Applied Intelligence (2021) 51:4214–4235
https://doi.org/10.1007/s10489-020-02057-9

A unified approach for detection of Clickbait videos on YouTube using cognitive evidences

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

Accepted: 31 October 2020 / Published online: 2 January 2021
© Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Clickbait is one of the form of false content, purposely designed to attract the user's attention and make them curious to follow the link and read, view, or listen to the attached content. The teaser aim behind this is to exploit the curiosity gap by giving information within the short statement. Still, the given statement is not sufficient enough to satisfy the curiosity without clicking through the linked content and lure the user to get into the respective page via playing with human psychology and degrades the user experience. To counter this problem, we develop a Clickbait Video Detector (CVD) scheme. The scheme leverages to learn three sets of latent features based on User Profiling, Video-Content, and Human Consensus, these are further used to retrieve cognitive evidence for the detection of clickbait videos on YouTube. The first step is to extract audio from the videos, which is further transformed to textual data, and later on, it is utilized for the extraction of video content-based features. Secondly, the comments are analyzed, and features are extracted based on human responses/reactions over the posted content. Lastly, user profile based features are extracted. Finally, all these features are fed into the classifier. The proposed method is tested on the publicly available fake video corpus [FVC], [FVC-2018] dataset, and a self-generated misleading video dataset [MVD]. The achieved result is compared with other state-of-the-art methods and demonstrates superior performance.

Keywords Clickbait detection · Cognitive evidence · Dataset · User profiling · YouTube

1 Introduction

In the era of instant gratification, people mostly used to communicate with each other via social media platforms. As social media platforms like Twitter, Facebook, YouTube, etc. providing ease of posting user opinions, lots of misleading and unreliable multimedia content are often posted and widely disseminated via popular social media platforms. Among all other platforms, YouTube is one of the leading ones for sharing videos, it has billions of users covering almost one-third of the internet population and reaches billions of views per day,[1] due to which it is plagued with clickbait videos that doesn't faithfully represent the situation that it refers to. Clickbait's are purposely designed to attract the user's attention and make them curious to follow the link and read, view, or listen to the attached content. In 1994, George Loewenstein has explained clickbait, *"as the information gap theory of curiosity"* [1]. We followed this definition and defined clickbait "*as the information gap theory of curiosity, that play with human psychology, to lure the user to view a content that does not faithfully represent the claim it presents and eventually degrades the user experience*". Whereas, "*Non-click baits can be defined as the content that is presenting the real news and faithfully giving the same picture of content to the viewer, that it is claiming for*". The paper provides a detailed description of clickbait video detection mechanisms in online social media platforms. Detecting clickbait videos is an intelligent task, as it analyses the video content automatically using clickbait video detection frameworks/tool/plugins, as well as in the future it can also be used as an intelligent warning system that can help to automatically report the credibility of video content to the user. Figures 1 and 2 shows the example of clickbait and non-clickbait video in brief. A recent example is of COVID- 19 pandemic, which affects the worldwide badly, and there is no shortage of people who are taking this crisis as an opportunity for malicious activities/gaining profit. A lot of health-related

[1] https://www.youtube.com/yt/about/press/

✉ Dinesh Kumar Vishwakarma
dvishwakarma@gmail.com

Deepika Varshney
deepikavarshney06@gmail.com

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

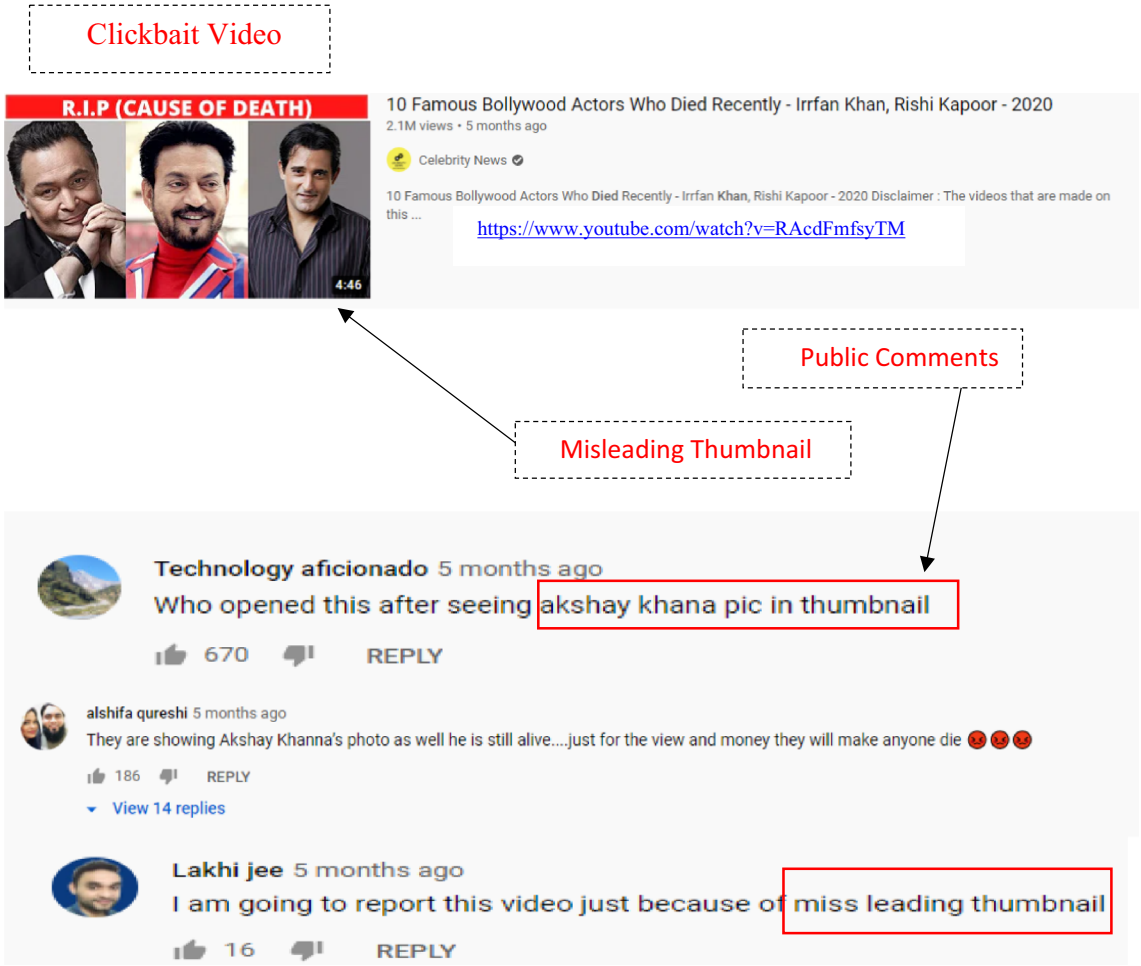

Fig. 1 Example of Clickbait Video

misleading information, some of the fake cures are suggested for COVID-19 have been posted by the malicious user, by adding catchy coronavirus headlines to increase their chance of a click, download, or purchase.

During this pandemic, people have their eye on any news announcement from the government official or some news that can help to get rid of COVID-19. The bad guy uses this opportunity to gain more views on their post, by adding catchy headlines with the news content that does not faithfully represent the event that it refers to and in this way it is spreading false information.[2] One of the fake YouTube video, gone viral having nearly half a million views falsely said that inhaling hot water from a hairdryer can help to cure the coronavirus,[3] which later turned out to be false. The presence of such misleading content over social media makes it more difficult for the user to discriminate the credible information from false stories, and it leads making it a challenging area in research. As the spread of clickbait videos not only degrades user experience, but also decreases the trustworthiness of video-sharing platforms. Few works have been reported in detecting clickbait on the video platform. There is a careful analysis required among the features extracted from the video. The current research has not addressed this problem fully, as they focus only on the content-based solution like the content of the video [2, 3], the image of the thumbnail [4, 5] or text of the title. Most of the text-based clickbait detection methods have adopted linguistic features [6], or word embeddings for the detection of clickbait news headlines, but those solutions cannot be employed to address the clickbait videos, as the only title may not be a reliable indicator, because two videos can share the same title with different content. In the same way, another sort of image-based approaches are employed that focus on thumbnail based features and are not found to be effective in solving the video clickbait detection problem.

In this work, we have proposed a novel mechanism by introducing three sets of evidential clues, identified and retrieved concerning each video, so that one can easily discriminate unsubstantiated information. The recent work addressed various text-based and comment-based features, however neglecting video speech-based features, as well as user profile

[2] https://www.buzzfeednews.com/article/janelytvynenko/coronavirus-fake-news-disinformation-rumors-hoaxes
[3] https://www.bbc.com/news/52124740

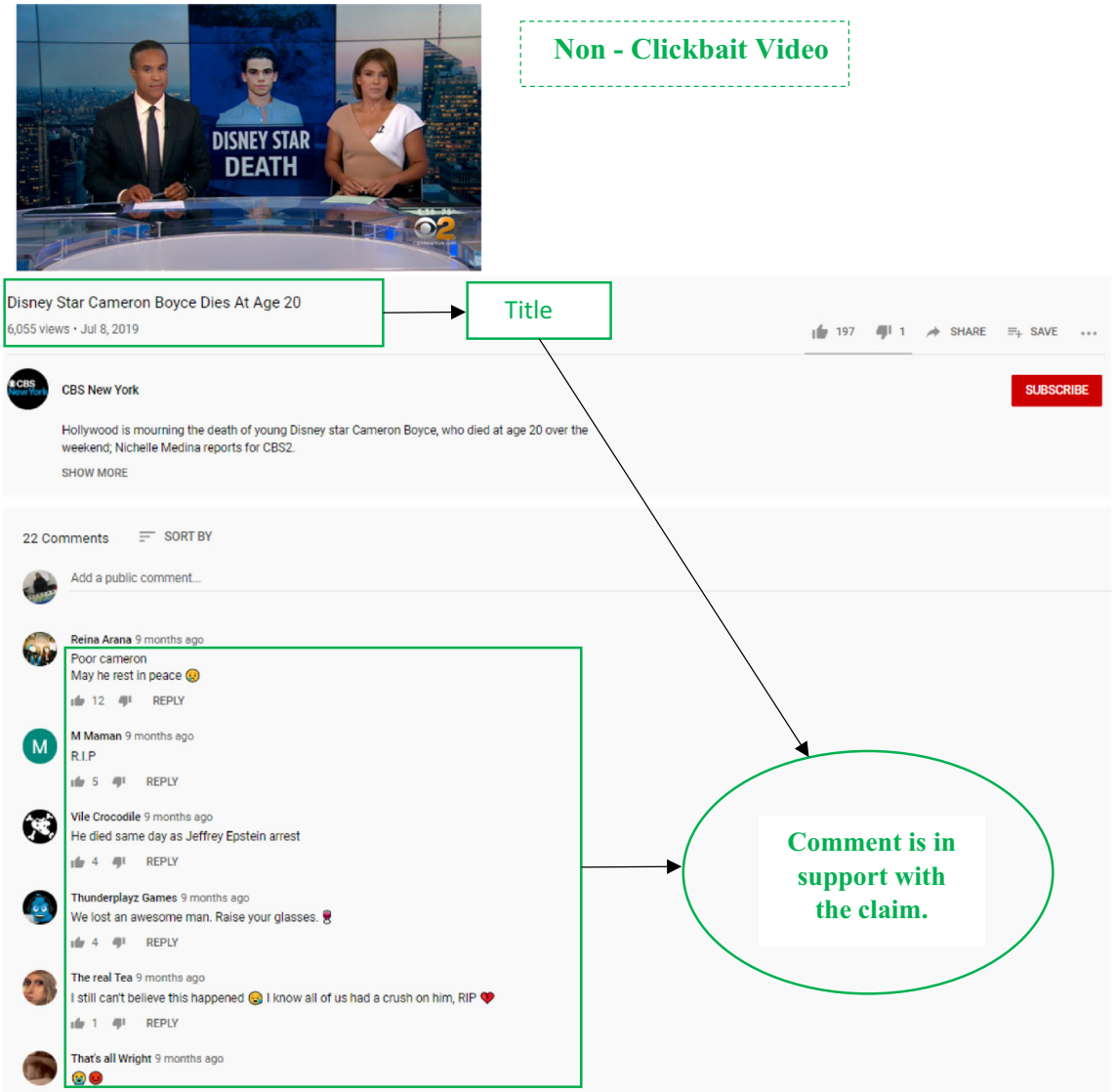

Fig. 2 Example of Non-Clickbait

features that are also not explored well and found to be very effective in detecting clickbait videos. To the best of our knowledge, speech-title similarity based features have not been explored well by the previous research, which can be an important clue to solve the problem. When we have two videos with a similar title but different content, the speech is converted into text, and then comparison has been made with the title to check how faithfully the speech is representing the title. Along with it, we have also addressed the problem, when the comments based features are not retrieved, as the uploader doesn't allow comments from the viewers. In that case, relying only on a certain set of features is not effective. To address this case, we have introduced another clue i.e. credible sources, through which we will be able to predict the credibility of the video, the detailed description has been given in the later sections. The proposed work makes the following significant contributions for the clickbait detection task:

- The proposed work gives a significant contribution in providing a novel methodology for the collection and annotation of YouTube videos of different categories. It builds a self-generated dataset[MVD] of Clickbait's and Non-Clickbait videos, as few and small datasets are publically available related to this area of research.
- To the best of our knowledge, our work is first to provide three sets of cognitive evidence for the clickbait's detection task, with the specification of desirable measures that are required for each specific possible detection case.
- The paper firstly includes speech-title similarity measures that haven't been included in the previous research. Some of the previous research is using thumbnail image and title of the video, which are not effective in the case when the two videos [clickbait and Non-Clickbait] having the same title and may also have the same thumbnail, then in that situation, it is critical to discriminate a clickbait and a Non-

Clickbait one. That's why to tackle this problem, the speech is also analysed, which gives a significant clue for discrimination [7].

- In some of the scenarios, the uploaders are not allowing the comments from the viewers, in that case, the comments part is not contributing, and we are losing some crucial features for the reliable prediction. To tackle this problem, we have also introduced other evidence that gathers important clue from the credible web links, when user responses are not there.
- We investigate the model performance with different classifiers, and comparative analysis reveals that our proposed method outperforms other state-of-the-art on the same dataset.

The remaining of this paper is organized as follows. In Section 2, we are going to discuss the previous work that has been done related to this field, where Section 3 gives a detailed description of how the data collection and annotation have been employed, where Section 4 describes the strategy/ method that we have used for the clickbait detection task, which followed by a discussion of experimental results in Section 5. Lastly, the paper is concluded with some suggested future work aspects.

2 Related work

In this section, the previous work on Clickbait, Rumour, Hoax, and Fake news detection are analyzed, and their performances are discussed. These are different prominent forms of misleading content that are available on social media [8] and used interchangeably concerning different contexts. For example, hoaxes can be prominently seen in the context of false celebrities' death stories. While the concept of Fake-news widely comes into picture during the US presidential election. The clickbait is also one of the eminent forms of misleading content available over various social media platforms. Initially, the research is going on the detection of clickbait's headlines that lure the user to view the webpage, now the concept is also moving towards detecting clickbait's videos and becoming an emergent area of research. In the following section, we briefly describe each of these categories.

2.1 Clickbait's detection

Detection of clickbait over the social media platforms, several techniques have been developed [[9, 10]]. For the automatic detection of clickbait, a browser extension has been developed by [10], which allows users with an option to block clickbait, as well as facilitate with the warning mechanism. The authors of [6] incorporate several set of handcrafted features like bag-of-words, n-grams, etc., to train the classifier and firstly introduce an automatic clickbait detector. In [6], authors have used the random forest for the prediction of clickbait tweets. Whereas, the authors of [11], proposed a technique to address the issue by employing SVM and Naïve Bayes classifier, while the authors of [12] address the problem of detecting clickbait's in the news articles, by incorporating Gradient Boosted Decision Trees. The authors of [1] have proposed an ensemble learner-based classification system for the prediction of clickbait and Non-Clickbait ones. From the analysis, it has been observed that random forest found to be the best classifier, with an accuracy of 91.16%. Whereas, [7] developed a novel content-agnostic scheme for effectively detecting the clickbait video by exploring and analyzing the comments from the viewers. Along with the machine learning technique, some of the researchers have also explored the deep learning methods. The authors of [13], firstly introduced the use of deep learning methods to counter the problem of detecting clickbait news articles. The recurrent neural network, in conjunction with word embeddings, has been employed as the proposed technique. In [[13, 14]], a deep neural network method has been used, however, the method considered text content for analysis i.e. the title and the descriptor, which is not quite effective in the case when the clickbait and Non-Clickbait videos are sharing same title and descriptor. From the survey, it has been observed that very few works have been reported clickbait detection solution that has considered video-based clickbait's. In recent work, the authors of [9], employed crowdsourcing based technique, where it has been predicted that the thumbnail of the video is clickbait or not. However, the technique found to be time-consuming, when the dataset is large. The further enhanced work has been proposed by [15], where they have used a combination of thumbnails and the statistical features of the user's comments and our work have not only introduced comment and user profile based features to retrieve useful evidence but also incorporate speech data from the videos for analysis, that further be compared with the title, to identify whether the title faithfully represents the video content or not.

Along with this, there are also some other forms of false information present over social media, like Rumours, Hoaxes, Fabricated, Conspiracy theories, Satire where research is going on [8].

2.2 Rumours detection

Rumours are also one of the forms of false information, which can be defined as the content/ post whose veracity is not verified at the time of posting and whose truthfulness is ambiguous or not confirmed. Many of the work has been done to detect whether a given post is a rumour or not [[16–21]]. The authors of [16], proposed two novel features, client-based, and location-based features. The classification result shows that SVM performs best in detecting fake news. The

authors of [22], proposed linguistic, temporal, and structural features from the tweet post and employed random forest, decision-tree, and SVM for the detection of rumours. The evaluation results reveal that the model achieves precision and recall scores of 87% and 92% respectively. Where in [23], user, structural, linguistic, and temporal features have been explored for the rumour classification, by analysing the task over varying time windows on twitter. In contrast, the authors of [24], learn a hidden representation of the input, by employing a recurrent neural network, without the need for hand-crafted features for rumour classification.

2.3 Hoax detection

Hoax is the news reports, whose facts are either false or incorrect, and they are representing it as a legitimate fact. Many times, Hoaxes have been seen in the context of false celebrities' death stories. The authors of [25], have proposed user interaction based features, and employ logistic regression for classification, that able to identify hoaxes with an accuracy of 99%. Whereas, in [26] the author proposes a technique using a random forest classifier to distinguish if an article is a hoax or not. The experimental results reveal that the proposed model achieves an accuracy of 92%. Whereas, the authors of [27], proposes a hoax detection scheme by employing user feedback features for news verification using a Naïve Bayes algorithm. However, Hoax detection is the less explored area, where few works have been reported and need further research.

2.4 Fake news detection

Fake news is another type of false information present over social media, and fake news can be defined as "a news article that is intentionally and verifiably false" [28]. Various textual and visual features have been employed for the detection of Fake news. The authors of [29], proposed a similarity-aware fake news detection method, that employed multimodal data(textual and visual) to investigate the relationship between the extracted features across modalities. The result reveals that the multimodal features and cross-modal similarity relationships are effective in detecting fake news. Whereas, in [30] authors proposed a technique for fake news detection by combining text mining techniques and supervised artificial intelligence algorithms, where the result shows that the best mean values in terms of precision, accuracy, recall, and f-measure have been obtained from the decision-tree, CVPS, ZeroR algorithms. In [31], the authors adopted a deep neural network(Convolution and Recurrent neural network) for the feature extraction process to predict fake news. Whereas, credible web sources are analyzed by [32] for the fake news prediction. They proposed a novel Rp (Non-Clickbait parameter) parameter, in which if the given threshold value exceeds, the event is classified as fake otherwise not. While many of the authors [[33–36]] provide a good literature survey by exploring techniques, features, datasets, and other analyses for further research exploration.

The original contribution of the work can be seen in Fig. 3, where the existing work and the proposed work has been demarcated. The main contribution of the work is as under:

- The major challenge in the area of clickbait detection is that few public datasets are available for the detection of clickbait videos, and the majority of the datasets are available incorporating clickbait headlines. The self-generated dataset[MVD] has been proposed in this paper, which further helps to explore the research in this field.
- Few methodologies have been proposed that aims to detect clickbait video [7]. This is an emerging field and largely unsolved problem, due to which still very few works have been reported in this area. This gives novelty to our work and also motivates us to provide an efficient solution for clickbait detection. Three sets of novel features (Video-based, Comment based, and Channel-based) are reported in this work that are found to be efficient and outperforms other state-of-the-art methods on the same dataset.
- To the best of our knowledge, no work has been reported on the concept of tackling videos having commenting off, where the uploader is not allowing users to comment on their video due to which no important clues can be retrieved from the comments section to predict the video as clickbait or real. Top 15 web headlines are fetched and analyzed to get some important evidence about a video and helps in efficient prediction.
- It has been observed that many of the existing works have considered video metadata instead of extracting some clues from the video transcripts. To the best of our knowledge, we have firstly included the video transcript based feature to get some informative evidence out of it as explained in section 4.2.1.
- The comparative analysis has been done with the other existing algorithms. The results clearly show that the proposed model is superior and outperforms the other existing state-of-the-art.

