-orientation geometric medical volumes segmentation using 3d multiresolution analysis. *Multimedia Tools and Applications* pp 1–26
- AlZu'bi S, Aqel D, Mughaid A, Jararweh Y (2019) A multi-levels geo-location based crawling method for social media platforms. *2019 Sixth international conference on social networks analysis, management and security (SNAMS)*, IEEE, pp 494–498
- Aphiwongsophon S, Chongstitvatana P (2018) Detecting fake news with machine learning method. *2018 15th international conference on electrical engineering/electronics, computer, telecommunications and information technology (ECTI-CON)*, IEEE, pp 528–531
- Attiya IA, Abd Elaziz M, Abualigah L, Nguyen TN, Abd El-Latif AA (2022) An improved hybrid swarm intelligence for scheduling iot application tasks in the cloud. *IEEE Trans Ind Inform*
- Bali APS, Fernandes M, Choubey S, Goel M (2019) Comparative performance of machine learning algorithms for fake news detection. In: *international conference on advances in computing and data sciences*, Springer, pp 420–430
- Batur Şahin C, Abualigah L (2021) A novel deep learning-based feature selection model for improving the static analysis of vulnerability detection. *Neural Comput Appl* 33(20):14049–14067
- Castillo C, Mendoza M, Poblete B (2011) Information credibility on twitter. In: *Proceedings of the 20th international conference on World wide web*, pp 675–684
- Choudhary A, Arora A (2021) Linguistic feature based learning model for fake news detection and classification. *Expert Syst Appl* 169:114171
- Duo M (2021) Alexa rank: everything you need to know about it. <https://kinsta.com/blog/alexa-rank/#what-is-alexa-rank/>
- Elbes M, Alrawashdeh T, Almaita E, AlZu'bi S, Jararweh Y (2020) A platform for power management based on indoor localization in smart buildings using long short-term neural networks. *Trans Emerg Telecommun Technol* 33(3):e3867
- Fayaz M, Khan A, Bilal M, Khan SU (2022) Machine learning for fake news classification with optimal feature selection. *Soft Comput* 1–9
- Freire PMS, da Silva FRM, Goldschmidt RR (2021) Fake news detection based on explicit and implicit signals of a hybrid crowd: an approach inspired in meta-learning. *Expert Syst Appl* 183:115414
- Gandomi AH, Chen F, Abualigah L (2022) Machine learning technologies for big data analytics. *Electronics* 11(3):421
- Goldani MH, Safabakhsh R, Momtazi S (2021) Convolutional neural network with margin loss for fake news detection. *Inf Process Manag* 58(1):102418
- Guo Z, Shen Y, Aloqaily M, Jararweh Y, Yu K (2021) Probabilistic inference-based modeling for sustainable environmental systems under hybrid cloud infrastructure. *Simul Model Pract Theory* 107:102215
- Hanselowski A, PVS A, Schiller B, Caspelherr F, Chaudhuri D, Meyer CM, Gurevych I (2018) A retrospective analysis of the fake news challenge stance detection task. *arXiv preprint arXiv:1806.05180*
- kaggle (2021a) Fake and real news dataset. <https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset>
- kaggle (2021b) Fake news, build a system to identify unreliable news articles. <https://www.kaggle.com/c/fake-news/data>
- Kaliyar RK, Goswami A, Narang P (2021) Fakebert: fake news detection in social media with a bert-based deep learning approach. *Multimed Tools Appl* 80(8):11765–11788
- Klein DO, Wueller JR (2018) Fake news: A legal perspective. *Australas Polic* 10(2)
- Monteiro RA, Santos RL, Pardo TA, De Almeida TA, Ruiz EE, Vale OA (2018) Contributions to the study of fake news in portuguese: New corpus and automatic detection results. In: *international conference on computational processing of the portuguese language*, Springer, pp 324–334
- Nguyen HV, Bai L (2010) Cosine similarity metric learning for face verification. In: *Asian conference on computer vision*, Springer, pp 709–720
- Otaïr M, Ibrahim OT, Abualigah L, Altalhi M, Sumari P (2022) An enhanced grey wolf optimizer based particle swarm optimizer for

- intrusion detection system in wireless sensor networks. *Wirel Netw* pp 1–24
- Oyelade ON, Ezugwu AE, Mohamed TI, Abualigah L (2022) Ebola optimization search algorithm: a new nature-inspired metaheuristic algorithm. *IEEE Access*
- Paka WS, Bansal R, Kaushik A, Sengupta S, Chakraborty T (2021) Cross-sean: a cross-stitch semi-supervised neural attention model for covid-19 fake news detection. *Appl Soft Comput* 107:10733
- Pérez-Rosas V, Kleinberg B, Lefevre A, Mihalcea R (2017a) Automatic detection of fake news. arXiv preprint [arXiv:1708.07104](https://arxiv.org/abs/1708.07104)
- Pérez-Rosas V, Kleinberg B, Lefevre A, Mihalcea R (2017b) Automatic detection of fake news. arXiv preprint [arXiv:1708.07104](https://arxiv.org/abs/1708.07104)
- Qazvinian V, Rosengren E, Radev D, Mei Q (2011) Rumor has it: identifying misinformation in microblogs. In: *Proceedings of the 2011 conference on empirical methods in natural language processing*, pp 1589–1599
- Rankapi (2021) Rank api and checker. <https://www.rankapi.net/>
- Reis JC, Correia A, Murai F, Veloso A, Benevenuto F (2019) Supervised learning for fake news detection. *IEEE Intell Syst* 34(2):76–81
- Rubin VL, Conroy N, Chen Y, Cornwell S (2016) Fake news or truth? using satirical cues to detect potentially misleading news. In: *Proceedings of the second workshop on computational approaches to deception detection*, pp 7–17
- Sadiku M, Eze T, Musa S (2018) Fake news and misinformation. *Int J Adv Sci Res Eng* 4(5):187–190
- Şahin CB, Dinler ÖB, Abualigah L (2021) Prediction of software vulnerability based deep symbiotic genetic algorithms: phenotyping of dominant-features. *Appl Intell* 51(11):8271–8287
- Saikh T, De A, Ekbal A, Bhattacharyya P (2020) A deep learning approach for automatic detection of fake news. arXiv preprint [arXiv:2005.04938](https://arxiv.org/abs/2005.04938)
- Sivasankari S, Vadivu G (2021) Tracing the fake news propagation path using social network analysis. *Soft Comput* pp 1–9
- Spradling M, Straub J, Strong J (2021) Protection from ‘fake news’: the need for descriptive factual labeling for online content. *Future Internet* 13(6):142
- Thorne J, Vlachos A, Christodoulopoulos C, Mittal A (2018) Fever: a large-scale dataset for fact extraction and verification. arXiv preprint [arXiv:1803.05355](https://arxiv.org/abs/1803.05355)
- Thota A, Tilak P, Ahluwalia S, Lohia N (2018) Fake news detection: a deep learning approach. *SMU Data Sci Rev* 1(3):10
- Thota A, Tilak P, Ahluwalia S, Lohia N (2018) Fake news detection: a deep learning approach. *SMU Data Sci Rev* 1(3):10
- Vlachos A, Riedel S (2014) Fact checking: task definition and dataset construction. In: *Proceedings of the ACL 2014 workshop on language technologies and computational social science*, pp 18–22
- Wu D, Jia H, Abualigah L, Xing Z, Zheng R, Wang H, Altalhi M (2022) Enhance teaching-learning-based optimization for tsallis-entropy-based feature selection classification approach. *Processes* 10(2):360

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.