Neural Computing and Applications (2023) 35:5999–6013
https://doi.org/10.1007/s00521-022-07938-3

ORIGINAL ARTICLE

Framework for detection of probable clues to predict misleading information proliferated during COVID-19 outbreak

Deepika Varshney[1] · Dinesh Kumar Vishwakarma[1]

Received: 11 August 2021 / Accepted: 11 October 2022 / Published online: 13 November 2022

Abstract

Spreading of misleading information on social web platforms has fuelled huge panic and confusion among the public regarding the Corona disease, the detection of which is of paramount importance. To identify the credibility of the posted claim, we have analyzed possible evidence from the news articles in the google search results. This paper proposes an intelligent and expert strategy to gather important clues from the top 10 google search results related to the claim. The N-gram, Levenshtein Distance, and Word-Similarity-based features are used to identify the clues from the news article that can automatically warn users against spreading false news if no significant supportive clues are identified concerning that claim. The complete process is done in four steps, wherein the first step we build a query from the posted claim received in the form of text or text additive images which further goes as an input to the search query phase, where the top 10 google results are processed. In the third step, the important clues are extracted from titles of the top 10 news articles. Lastly, useful pieces of evidence are extracted from the content of each news article. All the useful clues with respect to N-gram, Levenshtein Distance, and Word Similarity are finally fed into the machine learning model for classification and to evaluate its performances. It has been observed that our proposed intelligent strategy gives promising experimental results and is quite effective in predicting misleading information. The proposed work provides practical implications for the policymakers and health practitioners that could be useful in protecting the world from misleading information proliferation during this pandemic.

Keywords COVID-19 · Information pollution · Fake news detection

1 Introduction

In the recent scenario, a new coronavirus disease spread around the world. The disease emerges as a respiratory infection with significant concern for global public health hazards. Initially, it is suspected that the disease is transmitted from animal to humans, later the paradigm is shifted that the infection is transmitted toward human to human via droplets, close contacts creating huge panic with approximate 6,359,182 confirmed cases and 380,663 deaths[1] have been encountered till now and growth rate is still high which has alarmed the global authorities including world health organization (WHO) [1]. The COVID-19 pandemic affects worldwide badly. However, there is no shortage of people who are taking this crisis as an opportunity for malicious activities/gaining profit. Many health-related misleading information some of the fake cures suggested for COVID-19 have been posted by malicious users, creating confusion and misconceptions about the disease. During this pandemic, people have their eye on any new announcement from the government official or some news that can help to get rid of COVID-19. As the disease is deadly, the people are also desperate to know some cure and rush to find a new coronavirus disease. Some of the fake cures posted on social media are harmful and give bad health advice. The recent examples of fake cures are shown

✉ Dinesh Kumar Vishwakarma
dinesh@dtu.ac.in

Deepika Varshney
deepikavarshney06@gmail.com

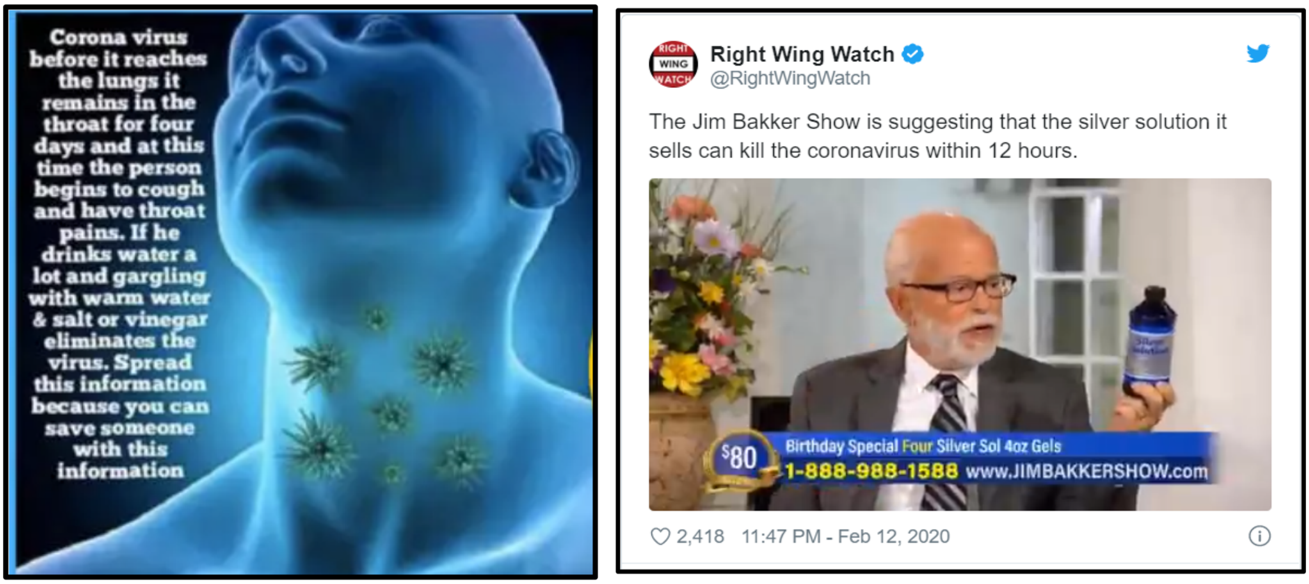
[1] Biometric Research Laboratory, Department of Information Technology, Delhi Technological University, Delhi 110042, India

[1] https://in.search.yahoo.com/search?fr=mcafee&type=E211IN885G0&p=coronavirus.

in Fig. 1. where Fig. 1a shows the image with a false claim is gone viral that drinking water a lot and gargling with warm water and salt or vinegar eliminates the coronavirus, however, there is no significant evidence has been found concerning to this claim. Another cure in Fig. 1b reported that the silver solution can kill coronavirus within 12 h. The proliferation of these misleading information creates many misconceptions in the mind-set of people related to coronavirus disease. Some of the users are spreading it without verification and fuelled panic among people regarding the COVID-19. According to [2], misleading or fake can be defined as any post that shares content that does not faithfully represent the event that it refers to. We followed this definition in our work and define *"misleading information as the content that does not faithfully represent the event that it refers to and having no significant evidence of proof to validate the claim."* Recent research shows that numerous misleading content is circulating about the coronavirus and it is becoming difficult to differentiate fake news from the real one [3]. The propagation of misleading content on the virus could also be deleterious to mankind. This has led to the dire need for a system that can differentiate fake from real. Earlier, many of the previous research has reported methods of detecting fake news in online social media considering various applications. Most of the previous research has counter fake news problems mainly in the following types: Image-based and Text-based algorithms. Earlier, many researchers have worked on fake news detection by applying a text-based approach. Text-based approaches mainly use text patterns and match them with already existing patterns of fake news. They are sometimes referred to as the linguistic approach. Along with this lots of researchers have shifted their interest in the credibility detection of posts/tweets using text-based features [4]. Like a text-based approach, research has also been done by employing an image-based approach. From the study, it has been seen that researchers have explored images based algorithms for the analysis of fake images or images attached with false claims in mainly following ways, Text additive images, and Manipulated images. *The manipulated images are termed as images whose piece/part or certain region is manipulated with respect to visual context.* Various image-based features have been explored for the classification of images. The authors of [5] propose 5 visual features and 7 statistical features for the verification of news events. Along with the manipulated images, some researchers have also considered text additive images to analyze misleading content. *The text additive images termed as images embedded with false claims instead of having any manipulation from visual context.* The authors of [6] have incorporated text additive images, where they have applied a rule-based algorithm for the prediction of fake news. From recent research, it has been observed that none of the works have shown and reported fake news prediction analysis propagated during one of the major pandemic *CORONAVIRUS*. Many people are sharing fake cures to get rid of coronavirus disease without any verification and create lots of misconceptions. Government and officials have also urged peoples to check the authenticity of the post before sharing [3]. This also motivates us to build an intelligent system for the prediction of fake news spreading during this pandemic. We, therefore, developed a generalized model of detecting misleading content on social media platforms, where we have considered COVID-19 as a special issue which is a huge pandemic and taken as one of the application case studies in this work. However, our model is generalized and works for other applications as well. COVID-19 is an emerging issue and very few research have been reported yet in this context that leads to motivates us to build an efficient framework to predict misleading content spreading during the COVID outbreak. One of the major lacks in this case that there is no dataset concerning corona fake news that is publically available to test our proposed algorithm. One of the novelty and contribution of the work is we have built a dataset with 335 claim samples having 168 fake and 167 real news related to COVID-19, also working to expand the dataset. Along with this to validate the generalizability of our model, it has also been tested over one of the standard dataset "LIAR" that further validates the effectiveness of the model. Three sets of novel features (N-gram, Levenshtein Distance, and Word Similarity) are proposed in this work that is found to be efficient in predicting misleading content and results reveal that the model outperforms the other state-of-the-art method, discussed in detail in the later sections. To the best of our knowledge, none of the proposed work has been employed these features concerning news article headlines and its content. The major key contributions of the work are highlighted in the following points.[2]