3 Dataset creation

One of the significant contributions of this work is the dataset creation since few datasets are available. Hence, a dataset of clickbait's and the Non-Clickbait videos have been developed by collecting a diverse set of videos using YouTube REST Data API v3. The details include video content (title, likes, dislikes, views, etc.), Number of comments, Channel details (Number of subscribers, registration date, video count, and view count). In the field of misleading video detection, very

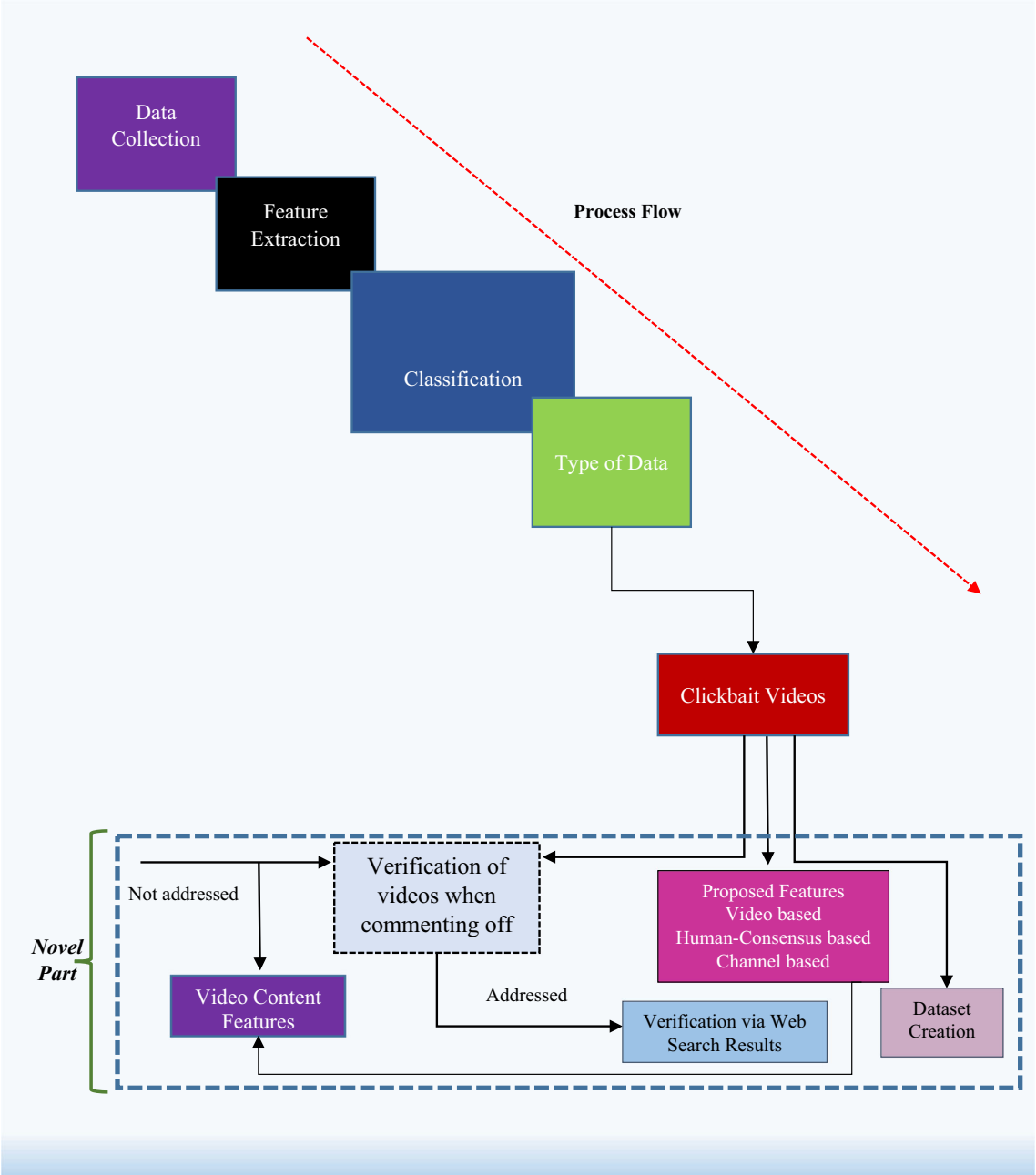

Fig. 3 Flow of Proposed Methodology

few datasets are available, which leads to give the aim to build a generalized dataset, incorporating various categories. From the list of 16 most popular videos category defined by YouTube,[4] we have collected 987 videos(474 Clickbait and 513 Non-clickbait). To collect clickbait videos, we manually crawled and annotated each of the 474 videos. We have analyzed some of the channels as well as their following channels that have prominently posted clickbait's, to lure the user to visit their video. Some of the channels[5] that are posting claim make users curious to visit a link for getting more impression on their video. YouTube has a good check algorithm, for detecting fraudulent videos and also have a blocking mechanism, then also most of the videos are still in their active stage and not removed. From the study, it has been analyzed that most of the clickbait's or hoaxes are posted concerning celebrity's death, which later found to be false and degrades the user experience. As no correct verification is provided, this posted news hamper the public emotions as well. That's why detecting clickbait's video is one of the prominent areas of research. To reduce the time of collecting fake videos, the strategy that we follow is the viral videos, because they are

[4] https://mediakix.com/blog/most-popular-youtube-videos/
[5] https://www.youtube.com/watch?v=zDa-HzCFoIo&t=8s

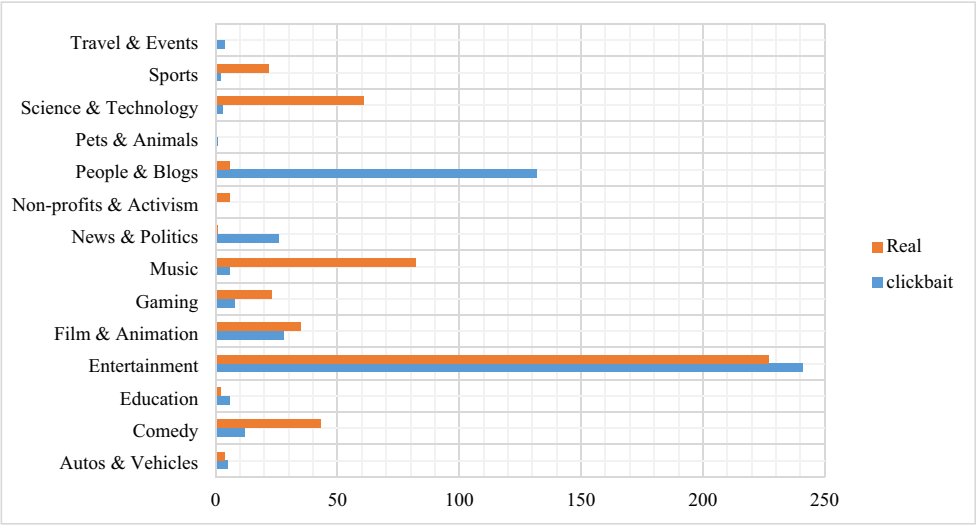

Fig. 4 Number of Videos by Category and Class

more likely to generate fake content, with having catchy headlines also prone to make them viral. We manually analyzed the channels/source for generating these videos.[6]

Additionally, the titles are also scrapped and analyzed to get some clickbait's phrases like some of them are "Shocking", "OMG", "Sad News", "Bad News"," Dukhd Khabar". To direct our search in the correct direction, we try to find out those channels that are following these channels, because it is more likely that the following channel is also posting fake content. Video response also plays an important role while segregating fake/clickbait videos, some of the phrases like "fake video", "bullshit", "galat", "hoax", "clickbait"," Alive"," fake news"," liar", "false", "falsely"," misinformation", "rumor"," clickbait"," hoax" are used, along with it the dislike to like ratio has also been used for further filtering, as it has been observed that clickbait/fake videos received more dislikes compare to likes. For the collection of Non-Clickbait videos, some popular authentic channels have been considered for analysis such as "TEDx Talks", "Harsh Beniwal", "Marvel Entertainment" etc. We call this dataset as "MVD" (Misleading Video Dataset), and the distribution of dataset is as given in Fig. 4 and Table 1. It can be seen from the Fig. 4 that 14 different categories have been considered for the dataset creation, whereas in the dataset, the number of clickbait is found more from the "Entertainment" and secondly with "people and blogs" category, as from the manual analysis it has been observed that most of the clickbait's are prominently available in these categories, while very few videos are considered from the categories("Pets and Animals", "Auto and Vehicles") where the possibility of clickbait generation is quite low.

[6] https://www.youtube.com/channel/UC_UdS7tWCwgBDoaM-Hmkzxg/videos

4 Cognitive evidence for Clickbait's detection

In this section, we present the problem definition for the clickbait detection task and describe the proposed method (CVD), a clickbait video detector to address the problem.

4.1 Problem definition

We define the clickbait's detection task as for a given set of videos, the system has to determine which of the videos are reporting clickbait's and are not faithfully representing the event it refers to. The identification of clickbait videos is ultimately meant to warn users that the given video content is not faithful about the claim it representing, and helps in countering the spreading of false content. In this paper, we have considered the following three detection cases shown below in Table 2.

A set of evidence is required to justify and verify the above cases and warn the user to think twice while believing and spreading false information. If these three cases are identified, then there is a possibility that the video is clickbait and does not faithfully represent the event that it refers to. Formally, the task takes a set of video ids $V_{ID} = V_{ID1}, V_{ID2}, V_{ID3}\ldots. V_{IDN}$ as an input, and the classifier has to determine whether each of these videos V_{IDi} is a clickbait or Non-Clickbait by assigning a label from $Y = \{C, R\}$. Hence, we formulate the task as a binary class classification problem, whose performance is analyzed and evaluated by computing the various performance measures like precision, recall, and F1 score for the target class, i.e., Clickbait. There are three types of cognitive evidence that have been considered for the detection of clickbait. Each of these evidence gives a significant contribution in finding the clues in predicting a video as clickbait or not. The three sets of cognitive evidence for a video are as follows.

Table 1 Detailed Description of the Self-Generated Dataset (MVD)

S.No.	Category	Number of Videos	Class
1	Autos & Vehicles	5	clickbait
2	Autos & Vehicles	4	Non-Clickbait
3	Comedy	12	clickbait
4	Comedy	43	Non-Clickbait
5	Education	6	clickbait
6	Education	2	Non-Clickbait
7	Entertainment	241	clickbait
8	Entertainment	227	Non-Clickbait
9	Film & Animation	28	clickbait
10	Film & Animation	35	Non-Clickbait
11	Gaming	8	clickbait
12	Gaming	23	Non-Clickbait
14	Music	6	clickbait
15	Music	82	Non-Clickbait
16	News & Politics	26	clickbait
17	News & Politics	1	Non-Clickbait
18	Non-profits & Activism	6	Non-Clickbait
19	People & Blogs	132	clickbait
20	People & Blogs	6	Non-Clickbait
21	Pets & Animals	1	clickbait
22	Science & Technology	3	clickbait
23	Science & Technology	61	Non-Clickbait
24	Sports	2	clickbait
25	Sports	22	Non-Clickbait
26	Travel & Events	4	clickbait
Total		474 clickbait 513 Non-Clickbait	

4.2 Detecting Clickbait's videos

In this section, we describe the proposed method, the CVD (Clickbait Video Detector) to address the problem formulated previously. The technique consists of three major pieces of evidence, retrieved using three feature components based on Video, Human-consensus, and User-Profile. The first feature component is used to extract video related features (e.g. speech-title similarity, number of likes, number of dislikes, dislike-like ratio). The second feature component is based on human-consensus. This module learns from individual human cognition and combined from the consensus response. The output has been retrieved, which gives the agreement of individuals towards the posted content. The third component is the user-profile feature extraction (e.g. video-age-ratio, channel views, registration age). This is directly related to the reputation of the video uploader. Lastly, we finish with the classification model, which finally does the binary classification (clickbait and Non-Clickbait) using features extracted from the first three components. The overview of CVD is shown in Fig. 5 where the three sets of features are extracted from the video, comments, and channel information.

Table 2 Possible cases for the clickbait's detection

S.No.	Detection Cases
1.	Title faithfully represents the video speech content and comments are not in contradiction
2.	Title does not faithfully represent the video speech content and both are in contradiction.
3.	Title faithfully represents the video speech content and comments are in contradiction.

