Journal of Network Intelligence
Taiwan Ubiquitous Information
 ISSN 2414-8105
Volume **8**, Number **3**, August **2023**

An Investigation of Contextual Features for Misleading Video Detection

Xiao-Jun Li[1], Shuai Fan[1*], Jun-Ping Yao[1], Shao-Chen Li[1]
[1]Xi'an Research Institute of High-Tech, 710000, Xi'an, China
xi_anlxj@126.com, xi_anfs@126.com, junpingy200225@163.com, lishaochen11112@163.com
*Corresponding author: Shuai Fan
Received December 1, 2022, revised February 5, 2023, accepted April 4, 2023.

Abstract. *As videos gain increasing popularity on social media, misleading videos are becoming a major challenge for the social media platforms. Though occasionally used in the misinformation detection, contextual cues are seldomly proposed in works on video detection and evaluated systematically. To address this research gap, we proposed a comprehensive set of contextual features (i.e., similarity range, similarity average, watch frequency, inverted likes, coin ratio, favorites ratio, forwards, bullet comment ratio, and watch review ratio) and evaluated their effect on the detection performance of misleading videos based on a naïve Bayesian model. The results show that the proposed contextual features are effective to fulfill the detection task and achieved an F1-score of 0.81. We also compared our model with other baselines, and found that it outperformed other baselines such as the support vector machine (SVM), K-nearest neighbors (k-NN), decision tree and random forest.*
Keywords: Contextual features; Naïve Bayesian model; Misleading video; Misinformation

1. **Introduction.** Misleading videos are becoming a major threat to video sharing websites. Close to 40% of healthcare videos released since 2019 on social media are found to contain misleading information or biases [1,2]. Such videos can endanger people's health, lead to missing of the best timing for medical treatments, and result in irreparable damage [3]. Automatically detecting misleading videos is aimed at eliminating the harm from misinformation for the viewers. Therefore, how to detect misleading videos automatically remains a challenge among researchers.

There are two main types of misleading videos we might come across on the Internet. The first type shows something that has really happened, but is mislabeled for political, advertising or commercial purposes. The second type refers to those that are not real, either because they have been staged or digitally doctored. In this study, we focus on detection of the first type of misleading videos (e.g., exaggerated health advertisements claiming to cure cancer). Since it is difficult to detect the truthfulness of claims or opinions embedded in the videos directly, contextual cues are more appropriate for this task. Therefore, we focus on different contextual cues and evaluate their effectiveness on detection of misleading videos.

However, the contextual cues are less investigated in the literature about misleading video detection, Wang et al. [4] used deep learning to understand image content and generate descriptive text. Among the few studies on contextual cues, Hussain et al. [5] used the cues from the comments to identify misleading videos. Inspired by this idea, we attempted to investigate how other contextual cues can be used to detect misleading videos. Therefore, we proposed several contextual cues and compared their effectiveness on detecting misleading videos.

The major contributions of our work are as follows. Firstly, nine contextual features were proposed and validated for the misleading video detection task. Secondly, we found that the best performance was achieved when all those nine features were considered. Thirdly, we proposed a misleading video detection model based on the naïve Bayesian model and demonstrated that it outperformed all the other baselines including naïve SVM, k-NN, decision tree and random forest.

The remainder of this paper are organized as follows. Section 2 reviews the related work. Section 3 introduces the features and the model in detail. The experiments and results are described in Section 4, followed by the conclusions in Section 5.

2. **Related Work.** In this section, we review the following related topics: misleading video detection; fake news detection, spam video detection, and naïve Bayesian classifier in fake content detection.

2.1. **Misleading Video Detection.** There are mainly two types of methods for misleading video detection: machine learning methods and deep learning methods.

Machine learning methods have been popular in works on misleading video detection. For example, Syed-Abdul et al. [6] studied videos related to anorexia on YouTube, and found that social and textual clues could be used to automatically identify pro-anorexia contents. Hou et al. [7] developed a classification model that could classify the videos based on linguistic, acoustic, and user engagement features to identify misleading videos related to prostate. Ghenai et al. [8] constructed a classifier to identify users who tended to spread misleading videos by extracting features of user attributes, writing style, and sentiment.