- The proposed work gives a significant contribution in providing a novel generalized framework for the collection and annotation of misleading content in an online social network where considering COVID-19 (fake news spreading during Corona outbreak) as one of the special case studies from the application perspective, where considered post in the form of images and text related to unverified cures to get rid out from the disease.
- For the analysis perspective, the data are collected related to Covid-19 and builds a self-generated dataset *"Covidfakenews 2019"* of misleading post spreading during corona. As the problem is new/emerging, no

[2] https://www.buzzfeednews.com/article/janelytvynenko/coronavirus-fake-news-disinformation-rumors-hoaxes.



Corona virus before it reaches the lungs it remains in the throat for four days and at this time the person begins to cough and have throat pains. If he drinks water a lot and gargling with warm water & salt or vinegar eliminates the virus. Spread this information because you can save someone with this information

Right Wing Watch
@RightWingWatch

The Jim Bakker Show is suggesting that the silver solution it sells can kill the coronavirus within 12 hours.

$80
Birthday Special Four Silver Sol 4oz Gels
1-888-988-1588 www.JIMBAKKERSHOW.com

♡ 2,418 11:47 PM - Feb 12, 2020

(a)

(b)

Fig. 1 Examples of misinformation

datasets are publically available for analysis related to this area of research. This leads to giving a significant contribution.

- To the best of our knowledge, the proposed work is first to provide three sets of novel features (N-gram, Levenshtein Distance, and Word Similarity) concerning to news article headlines and its content from the google search results, that further be used as an efficient clue for the prediction of misleading content.
- As the COVID-19 is one of the emerging issues, none of the work has been reported yet to predict the fake news propagating during this phase and gives a major contribution by providing the analysis, which greatly helps researchers for further study.
- We investigate the model performances with different classifiers and comparative analysis reveals that our proposed method outperforms other states-of-the-art on the same dataset.

The remaining of this paper is organized as follows. In Sect. 2, we are going to discuss the previous work that has been done related to this field, where Sect. 3, gives a detailed description of how the data collection and annotation have been employed as well as the strategy/method that we have employed for the misleading information detection, which followed by a discussion of experimental results in Sect. 4. Lastly, the paper is concluded with some suggested future work aspects.

2 Related work

In the current era spreading false information is one of the crucial problems nowadays, where it is quite difficult for the online user to discriminate fake news from the real one and that's why the development of an intelligent system is required. Most of the methods proposed in earlier states to detect misleading information considered it a task of classification problem intending to associate label as true or false with a particular claim/post. From the survey analysis, it has been observed that the classification approaches are turn divided into approaches based on machine learning and deep learning. A detailed description is given below.

2.1 Machine learning

From the study, it has been proven that machine learning algorithms are extremely useful in countering numerous problems in the information engineering field. In particular, many machine learning approaches implemented for misleading/fake information detection are applied as a supervised learning strategy. In machine learning classification algorithms support vector machines(SVMs) are one of the widely used methods for classifications. The authors of [7] have proposed a method to employ a graph-kernel-based SVM classifier to detect rumors using propagation structure and content features with an accuracy of 0.91 on the Sina-Weibo dataset. Whereas, in [8] the author reported a set of features to distinguish among fake news, real news, and satire. The SVMs are also employed for clickbait detection in [9]. The random forest has also been exploited in numerous works for fake news and rumor detection like SVM. Most of the studies have reported random forests as a strong performer among other machine learning algorithms [10–14]. In [11], the author has proposed a set of temporal, structural, and linguistic features for the classification of rumors in a tweet graph by employing a random forest with an accuracy of 0.90. The random forest has also been used for stance detection in [10, 15]. The comparative studies of a different approach in the context of rumor and fake news have shown competitive performance for logistic regression [12, 16–18]. The authors of [19], employed logistic regression for stance classification of news articles or headlines and claims. A decision tree is another widely studied algorithm proposed particularly for misleading content detection [20]. The effectiveness of decision tree algorithms like j48 with respect to other machine learning paradigms, including SVMs has been reported in [4, 12]. The authors of [4], have used the content and context-based features to perform credibility evaluation of tweets and the model is performing well with an accuracy of 0.86. In [21], to evaluate the trustworthiness of users in social media via decision tree, the author has proposed a series of user trust metrics and reported an accuracy of 0.75.

2.2 Deep learning

Deep learning is one of the prominent and widely explored research topics in machine learning. The main advantage of deep learning over traditional machine learning approaches is that they are not based on manually crafted features and reduce feature extraction time. Along with this, the deep learning framework can learn hidden representations from simpler inputs both in context and content variation [22]. The two prominent and widely used paradigms in Morden artificial neural network are RNN and CNN. In [22], authors have proposed novel RNN architectures, namely tanh-RNN, LSTM, and Gated Recurrent Unit(GRU), to detect rumors. The results show that GRU has obtained the best results in both the datasets considered with 0.88 and 0.91 accuracies, respectively. Whereas, in [23], the author has proposed a multi-task learning approach and designed a multi-task learning framework with an LSTM layer shared among all tasks to counter the problem of rumor classification. Like RNN, CNNs have also been explored and widely studied for image recognition and many other fields of computer vision. However, it is now gaining popularity in the NLP field [24]. The authors of [15], have explored a technique using CNN with single and multi-word embedding to counter problems concerning both stance and veracity classification of tweets. The author has reported an accuracy of 0.70 for the stance classification problem and 0.53 for the problem of veracity classification. Whereas, Paragraph embedding is explored to learn the representation of a small group of posts in a specific event and used them as input for their CNN model in [25] and achieved an accuracy of 0.93 for Sina Weibo and 0.77 for Twitter. From the study, it has been observed that most of the recent work has explored the combination of RNN and CNN in their model [26–28]. The authors of [28] proposed an architecture applied on the LIAR dataset that encodes text information via a CNN and metadata about the author of the text using an LSTM layer as well as it has been also found that the hybrid model has proved to outperform all other baselines along with a bi-LSTM architecture with an accuracy of 0.27 on the testing dataset. Whereas, in [26] author has proposed an approach based on repost sequence patterns for the detection of false rumors.

All the different approaches discussed above have considered different machine and deep learning methods for the prediction of misleading content. The previous study shows that very few works have reported the important clues concerning news article headlines and news article content to predict misleading information. The study has observed that crucial pieces of evidence can be identified from News article headlines and its content to predict false information. None of the previous work has been included these three sets of novel clues (N-gram, Levenshtein Distance, and Word Similarity) concerning news articles that were found to be efficient. That gives novelty to our proposed work.

3 Methodology

In this section, we elaborate on each phase of the proposed methodology. We first detail the data creation step adopted for evaluating our model. The self-generated dataset is the collection of claims in the form of images and text posted by users related to coronavirus. Here, images we mean images embedded with text. In the next phase, the given claim has been processed with the removal of stop words, punctuation, and unnecessary words for query building. The query is then given as an input to google search, where the top 10 search results are considered for analysis. To maintain the claim's relevance and retrieve related results, the top 10 links have been used for processing. In the next phase, important clues are identified from both news article headlines as well as news article content concerning Word Similarity, N-gram, Levenshtein Distance Ratio, and N-gram, Levenshtein Distance Ratio measures, respectively. Lastly, these important clues are used for binary class classification real or misleading.