4.2.1 Video-content based features

Video-Content based feature is the first component of our proposed method. This component is responsible for the extraction of video-content based feature (e.g. speech-title similarity, number of likes, number of dislikes, dislike-like ratio). The speech-title similarity is one of the crucial features and plays a major role in retrieving Evidence 1. The speech-title similarity is the similar to the speech text with respect to the title of the video, which identifies whether the given claim attached to the video, faithfully represents the event that it refers to. To identify how faithfully the video is representing a claim, speech has been extracted from each video. The Google speech to text API has been used for the speech part, that later be converted into text. The cosine similarity has been applied in between the text extracted from the speech part of the video and the title, to measure the similarity among them. Google's speech to text API has been used for speech recognition. The speech to text has three main methods to perform speech recognition (Synchronous, Asynchronous, and Streaming Recognition).[7] Here we have used the synchronous recognition method, as in our case we need to process the data of less than one 1 min and synchronous recognition requests are limited to audio data of 1 min or less in duration. The detailed description of how the complete process is followed is shown in Fig. 6. It shows the retrieving process of the speech-title similarity for Evidence 1. In the first step, the video is given as an input, which is translated into its audio format. The audio is converted into text format using Google Speech API. So, to get better similarity results, the audio is processed in parts. For each video, the 1 min segments have been analyzed, as we are getting enough information within this duration to predict a video is bogus or credible. The audio transcripts of 1 min are subdivided into 4 parts of each 15 s. The four text parts that are incorporated in Fig. 6 is to split the 1-min audio transcripts into smaller subparts to identify how similar the title/claim with respect to the content that is

[7] https://cloud.google.com/speech-to-text/docs/basics

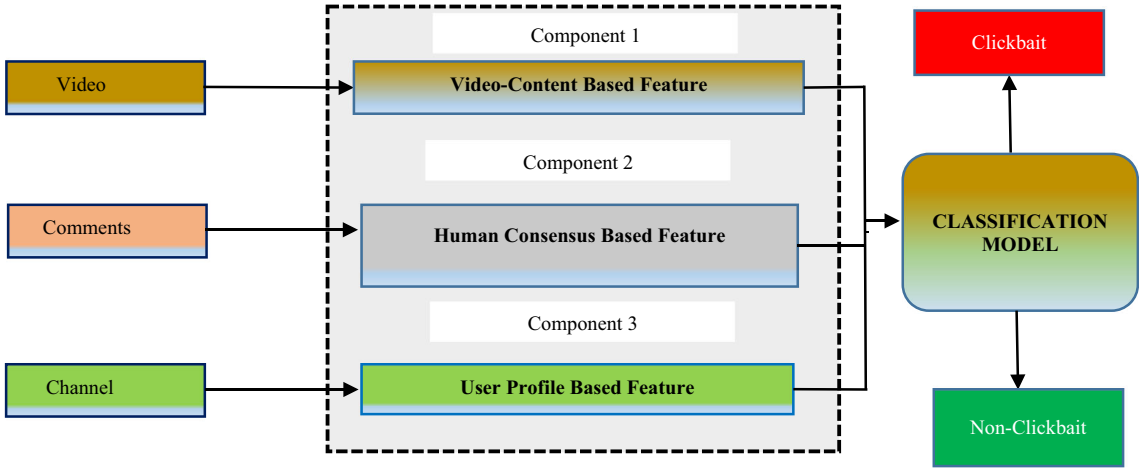

Fig. 5 An Overview of CVD

presented in the video. It has been noticed that in YouTube videos the given titles are too short in length, containing very few words and a 1-min audio transcript considered for analysis is quite lengthy and is not so effective to apply cosine similarity between them for reliable prediction. To address this problem, the audio transcript is split into small subpart (4 text parts of 15 s each) (Text part 1, Text part 2, Text part 3 and Text part 4), each text subpart contains 15-s audio transcripts in sequence after which the individual text similarity with respect to the title has been calculated, later the average value has been considered as the final similarity value.

The video-content based features are shown in Table 3. These features are evaluated to represents the statistical characteristics of the video content.

4.2.2 Human consensus-based feature

Individual human cognition can play an important role and gives a significant contribution in forming evidence for the detection of Clickbait's and the second component of our proposed approach. The individual viewer has their own cognition that has been come out in the form of expression/ emotions given as a video response. Many of the malicious users have not allowed the comments on their video, because the human consensus gives an initial clue about a video, and if any new user visiting the page, they can make an initial thought about video credibility via reading how an individual responds to a video. That's why most of the time, it can be seen that commenting is not enabled on videos created with

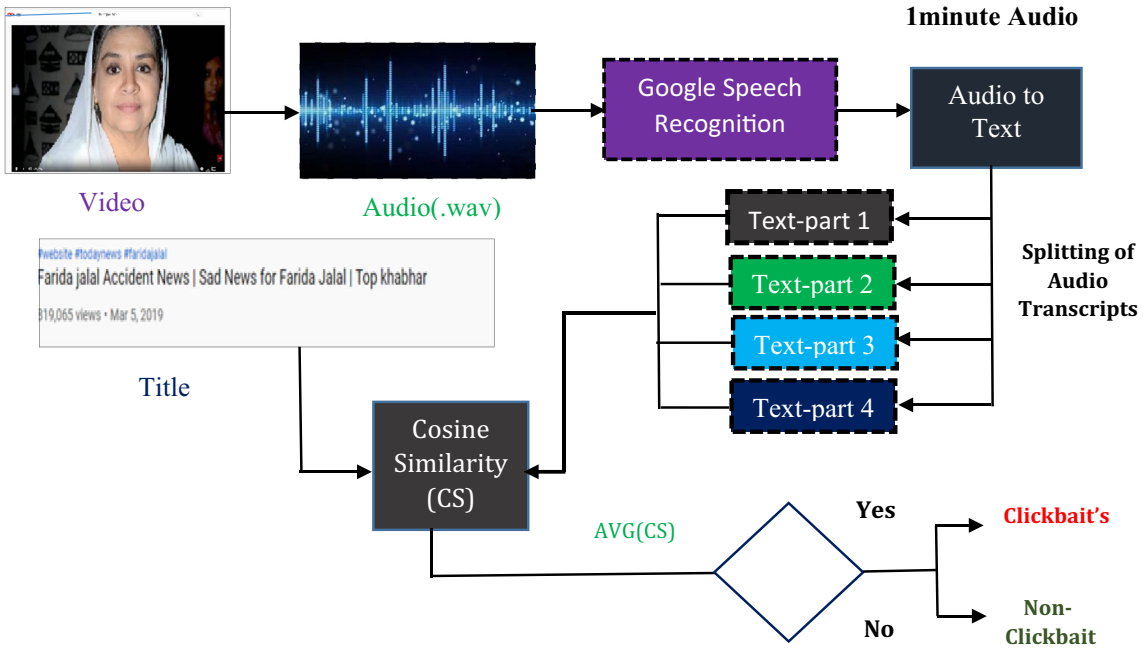

Fig. 6 The Figure represents the process of retrieving Evidence 1

Table 3 Video-Content based feature

Feature	Description
Audio Transcript based Features (Avg_cs)	The average cosine similarity measure between audio transcripts and the title of the video. This is one of the novel features and very few studies incorporate it.
Number of Likes $L(x)$:	This feature represents the number of likes on a video.
Number of Dislikes $D(x)$	This feature represents the number of dislikes on a video.
Dislike to Like Ratio $DL(x)$	The ratio of the number of dislikes to like count on a video. It has been observed that clickbait videos received more dislikes compared to likes. $DL(x)=\frac{D(x)}{L(x)}$
Number of Views	The number of views Received by a video.

some malicious intent. Figure 7 shows the process of retrieving Evidence 2. Here multilingual content has also been addressed. If any of the content is found to be in a different language, it is translated into English text using google translator, then further be used for analysis. As a result, we have also addressed the multilingual content.

A total of 6 Human consensus-based features are extracted. The Human-consensus based features are shown in Table 4. These features are evaluated to represent the statistical characteristics of the responses of the viewers.

4.2.3 User-profile based features

The reputation of the individual channel also plays an important role in identifying the credibility of the uploaded video. (e.g. video-age-ratio, channel views, registration age). A total of 7 User-Profile based features are extracted. The User-Profile based features are shown in Table 5. These features are evaluated to represent the statistical characteristics of the responses of the viewers.

From the all given set of features there are some important findings, it has been observed that the Non-Clickbait video channel has more number of subscribers with respect to registration age as compare to clickbait's video channel. The findings reveal that the average cosine similarity for Non-Clickbait videos is too less than that of the clickbait's, the one of the reason may be, because the clickbait videos probably repeating the same sentence as mentioned in the title, many of the times to extend the length of the video with redundant and bogus content. It has also been found that the number of dislikes is more than the number of likes for many of the videos belongs to clickbait's as compared to non-clickbait's.

4.3 Credible sources

Credible sources are the third sort of evidence, which plays an important role in news verification. The two sets of evidence that we have discussed earlier give significant information; however, they fail in certain situations. The first situation is when the speech-title similarity perfectly matches as well as

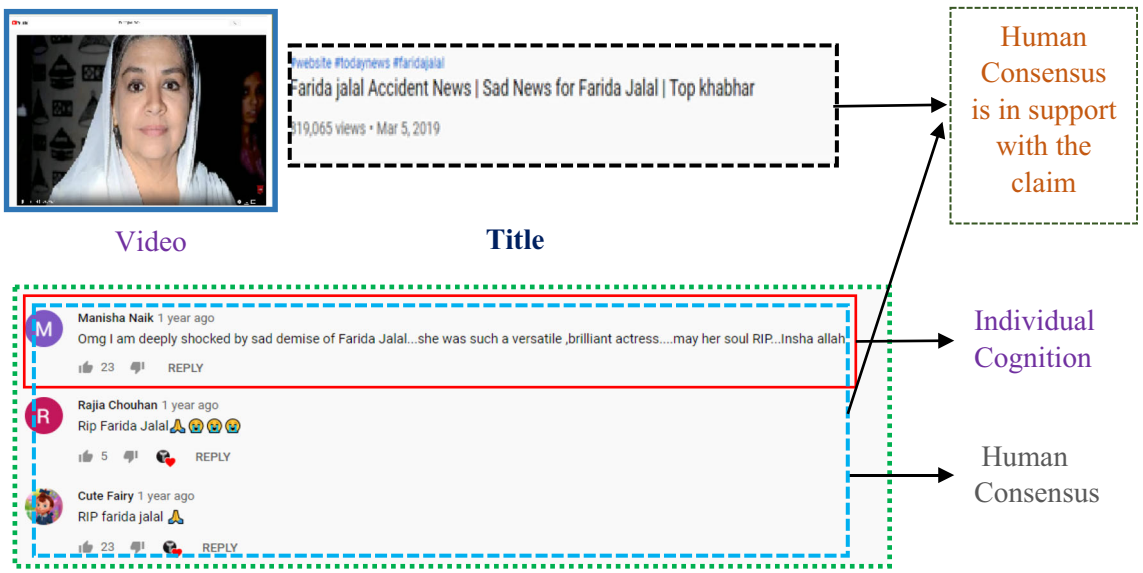

Fig. 7 The Figure represents the process of retrieving Evidence 2

Table 4 Human-Consensus based feature

Feature	Description
Number of Comment c(x)	This feature represents the number of comments received on a video. To restrict our search analysis, in total maximum of 200 comments have been considered. The below equation represent the comment count. $c(x) = \sum_{i=1}^{200}(\text{c i})$
Positive Polarity p(x)	This is the feature that indicates, how many comments showing a positive opinion towards a video. $p(\boldsymbol{x}) = positive\ polarity\ count/c(x)$
Negative Polarity $n(x)$	This is the feature that indicates, how many comments showing a negative opinion towards a video. $n(x) = negative\ polarity\ count/c(x)$
Positive-Negative Polarity Ratio $pn(x)$	This is the ratio of positive to negative comment polarity count. $pn(x) = \frac{p(x)}{n(x)}$
Fake_Comment_Count $FCC(x)$	The fake comment count is the number of comments having clickbait phrases. Clickbait's Phrases(CP) = {fake, bullshit, hoax, wrong… etc.} $FCC(x) = \sum_{i=1}^{200}(\text{CP})$
Fake_Comment_Count Ratio $FCCR(x)$	It is the ratio of the number of fake_comment_count to the total number of comments encountered. $FCCR(\boldsymbol{x}) = FCC/c(x)$

comments are also in supports. So, are these measures sufficient enough? It may happen that, these two pieces of evidence are in support, even though the information is false as shown in Fig. 8 it clearly shows that evidence 1 and evidence 2 are in support, but the information is false. So what is the breaking point here, through which we can reliably conclude about the credibility of news? None other but the credible news sources, we need to scrap the news headlines related to the specified claim by searching it on google, and searching for clickbait phrases in the headlines like {*'misleading', 'misinformation', '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',' not dead',' death rumours'*} for justifying the claim. There can be another case of it where it can be applied when the uploader doesn't allow any comments and make the commenting off, where it is quite needful to retrieve Evidence 3. The query is build using the video title concatenated with the fake news keywords *query= "title+ fake news"*, that goes as a search query to google. Top 15 URLs concerning the specific claim are scrapped and analyzed. These 15 web titles are considered as a replica of video comments, and the same measures are identified here, as evaluated over video comments (Human-consensus based features) to get the informative clues when comments are not available (Table 6).