Deep learning models have also been widely used in misleading video detection. For example, Palod et al. [9] constructed a deep learning network model which was found to outperform other baseline models and have good generality. Liu et al. [10] constructed a bidirectional LSTM (long short-term memory) model to classify the level of medical knowledge in videos based on word vector representations extracted from video descriptions with good results. Alsaeedi et al. [11] proposed a deep learning model based on traditional neural networks (CNN) for detecting misleading information on Twitter.

2.2. **Fake News Detection.** Fake news is news or stories created to deliberately misinform or deceive readers. In recent years, researchers have done a lot of work for the detection of fake news. The literature review shows that both traditional machine learning approaches and deep learning approaches have been used for fake news detection. For the traditional machine learning approach, Mahabub [12] proposed an integrated voting classifier for detection of fake news. Shah and Kobti [13] demonstrated a new approach to detect fake news using cultural algorithms with situational knowledge and non-fuzzy knowledge. Ahmad et al. [14] used integration techniques and Linguistic Inquiry and Word Count (LIWC) features to build a fake news detection model, which showed superior performance during a comparison with models such as support vector machines and convolutional neural networks on four real data sets.

For the deep learning approach, Nasir et al. [15] proposed a hybrid deep learning model that combined convolutional and recurrent neural networks for fake news classification. Ajao et al. [16] proposed a framework for detecting and classifying fake news messages from Twitter posts using convolutional neural networks and long-term recurrent neural network models. Agarwal et al. [17] used embedded words in text preprocessing to construct word vector spaces and establish linguistic relationships, which combined convolutional neural networks and recurrent neural network architecture to detect fake news.

2.3. **Spam Video Detection.** Spam videos are video content that is excessively posted, repetitive, or untargeted. Most works on spam video detection also rely on machine learning approach and deep learning approach.

For machine learning approach, Ashar et al. [18] proposed a method to classify videos as spam or legitimate videos based on YouTube video attributes, and found that certain linguistic features (the presence of certain terms in the title or description of YouTube videos) and temporal features can be used to predict video types. Kanodia et al. [19] proposed a Markov decision process approach to model the YouTube spam video detection problem. Aggarwal et al. [20] found that video language features, videos popularity, length of videos and videos category can be used to predict video types.

For deep learning approach, Seth and Biswas [21] used convolutional neural networks to classify emails independently to identify spam mails. Araujo et al. [22] proposed a very compact video classification model based on state-of-the-art network architecture. He et al. [23] demonstrated a form of deep learning, an architecture of linguistic attributes embedded in linguistic decision trees, which can improve the performance of spam detection.

2.4. **Naïve Bayesian Classifier in Fake Content Detection.** The plain Bayes classifier has been widely used in fake content detection. For example, Abdullah-All-Tanvir et al. [24] proposed a model to identify false news in the Twitter dataset and found that the plain Bayes model outperforms the five other machine learning algorithms. Granik and Mesyura [25] used the plain Bayes classifier to detect fake news in Facebook and finally achieved a classification accuracy of 74%. Reshmi et al. [26] proposed a method to detect fake news by using plain Bayes classifier during COVID19 outbreaks. Compared to other models, the plain Bayes model has lower model complexity while ensuring detection effectiveness. In addition, the Bayes classifier is not demanding in terms of the sample size, therefore it is suitable for research with a medium and small sample size. Therefore, the Naïve Bayes model is widely used in fake content detection.

In summary, most works on disinformation detection focus on the content rather than contextual cues, and despite the few exceptions, contextual cues have seldomly been analyzed in detail or evaluated systematically. For example, what are the specific variables of contextual cues? What are the effects of these different contextual cues on misleading video detection? These questions have not been systematically investigated before. To fill the research gap, we propose nine contextual features and test their effectiveness on misleading video detection in the present work.

3. **Features and Model.** This section introduces the contextual features and the proposed model for misleading video detection.