3.1 Problem definition

In this paper, we have considered a binary class classification problem. We assume that the posted source claim $c = \{c1, c2, c3...ck\}$ can be divided into two classes $\text{Class} = \{M, R\}$.: (1) Real (R), namely the posted claim faithfully represents the event that it refers to, (2) Misleading (M), namely the posted claim does not faithfully represent the event that it refers to. The claims that have been included here are related to coronavirus COVID-19. Take an example of the post related to coronavirus. The post that "chlorine and alcohol products cannot kill viruses within the body" is true information and belongs to the real class since many authoritative and authentic news media have reported relevant news and has also acknowledged by WHO.[3] While one of the posts says that "coronavirus is caused by 5G technology" is absolutely a false rumor, it not only lacks the factual sport but also deviates from the scientific principles. So, the goal is to learn a classifier from the labeled feature set, that is $f : X^k \mapsto Y^k$, where Y^k takes one of the two fine-grained classes: {R, M}. Given the input feature set X^k, the classifier f can output the classification result for the posted claim C^k.

3.2 Dataset creation and query searching

In the coronavirus outbreak, lots of misleading information was posted by users, and most of the misleading information reading fake cures, lockdown, and others have gone viral during the period. The posts have gone viral on different social media platforms like Twitter, Facebook, and others. The dataset is created during the covid-pandemic 2020. The dataset is the collection of post available in the form of text embedded images and text-only form. Here, text embedded images we mean that the claim is inbuilt in an image. Sometimes to represent certain event users post an image embedded with text. The example of text embedded images we can also find from Fig. 1. For building the dataset, the prominent claims are collected related to corona posted during the period from different web sources (News Media, WHO, Buzzfednews) in the form of images and text. Some of the sources of misleading news are Buzzfeed news,[4] Buzzfeednews provides many coronavirus hoaxes that are posted over various social web platform in different forms(images, audio, text), along with it they also provide veracity, mentioning whether the given claim is fake or not, another source that we have incorporated in this work is WHO (world health organization), the website is providing fake and real claims, the another web sources that are reporting myths regarding coronavirus.[5] The empirical analysis has been done on these web sources for the collection of fake samples (Fig. 2). The sources provide the running list of the latest hoaxes spreading about the coronavirus. Whereas most of the real claims are retrieved from the WHO website for more reliable collection. In all total of 335 queries are collected with having 168 misleading and 167 real claims (*Covid-fakenews 2019*). As there is no source providing clear-cut veracity until and unless we read the complete text of it. That's why after empirical analysis we prepare our ground truth dataset consisting of 335 samples, including fake and real as a target class. The world cloud representation of fake and real claims in the proposed dataset is shown in Fig. 3a and b, respectively.

There are 9837 fake samples and 8634 real samples are considered with respect to 335 claims. The text additive images are processed further using the OCR technique, where textual data has been extracted to build the query for further processing shown in Fig. 2. In Fig. 2, the input is taken in the form of either plain text or the form of text additive image. If it is a simple text, then it is directly passed to the query building phase; otherwise, if it is in the form of text additive image, the OCR technique has been applied to extract the text from an image, and further, it goes further to the query building phase. The query is searched on google, where top 10 google results $L_i = L_{1,} L_2 ... L_{10}$ are retrieved and used for further processing as shown in Fig. 2. Each link is passed to the processing phase, where important clues are identified for prediction.

3.3 Clues identification for misleading information prediction

In this phase, important clues are identified from the retrieved top 10 URLs that further be used to predict the misleading information. The pieces of evidence/clues are identified with respect to news article headlines/titles and news article content/text. In this phase, the clues are identified in two steps. Firstly, the titles are analyzed and important features are extracted with respect to Word Similarity Ratio, N-grams, and Levenshtein Distance Ratio. The LDR[6] and N-gram[7] are the earlier techniques that helps in reteriving efficient features in our work. LDR (Levenstein distance ratio) is one of the good method of reteriving similarity among two sentences. Similarly, N-Gram is also useful in finding important phrases in the

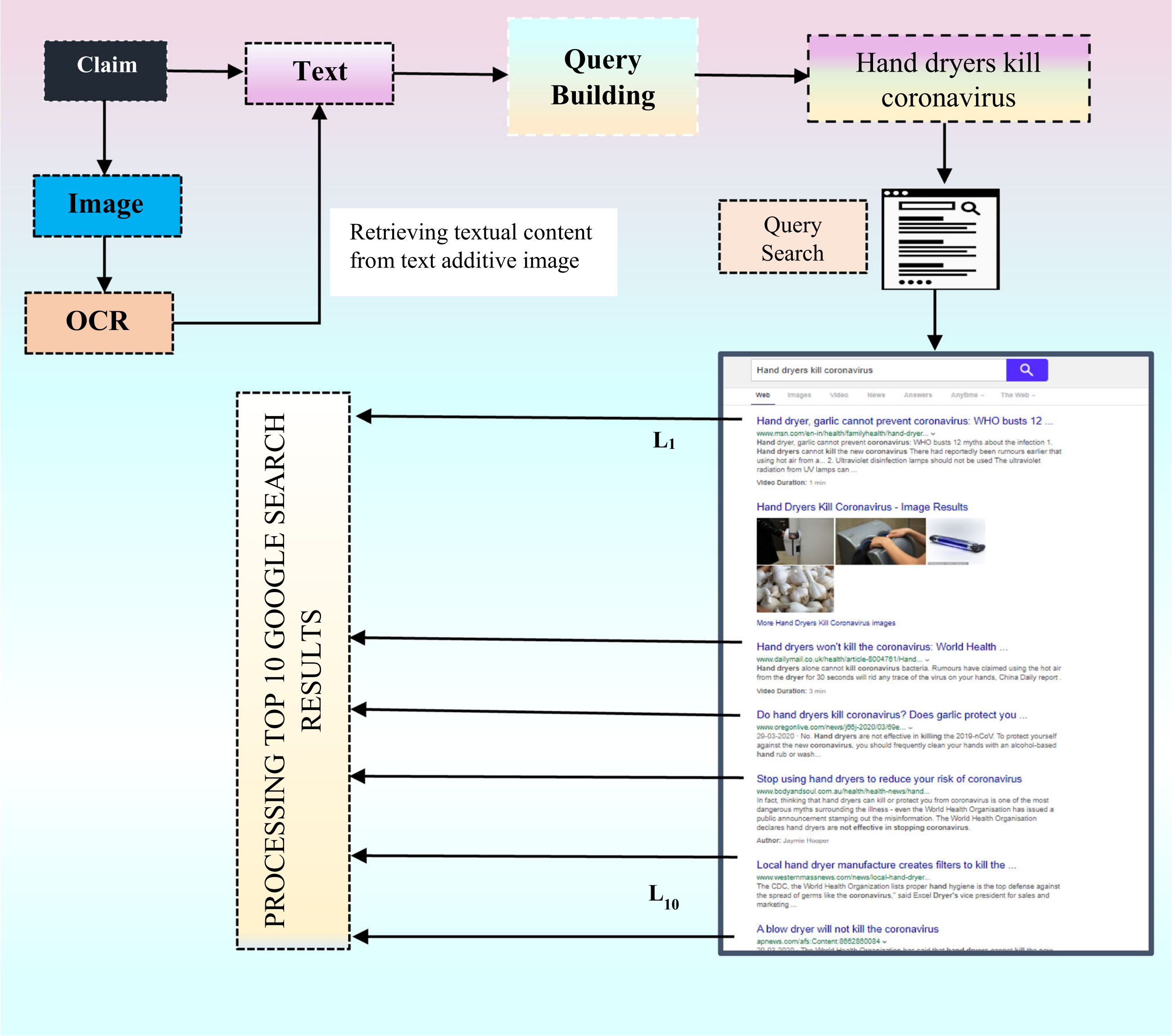
[3] https://www.indiatoday.in/world/story/drinking-alcohol-will-not-protect-you-from-covid-19-says-who-1653555-2020-03-08.