Algorithm 1 gives the detailed procedure of the retrieval process of all three pieces of evidence, and from all sets of features, some of the features that play a crucial role in retrieving the evidence are identified. The given algorithm takes a set of videos as an input and returns the status that whether a video is clickbait or not as an output, the given algorithm will further be used to create a Non-Clickbait-time prediction of clickbait. The crucial video-based features encountered are having Average cosine similarity (Avg_cs) and Dislike-like ratio (DL) for retrieving Evidence 1, and from the analysis

Table 5 User-Profile based feature

Feature	Description
Registration Age $r(x)$	The age of the user is an indicative measure of the rounded number of days that the user has spent on YouTube, i.e. from the day account was created up to the day of the current post.
Channel Views $CV(x)$	The total number of views received by the channel.
Total_no_of_Videos $V(x)$	The total number of videos has been posted by the channel till date.
Subscriber Count $SC(x)$	The total number of subscribers count on the channel.
Video_to_Age_ratio $VA(x)$:	This is the ratio of the total number of videos uploaded by the channel to its registration age. $VA(x) = \frac{V(x)}{r(x)}$
Subscribers_to_Age_ratio $SA(x)$	This feature represents the ratio of the number of subscribers on the channel to its registration age. $SA(x) = \frac{SC(x)}{r(x)}$
Channel_Views_to_Subscribers $CS(x)$	It is a ratio of the number of views received by the channel to its subscriber count. $CS(x) = \frac{CV(x)}{r(x)}$

of all set of sample the threshold values have been identified, if the Avg_cs >0.10 OR DL > =0.40 than the video is likely to be clickbait concerning the video content and make the value of Evidence 1 to be true or 1, otherwise 0. On the other hand, FCCR (Fake Comment Ratio) plays an important role in retrieving Evidence 2. After getting the value of Evidence 1 and Evidence 2, three different defined cases have been observed concerning these values and the output status has been predicted as Clickbait or Non-Clickbait.

Algorithm 1.(Clickbait Video Detection)
Input(Video_id) and Output(Status)

```
def. func1(Video_id)
    Evidence1= Video_based_feature ();
    Evidence2= Human_consensus_based_feature ();
    If (Evidence1==0) AND (Evidence2==0):
       Scrapped_urls= Google_search(query)
        Evidence 3= Processing(Scrapped_urls)
        IF (Evidence 3==1):
           Status= Print("Clickbait")
        Else:
           Status= Print("Non-Clickbait")
    Elif (Evidence1==1 and Evidence2==1):
       Status = Print("Clickbait")
def. Video_based_feature ()
  Evidence1 = 0
  Title= Extract_Title ();
  Audio_transcript= Extract_Transcript ();
  number_of_dislike= Count(Dislikes);
  number_of_like= Count(like);
 Averge_cosine_similarity(Avg_cs) = Cosine_similarity (Title, Audio_transcript);
 Dislike_like_ratio(DL)= (number_of_dislike)/ (number_of_like)
    If (Avg_cs>0.10 OR DL>=0.40):
        Evidence1= 1
        return (Evidence 1)
    Else:
        Evidence1=0
        return (Evidence 1)
def. Human_consensus_based_feature ()
    Evidence2= 0
    Number of comment c(x) = ∑_{i=1}^{200}(c i )
    Fake_Comment_Count_Ratio(FCCR)= Fake_Comment_Count /c(x);
    If(FCCR>=0.015)
      Evidence2= 1
      return(Evidence2)
    Else:
       Evidence2= 0
       return(Evidence2)
```

5 Experimental results

This section, we evaluate the performance of the CVD scheme in comparison with the state-of-the-art method. The result shows that the CVD technique significantly outperforms the baseline method with respect to different performance measures.

5.1 Comparison of classifiers on the self-generated dataset [MVD]

Tables 7 and 8 briefly describes the results for different classifiers using all sets of features on the Self-Generated Dataset (MVD) by employing two validation strategies, Cross-validation, and Percentage Split, respectively.

Fig. 8 The Figure represents the process of retrieving Evidence 3

From the rigorous analysis of all the proposed features, it has been observed that Human Consensus and User Profile features give a significant contribution and play a major role in predicting the video as clickbait or non-clickbait. Table 7 shows the performance analysis of the model by employing a cross-validation technique concerning different measures, TP (True Positive), False Positive(FP), PRE(Precision), REC(Recall), FM(F-Measure) and ACC(Accuracy). The analysis has been presented by considering all sets of features, video-based features, human consensus-based features, and user-profile features. The performance of the classifiers (Random-forest, Naïve-Bayes, Logistic, SVM, SGD, k-nearest, and J48 using all set of features and for each independent set of features suggests that J48 remarkably outperforms over the rest of the classifiers with the highest accuracy of 98.28 on both cross-validation and percentage-split mechanism. This can be seen clearly when we look at precision, where J48 performs substantially better than the rest. However, the k-nearest neighbor classifier performs better, only when considering user-profile features. It has been observed that SVM is not performing well when considering only video-based features, while significant improvement in the accuracy, when considering all set of features. Whereas, logistic regression performs worst in comparison to all other classifiers in each scenario. On the other hand, using the percentage split technique, the random-forest, and J48 both performing the same on all sets of features in terms of true positive, precision, recall, f-measure, and accuracy.

Figure 9 shows the comparative analysis of various classifiers (Random-Forest, Naïve-Bayes, Logistic, SVM, SGD, k-

Table 6 The table shows the cases and the needful evidence required for the detection

S.NO	Detection Cases	Essential Measure	Desirable Measure
1.	Title faithfully represents the video speech content and comments are not in contradiction	Evidence 1 Evidence 2 Evidence 3	Evidence 1 Evidence 2 Evidence 3
2.	Title does not faithfully represent the video speech content and both are in contradiction.	Evidence 1	Evidence 1 Evidence 2
3.	Title faithfully represents the video speech content and comments are in contradiction.	Evidence 1 Evidence 2	Evidence 1 Evidence 2

Table 7 Performance of the various classifier by employing Cross-Validation

All Set of Features								
Classifiers	Fold1	Fold2	TP	FP	PRE	REC	FM	ACC
Random Forest	10	–	0.974	0.025	0.975	0.974	0.974	97.37
Naïve Bayes	10	–	0.964	0.034	0.965	0.964	0.964	96.37
Logistic	10	–	0.909	0.085	0.922	0.909	0.909	90.92
SVM	10	–	0.955	0.042	0.959	0.955	0.955	95.46
SGD	10	–	0.955	0.043	0.957	0.955	0.955	95.46
k-nearest	10	–	0.964	0.037	0.964	0.964	0.964	96.37
J48	10	–	**0.983**	0.017	**0.983**	0.983	0.983	**98.28**
Random Forest	–	20	0.974	0.025	0.974	0.974	0.974	97.37
Naïve Bayes	–	20	0.965	0.033	0.966	0.965	0.965	96.47
Logistic	–	20	0.888	0.105	0.905	0.888	0.887	88.81
SVM	–	20	0.955	0.042	0.959	0.955	0.955	95.46
SGD	–	20	0.957	0.041	0.959	0.957	0.957	95.66
k-nearest	–	20	0.965	0.035	0.965	0.965	0.965	96.47
J48	–	20	**0.983**	0.017	**0.983**	0.983	0.983	**98.28**
Video-based features								
Random Forest	10	–	0.875	0.120	0.885	0.875	0.875	87.5
Naïve Bayes	10	–	0.772	0.243	0.830	0.772	0.760	77.21
Logistic	10	–	0.755	0.255	0.773	0.755	0.749	75.50
SVM	10	–	0.542	0.490	0.757	0.542	0.406	54.23
SGD	10	–	0.746	0.264	0.763	0.746	0.740	74.59
k-nearest	10	–	0.876	0.125	0.876	0.876	0.876	87.60
J48	10	–	**0.898**	0.102	**0.898**	0.898	0.898	**89.81**
Human Consensus-based Feature								
Random Forest	10	–	0.884	0.124	0.904	0.884	0.882	88.40
Naïve Bayes	10	–	0.856	0.154	0.883	0.856	0.852	85.58
Logistic	10	–	0.795	0.218	0.845	0.795	0.786	79.53
SGD	10	–	0.823	0.190	0.868	0.823	0.816	82.25
k-nearest	10	–	0.891	0.111	0.893	0.891	0.891	89.11
J48	10	–	**0.902**	0.104	**0.913**	0.902	0.901	**90.22**
User Profile-based features								
Random Forest	10	–	0.955	0.042	0.959	0.955	0.955	95.46
Naïve Bayes	10	–	0.949	0.049	0.952	0.949	0.949	94.85
Logistic	10	–	0.800	0.187	0.857	0.800	0.794	80.04
SVM	10	–	0.955	0.042	0.959	0.955	0.955	95.46
SGD	10	–	0.949	0.048	0.953	0.949	0.949	94.85
k-nearest	10	–	**0.973**	0.027	**0.973**	0.973	0.973	**97.27**
J48	10	–	0.971	0.030	0.971	0.971	0.971	97.07

nearest, and J48) on a different set of features by applying 10-fold cross-validation. The comparison results in term of accuracy measure, clearly show that the model outperforms when employing all three features combinedly, instead of applying individual features. However, it can also be observed that User-profile features individually perform better in comparison to Human-consensus and Video-content based features. At the same time, Video-content based features do not perform well individually. The experimental results reveal that the user-profile based features significantly improves the overall performance of the proposed model compared to other feature sets. One of the main reason identified from the observation that the reputation of an individual channel/account/user profile plays an important role in identifying the credibility of the uploaded video and gives an efficient clue for the verification of misleading content [16, 37]. Previous studies also reveal that user profile based features are efficient in detecting false content.