3.1. **Contextual Features.** In this study, we propose nine contextual features for misleading video detection. They are similarity range of video reviews, similarity average of video reviews, watch frequency, video's appeal, coins, likes and add-to-favorites, likes and shares (forwards), bullet comment-review ratio, and watch-review ratio. The construction of contextual features is based on original variables (e.g., views, bullet comments, upload time, viewers, thumbs-ups (likes), coins, favorites, forwards) on a social media platform, as shown in Figure 1. (1) Similarity Range (SR) measures the range of similarity scores for video reviews. Misleading videos are always associated with review manipulation because well-rated videos are more popular on media sharing sites [27]. Since the misleading videos are usually made intentionally for some political or commercial purpose, the uploaders

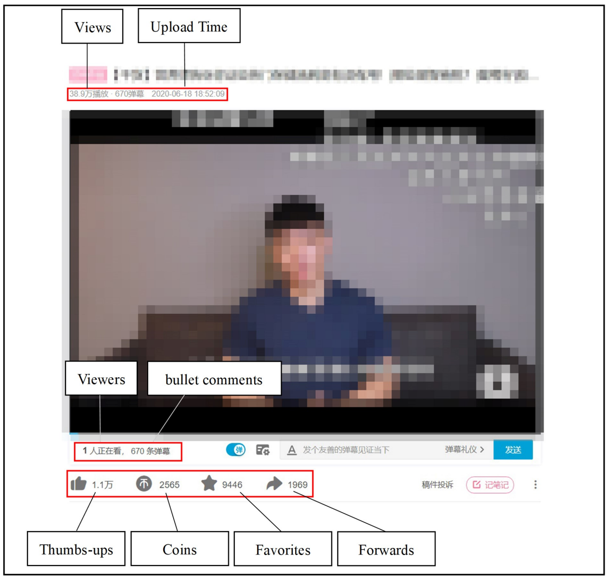

Figure 1. Some source of contextual features in a social medial platform

are likely to manipulate the video reviews. In addition, the manipulation of reviews can be detected by the similarity of reviews. For example, Day et al. [28] found that the similarity of reviews can be used to differentiate legitimate reviews and fake reviews. Inspired by this idea, we identify similarity range of reviews as contextual feature.

In this study, the SR of video reviews are calculated based on three most popular reviews $FV_{reviews}$, $FV_{review2}$, and $FV_{review3}$. We only consider the three most popular reviews because they represent the mainstream user feedbacks. Specifically, each review is represented as a word vector.

(1) We first calculate the similarity between the reviews. The SR is then obtained as the difference between the maximum similarity and minimum similarity.

$$S_{cos1} = cosine(PR_1, PR_2) \tag{1}$$

$$S_{cos2} = cosine(PR_1, PR_3) \tag{2}$$

$$S_{cos3} = cosine(PR_2, PR_3) \tag{3}$$

$$SR = Max_{(S_{\cos 1}, S_{\cos 2}, S_{\cos 3})} - Min_{(S_{\cos 1}, S_{\cos 2}, S_{\cos 3})} \tag{4}$$

where $FV_{comment1}$ stands for the hottest review, $FV_{comment2}$ is the second hottest review, and$FV_{comment3}$ is the third hottest review.

(2) Similarity Average (SA) measures the average of similarity scores among video reviews. A review is not likely to share high similarity with other reviews if not manipulated by the same person. In most cases, the first three hottest reviews represent the dominating views for a video. Therefore, SA is calculated based on the average similarity score among the three hottest reviews:

$$SA = \frac{(S_{cos1} + S_{cos2} + S_{cos3})}{3} \tag{5}$$

(3) Watch Frequency (WF) refers to the number of views that a video wins. Misleading videos often spread faster and wider than legitimate videos, so the former has more views than the latter. The calculation of WF is shown in the following equation:

$$WF = \frac{FV_{play\_times}}{|T_1 - T_0|} \tag{6}$$

where FV_{play_times} represents the number of times the video has been watched, T_1 means the release time of the video, and T_0 is the time when the crawler collects the video.

(4) Inverted Likes (IL) of a video measures the ratio of play times to the likes. It is easy to get a high number of plays, but it is not so easy to get a high number of likes. This is because on most video sharing platforms, you do not need to log in to "play", but you need to log in to "like". In addition, an account can only press the like button once. As a result, the scarcity of likes is further exacerbated. Therefore, the ratio of the play times to the likes of the manipulated video is usually greater. The calculation of IL is shown in the following equation:

$$IL = \frac{FV_{play\_times}}{FV_{likes}} \tag{7}$$

where FV_{likes} is the likes a video has won.