[4] https://www.buzzfednews.com/article/janelytvynenko/coronavirus-fake-news-disinformation-rumors-hoaxes.

[5] https://www.medicalnewstoday.com/articles/coronavirus-myths-explored.

[6] Levenshtein distance—Wikipedia.

[7] What are N-Grams?—Kavita Ganesan, PhD (kavita-ganesan.com).


Claim
Text
Query Building
Hand dryers kill coronavirus
Image
Retrieving textual content from text additive image
Query Search
OCR
PROCESSING TOP 10 GOOGLE SEARCH RESULTS
L1
L10
Web
Images
Video
News
Answers
Anytime
The Web

Hand dryers kill coronavirus

Hand dryer, garlic cannot prevent coronavirus: WHO busts 12 ...
www.msn.com/en-in/health/familyhealth/hand-dryer...
Hand dryer, garlic cannot prevent **coronavirus**: WHO busts 12 myths about the infection 1. **Hand dryers** cannot **kill** the new **coronavirus** There had reportedly been rumours earlier that using hot air from a... 2. Ultraviolet disinfection lamps should not be used The ultraviolet radiation from UV lamps can ...
Video Duration: 1 min

Hand Dryers Kill Coronavirus - Image Results
More Hand Dryers Kill Coronavirus images

Hand dryers won't kill the coronavirus: World Health ...
www.dailymail.co.uk/health/article-8004761/Hand...
Hand dryers alone cannot **kill coronavirus** bacteria. Rumours have claimed using the hot air from the **dryer** for 30 seconds will rid any trace of the virus on your hands, China Daily report .
Video Duration: 3 min

Do hand dryers kill coronavirus? Does garlic protect you ...
www.oregonlive.com/news/j68j-2020/03/69e...
29-03-2020 · No. **Hand dryers** are not effective in **killing** the 2019-nCoV. To protect yourself against the new **coronavirus**, you should frequently clean your hands with an alcohol-based **hand** rub or wash...

Stop using hand dryers to reduce your risk of coronavirus
www.bodyandsoul.com.au/health/health-news/hand...
In fact, thinking that hand dryers can kill or protect you from coronavirus is one of the most dangerous myths surrounding the illness - even the World Health Organisation has issued a public announcement stamping out the misinformation. The World Health Organisation declares hand dryers are **not effective in stopping coronavirus**.
Author: Jaymie Hooper

Local hand dryer manufacture creates filters to kill the ...
www.westernmassnews.com/news/local-hand-dryer...
The CDC, the World Health Organization lists proper **hand** hygiene is the top defense against the spread of germs like the **coronavirus**," said Excel **Dryer's** vice president for sales and marketing ...

A blow dryer will not kill the coronavirus
apnews.com/afs:Content:8662860084
29-03-2020 · The World Health Organization has said that **hand dryers cannot kill the new**

Fig. 3 Example of building query and search Process

sentences concerning false content. Some of the earlier studies have also mentioned that the N-gram method contributes to finding important clues from the news article title [2, 3]. We have created a corpus of most prominently used words to represent false information in our work. The N-Gram technique has been used to extract features (unigram, bigram, and trigram) concerning to each paragraphs. In the same way LDR method is one of the text-similarity measures between words, this gives a ratio indicating how similar the two sentences. Earlier studies have employed many different methods for text similarity (Cosine, LDR, Jaccard, etc.). The LDR is a good method for finding out or measuring the distance between sequence of characters.

The detailed description of the complete process is shown in the following steps.

3.3.1 Clues identification from news article headlines/titles

The effective clues are retrieved from news article title/headlines based on three proposed measures (Word Similarity Ratio, Levenshtein Distance Ratio, and N-Grams). A detailed description of each of these measures have shown below:

3.3.1.1 Word similarity ratio (WSR) This is the first measure considered for retrieving some important clues from the news headline. This is the ratio of the number of keywords (K_i) that are found to be similar in *news title* and the *query* to the total number of words in the query (t_i). The Ratio helps in identifying useful or relevant titles. The

(a)

(b)

Fig. 2 Word cloud representation for fake and real claims in the proposed dataset

more the word similarity ratio, the more relevant the title to the query.

$$\mathrm{WSR} = \frac{K_i}{t_i} \tag{1}$$

3.3.1.2 Levenshtein distance ratio (LDR) Levenshtein Distance Ratio has been used to find the similarity between the query and the news article headline/title. To calculate the ratio, Levenshtein distance needs to be calculated. Levenshtein distance is a metric to measure how apart are two sequences of words or in another way can say it is a measure of the minimum number of edits that you need to do to change a one-word sequence into the other.

The formal definition of the Levenshtein distance between two string a and b can be seen as follows:

$$\mathrm{lev}_{(a,b)}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j)=0 \\ \min\begin{cases} \mathrm{lev}_{a,b}(i-1,j)+1 \\ \mathrm{lev}_{a,b}(i,j-1)+1 \\ \mathrm{lev}_{a,b}(i-1,j-1)+1_{(ai\neq bj)} \end{cases} & \text{otherwise} \end{cases} \tag{2}$$

where, $1_{(ai\neq bj)}$ is an indicator function equals to 0 when $a_i = b_i$ and 1 otherwise. The Levenshtein distance ratio can be retrieved based on the Levenshtein distance and can be calculated using the following formula:

$$\mathrm{LDR} = \frac{(|a| + |b|) - \mathrm{lev}_{(a,b)}(i,j)}{|a| + |b|} \tag{3}$$

where, $|a|$ and $|b|$ are the length of sequence a and sequence b, respectively. Here a is query and b is news article headline.

The main purpose of including Levenshtein distance is for string matching, and it is helpful to find matches for short strings in many longer texts.[8] In our case, it will be helpful, as the string may not be of the same length. The length of the title and the length of the query may not be the same and quite effective than other distance measures.

3.3.1.3 N-grams N-grams is an ordered n-tuple of characters. N-gram also contributes to finding important clues from the news article title [29]. Unigram takes a sentence and gives all the words in that we fence. Whereas, Bigram takes a sentence and gives us a set of two consecutive words in the sentences. A trigram gives sets of three consecutive words in a sentence. We manually analyzed titles used to represent and find a prominent keyword used in false query search responses and build a *false phrases dictionary* with 20 keywords. Some examples are (*"misleading," "misinformation," "is it," "is it true," "not known," "no proof," "no known," "no scientific evidence," "no evidence," "not verified," "hoax," "clickbait," "not proven," "denied," "deny," "unverified," "false," "fake," "fake news," "falsely," 'myth," "ridiculous," "rumour"*).

The word cloud representation of the respective keyword is shown in Fig. 4. The thing to be noted here, the keywords considered here are mainly focused on the news reported related to coronavirus. To extract all possible combinations, we have considered three types of n-gram (Unigram, Bigram, Trigram) to reduce the chance of missing any possible case. The count of unigram, bigram, and trigram is used as a clue, which reveals the number of titles reporting false keyword phrases. The detailed

[8] https://en.wikipedia.org/wiki/Levenshtein_distance.

Fig. 4 Word cloud representing key phrases used for false information prediction

description of features belonging to each category concerning to News article headline/ title and News article content is shown in Table 1.