Table 8 Performance of the various classifier by employing Percentage Split

Classifiers	Split1	Split2	TP	FP	PRE	REC	FM	ACC
Random Forest	70:30	–	0.977	0.024	0.977	0.977	0.977	**97.65**
Naïve Bayes	70:30	–	0.963	0.037	0.964	0.963	0.963	96.30
Logistic	70:30	–	0.903	0.098	0.907	0.903	0.902	90.26
SVM	70:30	–	0.899	0.102	0.916	0.899	0.898	89.93
SGD	70:30	–	0.956	0.044	0.957	0.956	0.956	95.63
K-nearest	70:30	–	0.970	0.030	0.970	0.970	0.970	96.97
J48	70:30	–	**0.977**	0.023	0.977	0.977	0.977	**97.65**
Random Forest	–	80:20	0.975	0.026	0.975	0.975	0.975	**97.47**
Naïve Bayes	–	80:20	0.929	0.067	0.936	0.929	0.929	92.92
Logistic	–	80:20	0.899	0.105	0.907	0.899	0.898	89.89
SVM	–	80:20	0.949	0.054	0.954	0.949	0.949	94.94
SGD	–	80:20	0.960	0.042	0.961	0.960	0.960	95.95
k-nearest	–	80:20	0.970	0.030	0.970	0.970	0.970	96.96
J48	–	80:20	**0.975**	0.025	0.975	0.975	0.975	**97.47**

The authors of [16, 37–41], reports that the user profile/ account-based features plays a significant role in detecting false information. Whereas, the other two features are not performing well compared to this feature due to some constraints like the video-content based features are relying on the similarity of audio transcripts and the title of the video but there are some cases where the audio is too noisy due to which the clear transcripts can't be retrieved for matching or in a case when the video doesn't have any speech content present. These cases may be liable to degrade the performance of the model. Whereas, human-consensus based features are also

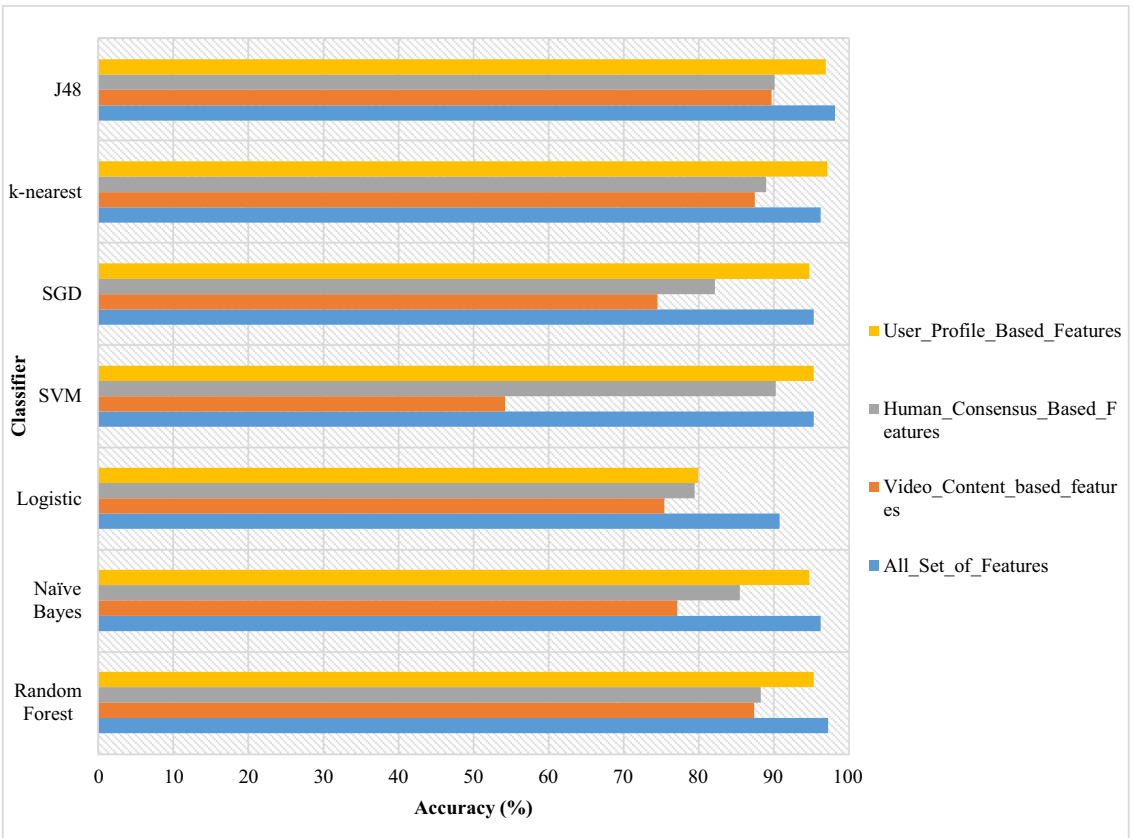

Fig. 9 Comparative analysis of various classifiers on different set of features

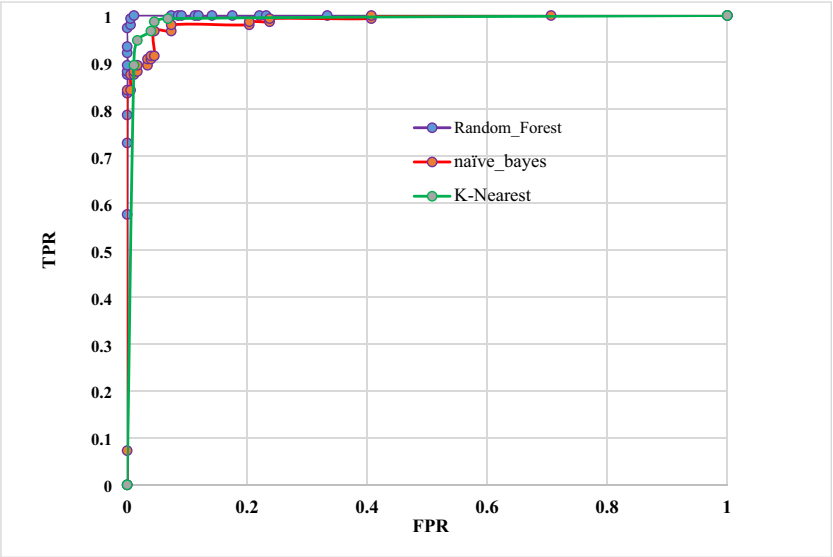

Fig. 10 The AUC-ROC Curve

performing well after user profile features, however in some cases when the sufficient clickbait phrases are not matched/ identified from the user responses or credible link sources in that case this feature may lack in performances.

The AUC-ROC curve using the Random Forest, Naïve-Bayes, and K-nearest neighbor classifier model is shown in Fig. 10. To get a good understanding of the performance of the model, we can also look at their receiver operating characteristics, ROC curves. Figure 10. represents the three ROC curves for the random forest, Naïve Bayes, and K-Nearest Neighbour classifier, trained on all sets of features (video-based, human-consensus, and user-profile based). Here we can see that the peak value for the random forest is achieved at the point (x = 0.01129, y = 1) with having a minimum false positive rate of 0.01. At this point, the model would correctly identify 100% of the true rumors, with only getting around 1% of the false rumors mistakenly classified as true. Whereas, for Naïve Bayes the peak value is achieved at the point (x = 0.401, y = 1), having a minimum false positive rate of 0.406. At this point, the model would correctly identify 100% of the true rumors, with only getting around 40% of the false rumors mistakenly classified as true. Depending upon the application purpose, user can choose or pick up different points on the ROC curve, like with respect to the normal user, who want to find the truthfulness of the specific news, the user could perhaps choose the point on the ROC curve with having minimum false positive rate, or the point where FPR is closer to zero, for getting reliable information.

Along with this, the Scatter plot matrix representation of the features against other features for clickbait and Non-Clickbait data sample is also represented to give a visual qualitative understanding of the correlation. We have created the scatter plot of one feature against another as shown in Fig. 11. Table 9 represents the features on the X and Y-axis of the plot. The plot visually represent the relationship between each of the feature on the set $X = X_1, X_2 \ldots X_9$ on X-axis to the set $Y = Y_1, Y_2 \ldots Y_9$ on Y-axis. The representation is useful, as it is showing the pattern in the relationship between attributes to visually explore the relationship between several numeric values. The dots in the scatter plot are colored by their class value (Clickbait and Non-Clickbait).[8] Like, it can be seen that the scatter plot matrix of subscriber_age_ratio feature on the X-axis against all other features shows an approximately clear separation between two classes (Clickbait's and Non-Clickbait's) and shows how the points are correlated with respect to different classes.

5.2 Comparison of classifiers on publicly available datasets

Along with the self-generated dataset, as we discussed previously, the analysis has also been applied over the other state-of-the-art. It has been observed that some of the work contributed by creating a public dataset of fake/misleading videos [4, 5, 42] . However, still very few datasets are available for comparative analysis, there is small, but some of the datasets of fake videos on YouTube are publicly available called as FVC (Fake Video Corpus) [2] and FVC- 2018 [42]. The dataset[9] is the collection of 381 videos in which 201 are fake, and 180 are Non-Clickbait. After analysis, it has been found that most of the videos are removed from YouTube. Due to

[8] https://machinelearningmastery.com/better-understand-machine-learning-data-weka/

[9] https://github.com/MKLab-ITI/fake-video-corpus/blob/master/FVC.csv

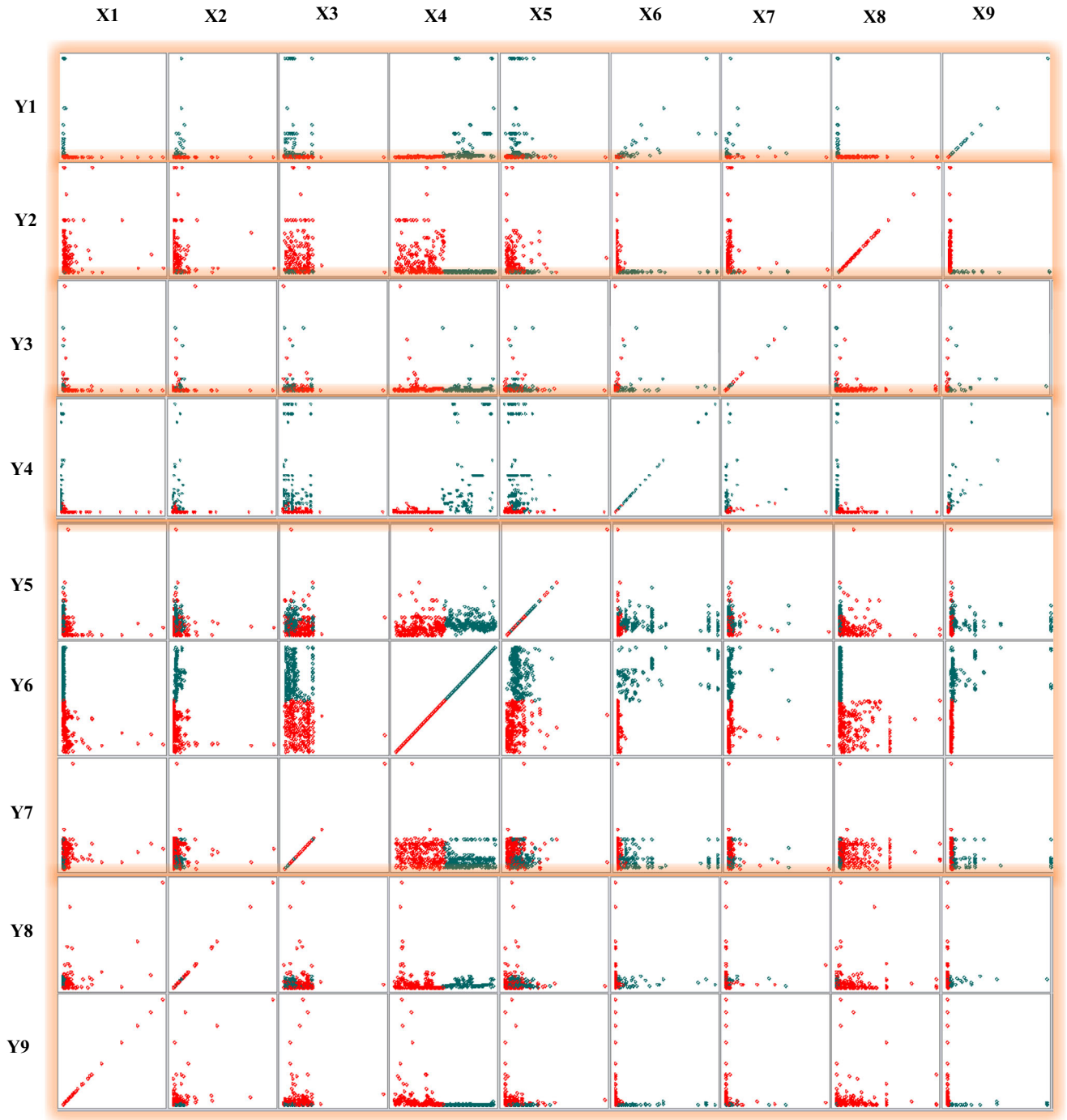

Fig. 11 Plot Matrix representation of proposed features against all other features for Clickbait's (Red) and Non-Clickbait's (Green)

which, we are only able to crawl 84 fake and 90 Non-Clickbait videos, we divide these videos into two disjoint sets, FVC (70:30), with having 70 videos for training and 30 videos for testing. The comparative analysis with the state-of-the-art on the same dataset is shown in Table 10. Whereas, the other dataset that is considered for analysis is FVC 2018. The FVC-2018 is the extended version of the FVC dataset i.e., the samples in the FVC-2018 is an order of magnitude larger than that of FVC, and much more varied. The dataset was extended with 3729 additional fake videos and 2283 real videos, published on YouTube, Facebook or twitter, and considering the time period of April 2006 and June 2018. However for analysis purpose we have considered only YouTube(YT) Videos i.e. 1675(Fake) and 993(Real), comparison analysis has been performed on the same.[10] From