(5) Coin Ratio (CR) measures the probability of a video to win coins from watchers. Similar to the likes, a watcher can give a coin to a video once if he/she holds a positive attitude toward the video. However, giving a coin is much more costly than giving a like because coins need to get paid. As a result, misleading videos should have less opportunities to win coins than legitimate ones. The calculation of CR is shown in the following equation:

$$CR = \frac{FV_{likes}}{FV_{coins}} \tag{8}$$

(6) Favorites Ratio (FaR) of a video refers to how many users add the video to favorites. Adding a video to favorites is also more costly than giving a like because a video added to the “favorites” folder stay much longer with the user than those that win only a “like”. Many users are not willing to add a video to favorites if they do not really like it. The calculation of FaR is shown in the following equation:

$$FaR = \frac{FV_{likes}}{FV_{favorites}} \tag{9}$$

where $FV_{favorites}$ is the number of users who have added the video to their favorites.

(7) Forwards (FoR) refers to how many times a video is forwarded. Misleading videos often need to be forwarded more than the legitimate videos to reach a wider audience, and this can be fulfilled by using bots or providing rewards for video watchers. The calculation of FoR is shown in the following equation:

$$FoR = \frac{FV_{likes}}{FV_{forwards}} \tag{10}$$

where $FV_{forwards}$ is the number of times that a video has been forwarded.

(8) Bullet Comment-Review Ratio (BRR) measures the ratio of the number of bullet comments number to the number of reviews. Normally, a bullet comment is sent by a user who is watching or has watched the video. For a misleading video, however, this is not necessarily true since the goal of the video uploader is to mislead the audience. Thus, the BRR can be used as an indicator to discriminate misleading videos from legitimate ones. The calculation of BRR is shown in the following equation:

$$BRR = \frac{FV_{bulletcomments}}{FV_{reviews}} \tag{11}$$

where $FV_{bulletcomments}$ is the number of bullet comments, and $FV_{reviews}$ is the number of reviews.

(9) Watch Review Ratio (WRR) is the ratio of number of play times to the number of reviews. In most cases, many reviews of a misleading video are posted by users who

have never watched the video, while the watcher of a legitimate video may not post any review even though he/she has watched the video. Therefore, there should be a significant difference between misleading videos and legitimate videos on WRR. The calculation of WRR is shown in the following equation:

$$WRR = \frac{FV_{play_t imes}}{FV_{reviews}} \tag{12}$$

3.2. **Modeling.** The contextual cues-based misleading video detection framework proposed in this study is shown in Figure 2. The nine features mentioned above are generated in the feature extraction step. In this study, the naïve Bayesian model is employed as the

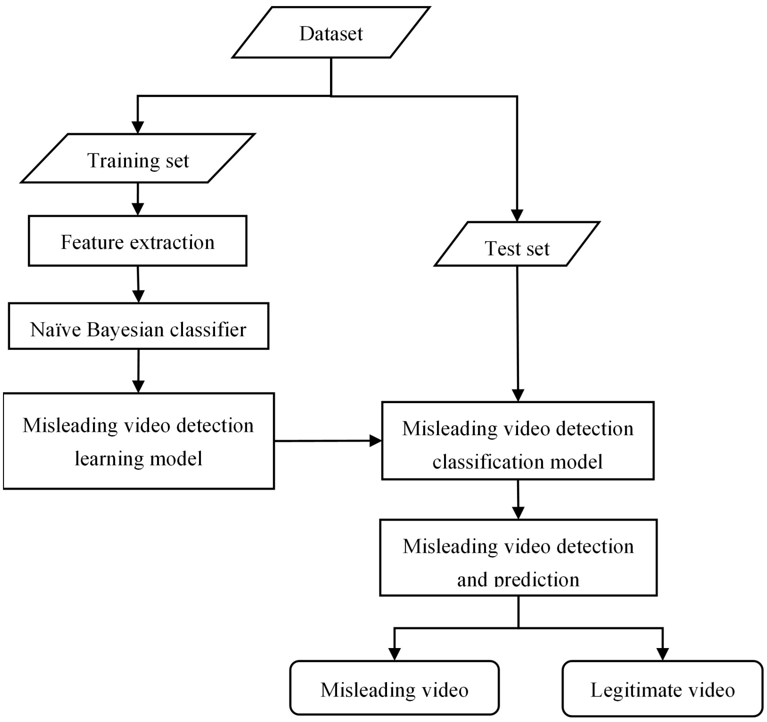

Figure 2. Contextual cues-based misleading video detection framework

classifier. Specifically, we assume a dataset $X = \{x_1, x_2, \ldots x_{n-1}, x_n\}$ is an unclassified set of misleading videos and legitimate videos, where $x_n = \{a_1, a_2 \ldots a_{n-1}, a_n\}$ represents a video in the set X, and $a_1, a_2 \ldots a_{n-1}, a_n$ are the values of contextual features of a video. The set $Y = \{y_1, y_2\}$ is the set of types of videos, where y_1 means a misleading video, and y_2 means a legitimate video.