3.3.2 Clues identification from news article content

As we have identified the clues in the previous section related to the top 10 news headlines/titles respective to each query. In the same way, the analysis has also been applied over new article content. New article content is one of the important parts that reveals what the article trying to convey, and whether the given content reporting the same thing as it claiming for. Analyzing news article content is needful to extract possible clues to predict whether the query is misleading or not. The article is quite long, analyzing it at once, and summarization is also not quite effective in getting an overview. To resolve this problem instead of analyzing the whole article at once or summarizing it, we extracted each article's first 8 paragraph (< p >) tags. The reason of considering first 8 < p > tags as from the analysis it has been found that most of the good articles covering the information within 8 paragraphs, as well as from the manual inspection it has also been observed that the 8 paragraphs are sufficient in getting the overview about an article. As discussed previously, each paragraph is inspected concerning two clue measures Levenshtein Distance Ratio and N-Grams. Both measures are applied in the same way but the context is changed, instead of applying it into news headline/title, news articles paragraphs are considered for fetching the possible clues. For each title, 8 paragraphs are employed to examine with respect to two clue measures as shown below:

3.3.2.1 Levenshtein distance ratio As we have previously discussed the Levenshtein distance ratio, the same is applied here. The LDR is calculated between the query and each paragraph text of the news article to inspect the similarity among them. This can be defined as for each query Q there are n paragraphs where the range is from $i = 1$ to $i = 8$ and the Levenshtein distance ratio is calculated between Q and n_i.

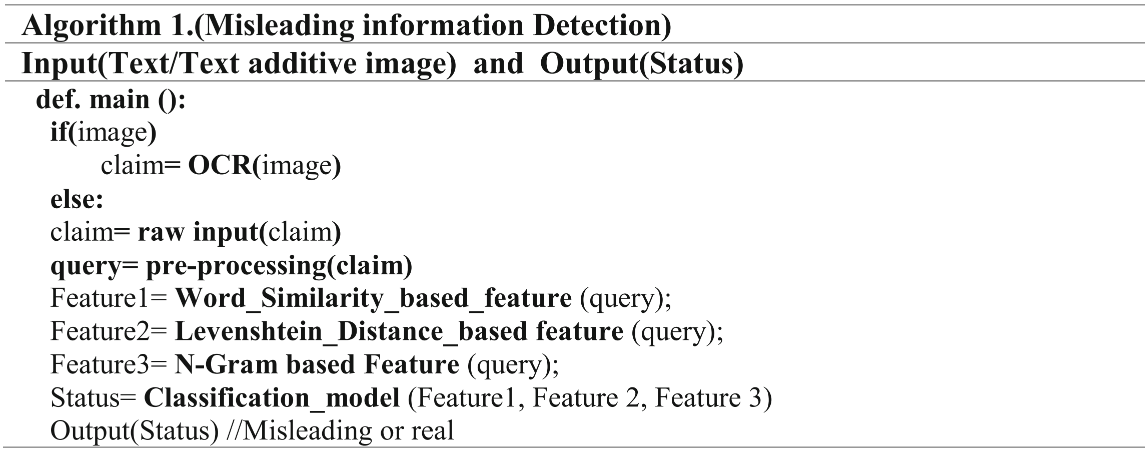
3.3.2.2 N- grams The N-gram measures are used in the same way as applied earlier. The n-grams are employed over each paragraph text The count of false keyword phrases is identified considering all three types 1-gram, 2-gram, and 3 grams that further be used to predict misleading information. For each $\mathrm{n_i}$ (n paragraphs), where the range is from $i = 1$ *to* $i = 8$, and for each paragraph, the count of keywords belonging to *false phrases dictionary* are calculated.

The proposed algorithm for the prediction of misleading content is given below in Algorithm 1: The proposed algorithm firstly checks for the user input whether it is an image or text. If it is an image embedded with text, then it first processed through the OCR algorithm, and the textual part is extracted which further be pre-processed to build the query for search. In the pre-processing part, stop words and punctuations have been removed. On the other hand, if the given input is text, it will go directly through the processing phase and the query building phase. The important clues for each query have been found out with respect to three feature sets based on word similarity, Levenshtein distance, and N-Gram. Finally, all the features are fed into the classification model for binary class classification and get the status (Misleading or real). Along with this, the

Table 1 Description of a set of features employed for misleading content detection

Feature category	Feature	Feature description	Clue identification type
Word similarity based	Word matching count (WMC)	The total number of words in the title/heading that are found to be similar and matching the words in the query	News Article Headline
Word similarity based	Query length (QL)	The word count of a query	News Article Headline
Word similarity based	Word matching ratio (WMR)	It is the ratio of word matching count and Query length	News Article Headline
Levenshtein Distance based	Title-query ratio	It is the Levenshtein distance ratio with respect to the title and the query	News Article Headline
N-gram based	1 Gram	Count of words(unigrams) found in the title/headline that belong to the false phrase dictionary	News Article Headline
N-gram based	2 Gram	Count of two consecutive words(Bigrams) found in the title/headline that belongs to the false phrase dictionary	News Article Headline
N-gram based	3 Gram	Count of three words(trigrams) found in the title/headline that belong to the false phrase dictionary	News Article Headline
Levenshtein distance based	Text-Query Ratio	It is the Levenshtein distance ratio with respect to the news article paragraph and the query	News Article Content
N-gram based	1 Gram(article paragraph)	Count of words(unigrams) found in each n article paragraph that belong to the false phrase dictionary	News Article Content
N-gram based	2 Gram(article paragraph)	Count of two consecutive words(Bigrams) found in each n article paragraph that belongs to the false phrase dictionary	News Article Content
N-gram based	3 Gram(article paragraph)	Count of three consecutive words(trigrams) found in each n article paragraph that belong to the false phrase dictionary	News Article Content

detailed flow diagram of the proposed approach as discussed above is shown in Fig. 5.



```
Algorithm 1.(Misleading information Detection)
Input(Text/Text additive image)  and  Output(Status)
 def. main ():
  if(image)
      claim= OCR(image)
  else:
  claim= raw input(claim)
  query= pre-processing(claim)
  Feature1= Word_Similarity_based_feature (query);
  Feature2= Levenshtein_Distance_based feature (query);
  Feature3= N-Gram based Feature (query);
  Status= Classification_model (Feature1, Feature 2, Feature 3)
  Output(Status) //Misleading or real
```

4 Experimental results

This section describes the experiments conducted to respond to the questions we draw up in the introduction part (Sect. 1). The performance experiment is made with the self-generated dataset *Covid-fakenews2019* dataset and publically available dataset (Liar Dataset). Currently, no fake news dataset is available for COVID-19, the experiment is conducted on the proposed dataset, and the model's performance is evaluated on the same. To validate the model's generalizability, the testing has also been performed on the publicly available dataset "Liar." The metric used in this paper is Accuracy(Acc), Precision(Pre),