[10] https://github.com/MKLab-ITI/fake-video-corpus/blob/master/FVC_dup.csv

Table 9 Description of features used for the plot of Scatter matrix representation, as shown in Fig. 11

Feature No	Feature	Feature No	Feature
X1	DL(x)	Y1	CV(x)
X2	CS(x)	Y2	FCCR(x)
X3	Avg_cs	Y3	VA(x)
X4	Video_id	Y4	SA(x)
X5	PN(x)	Y5	PN(x)
X6	SA(x)	Y6	Video_id
X7	VA(x)	Y7	Avg_cs
X8	FCCR(x)	Y8	CS(x)
X9	CV(x)	Y9	DL(x)

the previous studies, it has been found that a very limited number of baselines are available for clickbait's video detection [7]. The online clickbait video detection problem is an emerging field and is a largely unsolved research problem. Due to which very few works have been reported yet. Along with this, very limited datasets are publicly available for the evaluation of the proposed algorithm. Many of the works have not released the source code as well. Due to all these research constraints, limited baselines are available for comparative analysis. Some of the prominent methods that are closely related to our work are described in the following paragraph.

The authors of [42], proposed a verification algorithm to detect fake videos. They have also created the FVC-2018 dataset to train and evaluate the proposed method. In the verification algorithm, the author applied the same process that was used in [5], along with it two model variants: a concatenation of the two feature sets(videos metadata and comments feature) and the agreement-based approach given in [43] was used. Their proposed algorithm has been evaluated using 10-fold cross-validation on the dataset proposed by Papadopoulos [5]. and the FVC 2018 dataset with an F1 score of 0.85 and 0.69 respectively. Whereas, the authors of [5], build a classification model using two sets of features(video metadata and comments). Video metadata that specifically considered linguistic features extracted from the video description text and statistics extracted from the video channel. Whereas, the second feature is based on the comments by incorporating a two-level approach. In the first level, features

Table 10 Comparative Analysis with the State-of-the-art

Method	Classifier	Split/ fold	PRE	REC	FM	Dataset
[42] 2019	SVM(Video feature)	10 fold	0.88	0.79	0.82	FVC
	SVM(Comment Feature)	10 fold	0.88	0.74	0.79	FVC
	SVM(Fusion)	10 fold	0.88	0.82	0.85	FVC
	SVM(Video feature)[YT]	10 fold	0.87	0.59	0.70	FVC-2018
	SVM(Comment Feature)[YT]	10 fold	0.91	0.53	0.67	FVC-2018
	SVM(Fusion)[YT]	10 fold	0.79	0.61	0.69	FVC-2018
[5] 2017	SVM (Video Feature)	10 fold	0.88	0.79	0.82	FVC
	SVM (Comment Feature)	10 fold	0.88	0.74	0.79	FVC
	SVM RBF(Fusion)	10 fold	1.00	0.83	0.90	FVC
[44] 2019	Random Forest	70:30	0.74	0.73	0.73	FVC
	Decision- Tree	70:30	0.73	0.67	0.67	FVC
	SVM	70:30	0.56	0.55	0.54	FVC
	Logistic Regression	70:30	0.53	0.53	0.53	FVC
	UCNet	70:30	0.82	0.82	0.82	FVC
Our Method	Random Forest	70:30	0.84	0.78	0.77	FVC
	Decision- Tree	70:30	0.77	0.75	0.73	FVC
	SVM	70:30	0.65	0.63	0.63	FVC
	Logistic Regression	70:30	0.65	0.65	0.65	FVC
	SVM (Video Feature)	10 fold	0.87	0.83	0.83	FVC
	SVM (Comment Feature)	10 fold	0.87	0.83	0.83	FVC
	SVM (Fusion)	10 fold	0.87	0.85	0.85	FVC
	SVM(Video Feature)[YT]	10 fold	0.57	0.57	0.57	FVC-2018
	SVM(Comment Feature)[YT]	10 fold	0.57	0.57	0.57	FVC-2018
	SVM (Fusion)[YT]	10 fold	0.69	0.69	0.69	FVC-2018

are extracted from the individual comment, later the credibility of each comment is evaluated independently using a pre-trained model proposed by the authors of [43]. These two sets of features are used to train the support vector machine classifier. The algorithm is evaluated using 10 fold cross-validation on their proposed dataset with an F1 score of 0.90 on the fusion of both the features. The other comparative analysis has been done concerning the algorithm proposed by the authors of [44], the method is evaluated on 70:30 percentage split scheme, where it has been observed that our proposed work outperforms the method given in [44], considering all three measures(Precision, Recall, and F-Measure) by employing FVC dataset. In [44], the author has proposed an algorithm to counter misleading videos as a supervised classification task. A deep learning-based approach UCNet has been developed, along with it, some simple features are used for the detection of fake videos. It can be seen that the Decision-Tree, SVM, and Logistic Regression classifiers on the proposed approach outperforms the state-of-the-art, except the Random forest classifier. From the above study, it has been observed that most of the reported work mainly employed video metadata and comments based features for the prediction of clickbait videos, however to the best of our knowledge, none of the above-mentioned work has included video related features including video transcript as well as not discussed the similarity among the video title and its content (video transcript), due to which unable to identify how faithfully the video representing the text it claiming to. It has been found from the results outcomes that transcript based features are efficient in improving the model performance and also helps in identifying certain clues about the video credibility that whether it is faithfully representing the same as it claims to. Along with this some of the crucial and novel features are proposed concerning to different feature categories video, comments, and channel that jointly helps in achieving efficient results as described in Section 4.2. The comparative analysis is shown in Table 1; from Table 1 it can be observed that the proposed algorithm outperforms the existing state-of-the-art methods. The proposed approach outperforms the method proposed by [5, 42], concerning both Recall and F-score on 10 fold cross-validation scheme over the FVC dataset. The comparison has been applied to video, comment, and all set of features(fusion), where the proposed model outperforms existing work. On the other hand, while considering the FVC-2018 dataset, it performs better with respect to recall. From the comparative analysis shown in Table 10, the SVM classifier performs better than the previous approach with respect to recall and f-measure and approximate similar with respect to precision when considering the FVC dataset. In addition to this, the SVM also performs well on our proposed dataset (MVD) with an accuracy of 95.4% on 10-fold cross validation as shown in Table 7. However, in case of FVC-2018 dataset, it is less effective on video-based and comment based features individually, whereas by applying fusion of features the classifier performs better with respect to recall and f-score compared to previous approaches. The reason for worse

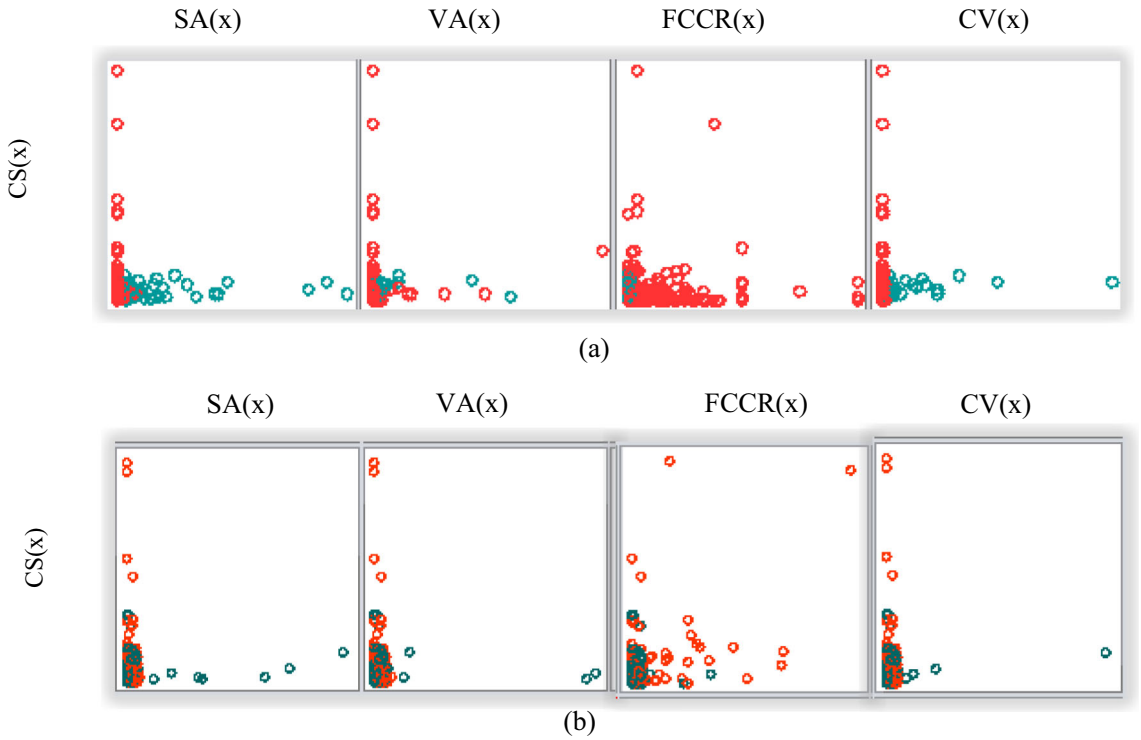

Fig. 12 Plot Matrix representation of features CS(x) against four other features on **a** MVD and **b** FVC-2018 dataset for Clickbait's (Red) and Non-Clickbait's (Green)

performance of the SVM classifier on the FVC-2018 dataset is existence of noise in the feature data. The study also reports that the SVM doesn't perform very well, when the data set has more noise i.e. target classes are overlapping that leads to misclassification of samples. The visiulaization of feature data is as shown in Fig. 12, where scatter plot matrix is presented as an example for some features to show the distribution of target samples on MVD and FVC-2018 Dataset.

The channel view to subscriber ratio(CS(x)) feature has been visualized w.r.t to four other features (subscriber_to_age_ratio SA(x), video_age_ratio VA(x), fake_comment_count_ratio FCCR(x) and channel views CV(x)) to see the distribution of target sample points. From the Fig. 12, it can be clearly noted that the samples of target class are noisy and overlapping in case of FVC-2018 dataset reported in Fig. 12b in comparison to MVD dataset reported in Fig. 12a and due to which we can not be able to get a better decision boundary for classification, and many of the real samples are misclassified as fake, as a result of which it may not end up performing well.