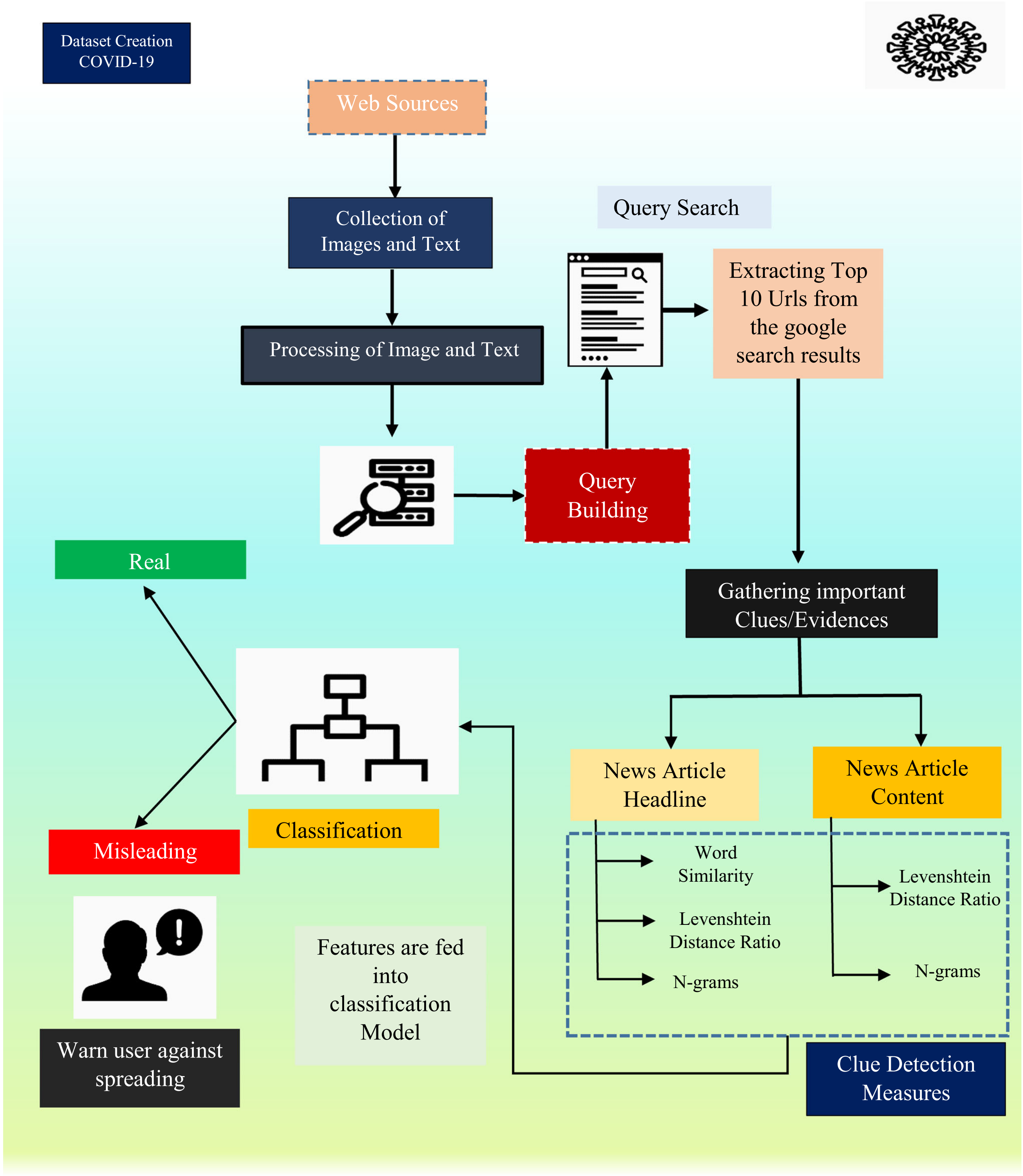

Dataset Creation COVID-19
Web Sources
Collection of Images and Text
Processing of Image and Text
Query Search
Extracting Top 10 Urls from the google search results
Query Building
Gathering important Clues/Evidences
Real
Misleading
Classification
News Article Headline
News Article Content
Word Similarity
Levenshtein Distance Ratio
N-grams
Levenshtein Distance Ratio
N-grams
Warn user against spreading
Features are fed into classification Model
Clue Detection Measures

Fig. 5 The figure shows the flow diagram of the proposed method

Recall(Rec), and F1-Score(F1). The experiment is conducted to analyze the performance of the proposed model. The evaluation results analysis on the self-generated and Liar datasets is discussed in Sects. 4.1 and 4.2.

4.1 Evaluation on self-generated dataset (*Covid-fakenews2019)*

To evaluate the proposed algorithm on the Self-Generated Dataset, the percentage split technique has been

Table 2 Effectiveness of proposed model on the Covid-fakenews2019 dataset with respect to New article headline

Classifier	Word similarity				Levenshtein distance ratio				N-Grams				All- features			
	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1
Logistic Regression	**99.98**	0.99	0.99	0.99	**99.98**	0.99	0.99	0.99	99.93	0.99	0.99	0.99	99.85	0.99	0.99	0.99
Naïve Bayes	99.02	0.99	0.99	0.99	99.02	0.99	0.99	0.99	99.03	0.99	0.99	0.99	98.62	0.98	0.98	0.98
Random forest	**99.99**	0.99	0.99	0.99	**99.99**	0.99	0.99	0.99	**99.99**	0.99	0.99	0.99	**99.99**	0.99	0.99	0.99
Adaboost	67.46	0.67	0.67	0.67	59.01	0.59	0.59	0.56	56.74	0.66	0.56	0.51	70.16	0.70	0.70	0.69
K-Nearest	**99.98**	0.99	0.99	0.99	**99.98**	0.99	0.99	0.99	**99.99**	0.99	0.99	0.99	99.96	0.99	0.99	0.99
SGD	**99.98**	0.99	0.99	0.99	**99.98**	0.99	0.99	0.99	**99.99**	0.99	0.99	0.99	**99.99**	0.99	0.99	0.99

Table 3 Effectiveness of proposed model on the Covid-fakenews 2019 dataset with respect to New article content

Classifier	Levenshtein distance ratio				N-grams				All- features			
	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1
Logistic regression	99.94	0.99	0.99	0.99	99.96	0.99	0.99	0.99	**99.94**	0.99	0.99	0.99
Naïve bayes	99.94	0.99	0.99	0.99	99.87	0.99	0.99	0.99	99.80	0.99	0.99	0.99
Random forest	**99.99**	0.99	0.99	0.99	**99.99**	0.99	0.99	0.99	99.89	0.99	0.99	0.99
Adaboost	59.48	0.60	0.59	0.57	56.25	0.67	0.56	0.50	60.54	0.62	0.60	0.57
K-nearest	**99.99**	0.99	0.99	0.99	99.92	0.99	0.99	0.99	99.89	0.99	0.99	0.99
SGD	**99.99**	0.99	0.99	0.99	99.98	0.99	0.99	0.99	99.89	0.99	0.99	0.99

Table 4 Effectiveness of proposed model on the Covid-fakenews2019 dataset with respect to both new article headline and content

Classifier	10-fold cross-validation				20-fold cross-validation				Percentage split(70:30)			
	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1
Logistic regression	99.95	0.99	0.99	0.99	99.96	0.99	0.99	0.99	99.92	0.99	0.99	0.99
Naïve bayes	98.41	0.98	0.98	0.98	98.41	0.98	0.98	0.98	98.33	0.98	0.98	0.98
Random forest	**99.99**	**0.99**	**0.99**	**0.99**	**99.99**	**0.99**	**0.99**	**0.99**	**99.99**	**0.99**	**0.99**	**0.99**
Adaboost	70.42	0.71	0.70	0.69	69.98	0.70	0.70	0.69	69.98	0.70	0.70	0.69
K-nearest	99.96	0.99	0.99	0.99	99.97	0.99	0.99	0.99	99.96	0.99	0.99	0.99
SGD	99.98	0.99	0.99	0.99	99.98	0.99	0.99	0.99	99.92	0.99	0.99	0.99
Linear SVM	**99.99**	**0.99**	**0.99**	**0.99**	**99.99**	**0.99**	**0.99**	**0.99**	**99.99**	**0.99**	**0.99**	**0.99**

employed with 70% samples are used to train the model and 30% to test the data. The performance has been analyzed based on three different ways. Firstly, the effectiveness of the proposed model is evaluated with respect to news article title/headline using percentage split technique (70:30) as shown in Table 2. Secondly, performance concerning to the news article content is evaluated and lastly, the effectiveness is evaluated by employing percentage split as well as cross-fold validation technique (tenfold and 20 fold) by combining the clues identified from both news article headline and its content as shown in Tables 3 and 4, respectively. The effectiveness of the proposed model has been evaluated with respect to the features employed for the news article headline which shows that the SVM classifier outperforms all other classifiers when considering all sets of features. Form the observation analysis shown in Tables 2, 3, and 4, most of the classifiers perform best on our proposed method by employing different set of categories, which clearly shows the effectiveness of the proposed model. The effectiveness of the proposed model by incorporating all set of features in terms of accuracy measure is shown in Fig. 6. From the figure, it can be clearly seen that most of the classifiers perform well with respect to different testing schemes (tenfold, 20 fold, 70:30 split). The highest value achieved by the different models are shown in Bold text in Table 2, 3, 4, and 5.

4.2 Evaluation on Liar dataset

To validates the generalizability of the proposed model and to do a comparative analysis with the state-of-the-art methods to show the effectiveness of the proposed

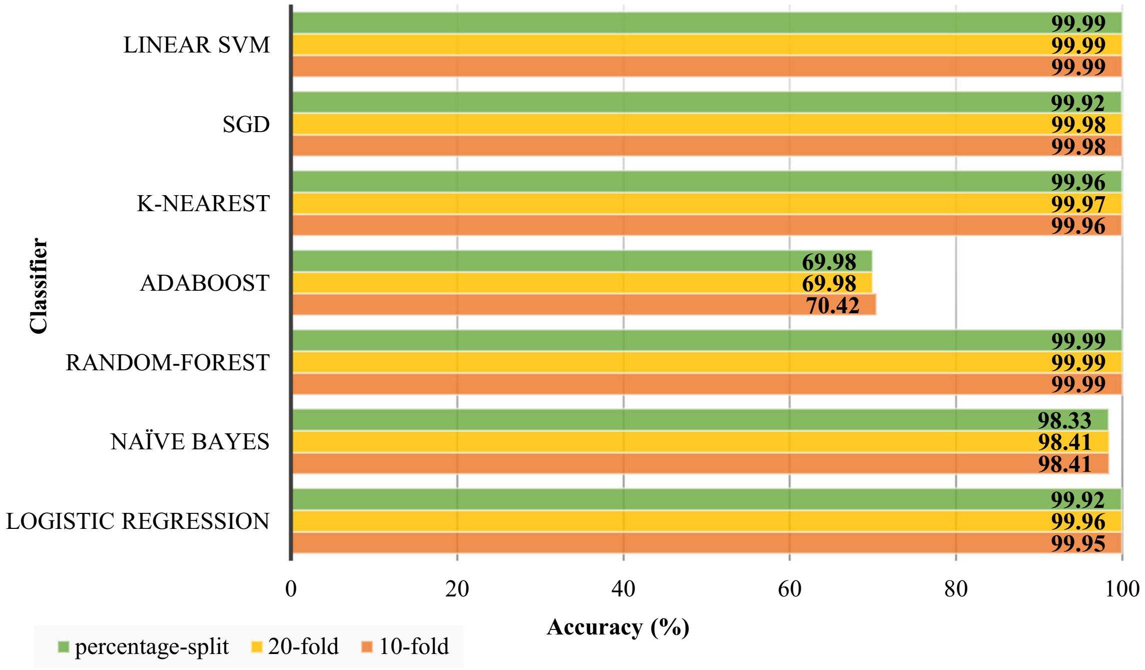


Fig. 6 Effectiveness of the proposed model on Covid-fakenews 2019 dataset

algorithm for predicting the misleading post on the social media platform, we have employed another standard dataset "Liar Dataset" to evaluate the performance of the proposed algorithm. Liar dataset was created by [28], which is the collection of 12,836 short-statements from Politifact platform[9](Fact checking website) which verify and label any text claim into six categories "true," "mostly-true," "half-true," "mostly-false," "false" and "pants-fire." As per the probabilistic analysis, the Politifact considered the true and mostly true sample in the credible category and labeled as true, whereas half true, mostly false, false and pants-on-fire samples are considered in the false category.[10] This categorization is based on the Politifact description given for the true and false label, which classified true and mostly true as accurate statements and others as inaccurate. Accordingly, we have considered these six label categories into true and false samples for the binary classification. The authors of [1], also applied the same way for the binary classification. The proposed solution has been applied on the given dataset, and the detailed evaluation measure like Accuracy, Precision, Recall, F-Measure, and analysis results is shown in Table 5. From Table 5, it can be observed that the proposed technique outperforms the existing algorithm by employing different classification algorithms with respect to Accuracy, Precision, Recall, and F1score. Most of the classifiers perform best, whereas SVM outperforms all other with an F1 score of 0.99 on both percentages split and cross-validation scheme. The existing work analysis also reports SVM as an effective classifier for the prediction of fake news [29, 30].

Table 5 Comparative study of the proposed method with the existing state-of-the-art methods

References	Model	Acc	Pre	Rec	F1
[29]	SVM	0.56	0.57	0.56	0.48
	LR	0.56	0.56	0.56	0.51
	Decision tree	0.51	0.51	0.51	0.51
	Ad boost	0.56	0.56	0.56	0.54
	Naïve-bayes	0.60	0.59	0.60	0.59
	K-NN	0.53	0.53	0.53	0.53
[28]	SVM	0.25	–	–	–
	LR	0.24	–	–	–
Our method	Random forest	0.98	0.98	0.98	0.98
	LR	0.97	0.98	0.97	0.96
	Naïve bayes	0.97	0.98	0.97	0.95
	Decision tree	0.97	0.98	0.97	0.96
	k-NN	0.97	0.98	0.97	0.95
	SVM	**0.99**	**0.99**	**0.99**	**0.99**

In the field of misleading information detection, very few standard datasets are available and one of the popular datasets that have been widely used for analysis is

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

[10] https://towardsdatascience.com/identifying-fake-news-the-liar-dataset-713eca8af6ac.

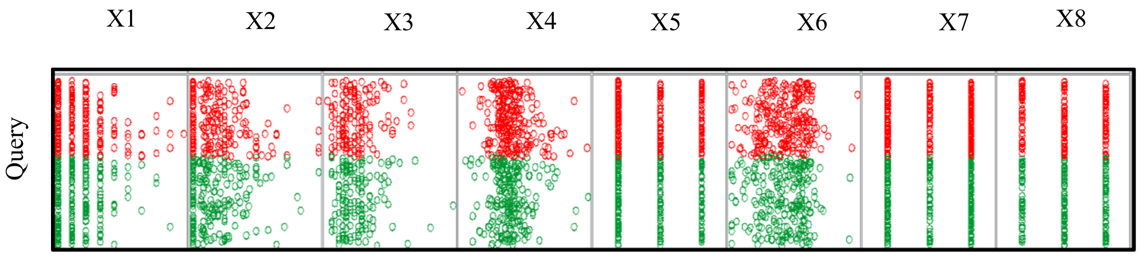


Fig. 7 Scatter plot of each features corresponding to the query

Table 6 List of features and their descriptions

Feature no	Feature description
X1	Word_matching_title_query
X2	Word_match_ratio
X3	Query_length
X4	Title_query_ratio
X5	Title_sentiment
X6	Query_text_ratio
X7	Text_sentiment
X8	Query_sentiment

the LIAR dataset. The comparative analysis has been performed with the existing algorithm on the same dataset. The authors of [29], have employed lexical and sentiment features as well as also included n-grams features for fake news detection. Lexical and sentiment features have included average word length, article length, and exclamation mark count as lexical. Every article's positive, negative, and neutral sentiment have been measured. Unigrams, bigrams, and TF-IDF vectorizer functions have been incorporated to generate TF-IDF n-gram features. The performance analysis over traditional machine learning models reveals that Naïve Bayes with n-gram features perform best in comparison with all other classifiers with an accuracy of 0.60. In [28], the author has exploited text-based features by employing SVM and Logistic regression classifier and achieving an accuracy of 0.25 and 0.24, respectively. The proposed algorithm outperforms the state-of-the-art methods that have been discussed above and SVM found to be the best with an F1 score of 0.99. From the proposed feature analysis, it has been found that Levenshtein distance and N-grams based features are quit efficient and significant for the prediction of misleading information [29, 29]. It has also been observed that news article title and paragraph wise content analysis gives efficient clues for the prediction and is useful in improving the model's performance. None of the previous work has included these paradigms.

It is seen from Table 5, that approximately all given classifiers are performing well on our data. To conclude, the method suggested in the study can identify misinformation at 99% confidence level. The procedure we have applied to show statistical analysis is as follows. Firstly, we identify and depict the nature of our data samples. Next we establish a relation between the data samples concerning to query w.r.t each derived feature $X1, X2, X3...X8$. To show the clear pattern of data samples concerning fake and real class, a scatter plot is shown in Fig. 7. As in the diagram, it can be clearly visualize that the data samples are linearly separable. The decision boundary can distinguish the real and fake samples correctly because the data does not contain noise and less overlapping. The classification can be easily done as shown in Fig. 7. Here, green samples are considered as true and the red samples are considered as fake. The feature $X1, X2, X3...X8$ are described briefly in Table 6.