6 Conclusion and future work

In this paper, we develop the CVD scheme to detect clickbait video. The scheme leverages on three components for learning three sets of latent features based on User Profiling, Video-Content, and Human Consensus that further be used to retrieve three sets of cognitive evidence, as an innovative idea for the detection of clickbait videos on YouTube. The set of features are given as an input to machine learning model and performance has been analyzed by considering all set of features and each feature independently by employing a different set of the classifier, and it has been observed that J48 outperforms all other with an accuracy of 98.89% by applying all set of features using cross-validation technique, while 97.47% using percentage split technique on the self-generated dataset [MVD]. The proposed method also performs well on the FVC, FVC-2018 dataset, and outperforms the state-of-the-art with an improved Recall and F score. From the analysis, it has been observed that non-clickbait's video channel has more number of subscribers with respect to registration age as compare to clickbait's video channel. The findings reveal that the average cosine similarity for Non-Clickbait videos is too less than that of the clickbait's, the one of the reason may be because the fake video probably repeating the same sentence as mentioned in the title, of many times to extend the length of the video with redundant and bogus content.

Further work can be enhanced by generating large datasets and employing more video-related features like image frames from the video to get more efficient clues, as well as the clickbait's headlines, can also be analyzed for different other applications like at the time of natural disasters, political elections, healthcare, etc. Along with this, we further extend the work by creating a Non-Clickbait-time application of it.

References

1. Sisodia DS (2019) Ensemble learning approach for Clickbait detection using article headline features. Informing Sci Int J an Emerg Transdiscipl 22:31–44
2. Papadopoulou O, Zampoglou M, Papadopoulos S, Kompatsiaris Y, & Denis Teyssou. (2018). InVID Fake Video Corpus v2.0 (Version 2.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.1147958
3. Zhang DY, Song L, Li Q, Zhang Y, Wang D (2018) Streamguard: A bayesian network approach to copyright infringement detection problem in large-scale live video sharing systems. In 2018 IEEE International Conference on Big Data, Seattle, pp 901–910. https://doi.org/10.1109/BigData.2018.8622306
4. Papadopoulos SA (2019) Towards Automatic Detection of Misinformation in Social Media. arXiv:1909.01543. https://export.arxiv.org/pdf/1909.01543
5. Papadopoulou O, Zampoglou M, Papadopoulos S, Kompatsiaris Y (2017) Web video verification using contextual cues. In Proceedings of the 2nd International Workshop on Multimedia Forensics and Security, pp 6–10. https://dl.acm.org/doi/10.1145/3078897.3080535
6. Potthast M, Köpsel S, Stein B, Hagen M (2016) Clickbait detection. In: Ferro N. et al. (eds) Advances in Information Retrieval. European Conference on Information Retrieval (ECIR) 2016. Lecture Notes in Computer Science, vol 9626. Springer, Cham, pp 810–817. https://doi.org/10.1007/978-3-319-30671-1_72
7. Shang L, Zhang DY, Wang M, Lai S, Wang D (2019) Towards reliable online clickbait video detection: a content-agnostic approach. Knowledge-Based Syst 182:104851
8. Zannettou S, Sirivianos M, Blackburn J, Kourtellis N (2019) The web of false information: rumors, fake news, hoaxes, clickbait, and various other shenanigans. J Data Inf Qual 11(3):1–37
9. Qu J, Hißbach AM, Gollub T, Potthast M (2018) Towards Crowdsourcing Clickbait Labels for YouTube Videos. In HCOMP (WIP&Demo). http://ceur-ws.org/Vol-2173/paper11.pdf
10. Chakraborty A, B. Paranjape, S. Kakarla, and N. Ganguly (2016) Stop clickbait: Detecting and preventing clickbaits in online news media. In IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. pp 9–16. https://ieeexplore.ieee.org/abstract/document/7752207/
11. Chen Y, Conroy NJ, Rubin VL (2015) Misleading online content: recognizing clickbait as false news. In Proceedings of ACM on workshop on Multimodal Deception Detection. pp 15–19. https://dl.acm.org/doi/10.1145/2823465.2823467
12. Biyani P, Tsioutsiouliklis K, Blackmer J (2016) 8 amazing secrets for getting more clicks: detecting clickbaits in news streams using article informality. In Thirtieth AAAI Conference on Artificial Intelligence. https://dl.acm.org/doi/10.5555/3015812.3015827
13. Anand A, Chakraborty T, Park N (2017) We used neural networks to detect clickbaits: You won't believe what happened next! In European Conference on Information Retrieval, pp 541–547. https://doi.org/10.1007/978-3-319-56608-5_46
14. Thomas P (2017) Clickbait identification using neural networks. arXiv preprint arXiv:1710.08721. https://arxiv.org/pdf/1710.08721.pdf
15. Zannettou S, Chatzis S, Papadamou K, & Sirivianos M (2018) The good, the bad and the bait: Detecting and characterizing clickbait on youtube. In IEEE Security and Privacy Workshops, pp 63–69. https://doi.org/10.1109/SPW.2018.00018

16. Yang F, Liu Y, Yu X, Yang M (2012) Semantics, pp 1–7. https://doi.org/10.1145/2350190.2350203
17. Wu K, Yang S, Zhu KQ (2015) False rumors detection on sina weibo by propagation structures. In IEEE International Conference on Data Engineering, pp 651–662. https://doi.org/10.1109/ICDE.2015.7113322
18. Zhang Q, Zhang S, Dong J, Xiong J, Cheng X (2015) Automatic detection of rumor on social network. In Natural Language Processing and Chinese Computing, pp 113–122. https://doi.org/10.1007/978-3-319-25207-0_10
19. Qin Y, Wurzer D, Lavrenko V, Tang C (2016) novelty detection. arXiv preprint arXiv:1611.06322. https://arxiv.org/abs/1611.06322
20. Liu X, Nourbakhsh A, Li Q, Fang R, Shah S (2015) Conference on Information and Knowledge Management, pp 1867–1870. https://doi.org/10.1145/2806416.2806651
21. Wu L, Li J, Hu X, Liu H (2017) Gleaning wisdom from the past: Early detection of emerging rumors in social media. In Proceedings of the 2017 SIAM international conference on data mining, pp 99–107. https://doi.org/10.1137/1.9781611974973.12
22. Kwon S, Cha M, Jung K, Chen W, Wang Y (2013) Conference on Data Mining, pp 1103–1108. https://doi.org/10.1109/ICDM.2013.61
23. Kwon S, Cha M, Jung K (2017) Rumor detection over varying time windows. PLoS One 12:1–19
24. Ma J, Gao W, Mitra P, Kwon S, Jansen B J, Wong K F, Cha M (2016) Detecting rumors from microblogs with recurrent neural networks. In Proceedings of the 25th International Joint Conference on Artificial Intelligence, pp 3818–3824. https://ink.library.smu.edu.sg/sis_research/4630
25. 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 preprint arXiv:1704.07506. https://arxiv.org/abs/1704.07506
26. Kumar S, West R, Leskovec J (2016) Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes. In Proceedings of the 25th international conference on World Wide Web, pp 591–602. https://doi.org/10.1145/2872427.2883085
27. Zaman B, Justitia A, Sani KN, Purwanti E (2020) An Indonesian hoax news detection system using reader feedback and Na{\"\i}ve Bayes algorithm. Cybern Inf Technol 20(1):82–94
28. Cao J, Qi P, Sheng Q, Yang T, Guo J Li J (2020) Exploring the Role of Visual Content in Fake News Detection. arXiv preprint arXiv:2003.05096. https://arxiv.org/abs/2003.05096
29. Zhou X, Wu J, Zafarani R (2020) SAFE: Similarity-aware multi-modal fake news detection. arXiv preprint arXiv:2003.04981. https://arxiv.org/abs/2003.04981
30. Ozbay FA, Alatas B (2020) Fake news detection within online social media using supervised artificial intelligence algorithms. Phys A Stat Mech its Appl 540:123174
31. Agarwal A, Mittal M, Pathak A, Goyal LM (2020) Fake news detection using a blend of neural networks: an application of deep learning. SN Comput Sci 1:1–9
32. 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. https://doi.org/10.1016/j.cogsys.2019.07.004
33. Meel P, Vishwakarma DK (2020) Fake news, rumor, information pollution in social media and web: a contemporary survey of state-of-the-arts, challenges and opportunities. Expert Syst Appl 153:1–26. https://doi.org/10.1016/j.eswa.2019.112986
34. Bondielli A, Marcelloni F (2019) A survey on fake news and rumour detection techniques. Inf Sci (Ny) 497:38–55. https://doi.org/10.1016/j.ins.2019.05.035
35. Kumar S, Shah N (2018) False information on web and social media: A survey. arXiv preprint arXiv:1804.08559. https://arxiv.org/abs/1804.08559
36. Zhou X, Zafarani R (2018) A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. ACM Comput Surv 53(5):101–109. https://doi.org/10.1145/3395046
37. Qazvinian V, Rosengren E, Radev DR, Mei Q (2011) misinformation in microblogs. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, pp 1589–1599. https://dl.acm.org/doi/10.5555/2145432.2145602
38. Shu K, Sliva A, Wang S, Tang J, Liu H (2017) Fake news detection on social media: a data mining perspective. ACM SIGKDD Explor Newsl 19(1):22–36
39. Shu K, Wang S, Liu H (2018) Information Processing and Retrieval, pp 430–435. https://doi.org/10.1109/MIPR.2018.00092
40. Shu K, Zhou X, Wang S, Zafarani R, Liu H (2019) The role of user profiles for fake news detection. In Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp 436–439. https://dl.acm.org/doi/10.1145/3341161.3342927
41. Shu K et al (2018) Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media. Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining 8:430–435
42. Papadopoulou O, Zampoglou M, Papadopoulos S, Kompatsiaris I (2019) A corpus of debunked and verified user-generated videos. Online information review 43(1):72–88. https://doi.org/10.1108/OIR-03-2018-0101
43. 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
44. Palod P, Patwari A, Bahety S, Bagchi S, Goyal P (2019) Misleading Metadata Detection on YouTube. In European Conference on Information Retrieval, pp 140–147. https://doi.org/10.1007/978-3-030-15719-7_18

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

Deepika Varshney is a Research Scholar inthe Department of Information Technology,Delhi Technological University India. She has done M.Tech. from IndiraGandhi Delhi Technical University for Women, NewDelhi, India 2017. Her research interest includes OnlineSocial Media Privacy and Security, Machine Learning, and DataScience. She has been awarded with a commendable researchaward in the year 2020 by the Delhi Technological University.

Dinesh Kumar Vishwakarma received the B.Tech. degree from Dr. Ram Manohar Lohia Avadh University, Faizabad,India, in 2002, the M.Tech. degree from the Motilal NehruNational Institute of Technology, Allahabad, India, in 2005,and the Ph.D. degree in the field of Computer Vision and Machine Learning from Delhi Technological University (Formerly Delhi College of Engineering), New Delhi, India, in 2016. He iscurrently an Associate Professor with the Department ofInformation Technology, Delhi Technological University, New Delhi. His current research interests includeComputer Vision, Machine Learning, DeepLearning, Sentiment Analysis, Fake News and Rumour Analysis, Crowd Behaviour Analysis,Person Re-Identification, Human Action and Activity Recognition. He is a reviewer of variousJournals/Transactions of IEEE, Elsevier, and Springer. He has been awarded with "ResearchExcellence Award" by Delhi Technological University, Delhi, India in 2018, 2019 and2020. He is Senior Member of IEEE and life member of ISTE.