A statistical analysis of the result is performed through a hypothesis test, where we have considered the level of confidence is 99%. We considered a Null Hypothesis (H0) and Alternative Hypothesis (H1). The alternative hypothesis is the initial hypothesis that predicts a relationship between variables. Whereas the null hypothesis is a prediction of no relationship between the variables. The hypothesis for our case is as follows:

H0: None of the given features are effective in predicting misinformation.

H1: The given features can predict misinformation effectively.

The confidence interval method has been applied here, the test statistic between or outside the confidence interval. We performed the Z-Test,[11] were two samples(query_text_ratio, title_query_ratio) for mean on 99% confidence interval, and it can be clearly seen from Table 7 that the test is rejecting the null hypothesis because p (significance value) < 0.01, this implies it accept the alternative hypothesis (The given features can predict misinformation

[11] https://www.geeksforgeeks.org/z-test/.

Table 7 Statistical analysis results on Z-Test: Two Sample for Means

Parameters	Query_text_ratio	Title_query_ratio
Mean	0.282686586	0.402105109
Known variance	1.38E−02	1.10E−02
Observations	22,366	22,366
Hypothesized mean difference	0	
Z	− 113.366	
$p(Z < \ = z)$ one-tail	0.000	
Z Critical one-tail	2.326	
$p(Z < \ = z)$ two-tail	0.000	
Z Critical two-tail	2.576	

effectively) with the confidence level of 99%. Hence, the results are statistically significant and accept H1 (The given features can predict misinformation effectively).

5 Conclusion

In this paper, we developed an intelligent generalized strategy for identifying possible clues to predict misleading information where fake news proliferated during the COVID-19 outbreak is considered as a special case study and detailed analysis has been discussed. To extract important clues for the verification of news content, the scheme leverages three sets of novel features proposed in this paper that are based on Word Similarity, Levenshtein Distance, and N-Grams. These features are extracted from the news article headlines/titles and their content. The set of features is given as an input to the machine learning model. Performance has been analyzed by considering different variants of feature category by employing a different set of classifier applied on self-generated dataset *Covid-fakenews2019*. and the *Liar dataset*. It has been observed that most of the classifiers are performing well by employing both the dataset, where SVM outperforms all other classifiers with an f1-score of 0.99. In future, we are planning to build the real-time application of the proposed work and also extend the proposed dataset by incorporating a large number of samples with different categories of news that proliferated during the COVID-19 outbreak, like in this work, we are more focusing into the fake news propagated with respect to cures/treatment of the disease. In future, one can also include other types of news like false information regarding lockdowns in the cities, fake claims that are pretended to be posted by some government officials, etc.

Data availability Data will be made available upon request.

Declarations

Conflict of interest It is declare that we do not have any conflict of interest.

References

1. Sahu KK, Mishra AK, Lal A (2020) Comprehensive update on current outbreak of novel coronavirus infection (2019-nCoV). Ann Transl Med 8(6)
2. Boididou C, Papadopoulos S, Zampoglou M, Apostolidis L, Papadopoulou O, Kompatsiaris Y (2018) Detection and visualization of misleading content on Twitter. Int J Multimed Inf Retr 7(1):71–86
3. Zhou W, Wang A, Xia F, Xiao Y, Tang S (2020) Effects of media reporting on mitigating spread of COVID-19 in the early phase of the outbreak. Math Biosci Eng 17(3):2693–2707
4. 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
5. Jin Z, Cao J, Zhang Y, Zhou J, Tian Q (2017) Novel visual and statistical image features for microblogs news verification. Trans Multi 19(3):598–608
6. Vishwakarma DK, Varshney D, Yadav A (2019) Detection and veracity analysis of fake news via scrapping and authenticating the web search. Cogn Syst Res 58:217–229
7. Wu K, Yang S, Zhu KQ (2015) False rumors detection on Sina Weibo by propagation structures. In: 2015 IEEE 31st Int. Conf. Data Eng. pp 651–662
8. Horne BD, Adali S (2017) This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In: Eleventh international AAAI conference on web and social media
9. Chakraborty A, Paranjape B, Kakarla S, Ganguly N (2016) Stop clickbait: detecting and preventing clickbaits in online news media. In: 2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM). pp 9–16
10. Aker A, Derczynski L, Bontcheva K (2017) Simple open stance classification for rumour analysis. arXiv Prepr arXiv:1708.05286
11. Briscoe EJ, Appling DS, Hayes H (2014) Cues to deception in social media communications. In: 2014 47th Hawaii international conference on system sciences. pp 1435–1443
12. Giasemidis G et al. (2016) Determining the veracity of rumours on Twitter. CoRR, vol. abs/1611.0
13. Swon S, Cha M, Jung K, Chen W, Wang Y (2013) Prominent features of rumor propagation in online social media. In: 2013 IEEE 13th Int. Conf. Data Min. pp 1103–1108
14. Zeng L, Starbird K, Spiro ES (2016) "# unconfirmed: classifying rumor stance in crisis-related social media messages. In: Tenth international AAAI conference on web and social media
15. Derczynski L, Bontcheva K, Liakata M, Procter R, Hoi GWS, Zubiaga A (2017) SemEval-2017 task 8: RumourEval: determining rumour veracity and support for rumours. CoRR, vol. abs/1704.0
16. Tacchini E, Ballarin G, Della Vedova ML, Moret S, de Alfaro L (2017) Some like it hoax: automated fake news detection in social networks. arXiv Prepr arXiv:1704.07506
17. Zhou L, Twitchell DP, Qin T, Burgoon JK, Nunamaker JF (2003) An exploratory study into deception detection in text-based computer-mediated communication. In: 36th annual Hawaii international conference on system sciences. pp 10

18. Ferreira W, Vlachos A (2016) Emergent: a novel data-set for stance classification. In: Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies. pp 1163–1168
19. Breiman L (2017) Classification and regression trees. Routledge, London
20. Loh W-Y (2011) Classification and regression trees. Wiley Interdiscip Rev Data Min Knowl Discov 1(1):14–23
21. Bodnar T, Tucker C, Hopkinson K, Bilén SG (2014) Increasing the veracity of event detection on social media networks through user trust modelling. In 2014 IEEE international conference on Big Data (Big Data). pp 636–643
22. Yu F et al (2016) Detecting rumors from microblogs with recurrent neural networks. CoRR 8(1):1435–1443
23. Kochkina E, Liakata M, Zubiaga A (2018) All-in-one: multi-task learning for rumour verification. In: Proceedings of the 27th international conference on computational linguistics. pp 3402–3413
24. Jacovi A, Shalom OS, Goldberg Y (2018) Understanding convolutional neural networks for text classification. arXiv Prepr arXiv:1809.08037
25. Yu F, Liu Q, Wu S, Wang L, Tan T (2019) Attention-based convolutional approach for misinformation identification from massive and noisy microblog posts. Comput Secur 83:106–121
26. Song C, Yang C, Chen H, Tu C, Liu Z, Sun M (2019) CED: Credible early detection of social media rumors. IEEE Trans Knowl Data Eng 33(8):3035–3047
27. Ajao O, Bhowmik D, Zargari S (2018) Fake news identification on twitter with hybrid cnn and rnn models. In: Proceedings of the 9th international conference on social media and society. pp 226–230
28. Wang W (2017) Liar, Liar Pants on Fire': a new benchmark dataset for fake news detection
29. Khan JY, Khondaker M, Islam T, Iqbal A, Afroz S (2019) A benchmark study on machine learning methods for fake news detection. arXiv Prepr arXiv:1905.04749
30. Bondielli A, Marcelloni F (2019) A survey on fake news and rumour detection techniques. Inf Sci 497:38–55
31. 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. pp 127–138

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