As per the Bayes theorem, the following equation is obtained:

$$P\left(y_i|x_n\right) = \frac{P\left(x_n|y_i\right) P\left(y_i\right)}{P\left(x_n|y_1\right) P\left(y_1\right) + P\left(x_n|y_2\right) P\left(y_2\right)}, i = 1, 2; n = 1, 2, \ldots n \tag{13}$$

Thus, when $P(y_1|x_n) > P(y_2|x_n)$, the video x_n is defined as a misleading video; and when $P(y_1|x_n) < P(y_2|x_n)$, the video x_n is defined as a legitimate video.

In the Bayesian model, contextual features are independent from each other, so $P(x_n|y_1)P(y_1)+P(x_n|y_2)P(y_2)$ is a constant. Meanwhile, as the number of legitimate and misleading videos in a dataset is fixed in a given dataset, $P(y_i)$ is also a constant. When $P(x_n|y_i)$ reaches the maximum, $P(y_i|x_n)$ reaches the maximum as well, i.e.,

$$\max_{0<x\leq 2} P\left(x_n|y_i\right) = \max_{0<x\leq 2} \Pi_1^n P\left(a_n|y_i\right) \tag{14}$$

In construction of the naïve Bayesian model, the prior distribution can be defined based on the a-priori probabilities to obtain the posterior distribution, which facilitates classifying the prediction target to a class. However, there is no a-priori information available about the falsehood of a video before detection. Thus, the Jeffreys' prior is employed to determine prior distribution.

For a video contextual feature a, the probability that it corresponds to a class in the set Y conforms to binominal distribution. Thus, we assume Z conforms to binominal distribution $B(n, \theta)$, and the following is obtained:

$$P(Z = a_z) = C_n^z \theta^z (1-\theta)^{n-z}, z = 1, 2, \ldots, n \tag{15}$$

The logarithm of the likelihood function of Eq. (13) is:

$$L = zln\theta + (n-z)ln(1-\theta) + lnC_n^z \tag{16}$$

Hence,

$$\frac{\partial^2 L}{\partial L^2} = -\frac{z}{\theta^2} - \frac{n-z}{(1-\theta)^2} \tag{17}$$

$$I(\theta) = E\left(-\frac{\partial^2 L}{\partial L^2}\right) = \frac{n}{\theta} - \frac{n}{1-\theta} = \frac{n}{\theta(1-\theta)} \tag{18}$$

Therefore, as per Jeffreys' theory, the density function of the noninformative prior distribution of θ is:

$$\pi(\theta) = [I(\theta)]^{\frac{1}{2}} \alpha \theta^{-\frac{1}{2}} \left(1 - \theta^{-\frac{1}{2}}\right) \tag{19}$$

That is, the noninformative prior distribution of θ is the Beta distribution $Be\left(\frac{1}{2}, \frac{1}{2}\right)$.

In a Beta distribution $Be(\alpha, \beta)$,

$$E(X) = \frac{\alpha}{\alpha + \beta} \tag{20}$$

$$Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{21}$$

thus, $E(\theta) = \dfrac{1}{2}, Var(\theta) = \dfrac{1}{2}$.

The conditional probability that the contextual feature a_z corresponds to a given type in the set Y is:

$$P(a_n|y_i) = \frac{1}{\sqrt{2\pi}E(\theta)} \exp\left(-\frac{[a_n - E(\theta)]^2}{2[Var(\theta)]^2}\right) \tag{22}$$

Then, we calculate the type that a contextual feature belongs to, and identify whether the video x_n is a misleading video or a legitimate one.

4. **Experiment.** This section introduces the dataset, experiment procedures and analysis results.

4.1. **Dataset.** The dataset used in this study consists of videos and contextual variables collected by spiders from Bilibili.com, a leading social medial video website in China. The videos are about two topics — traditional Chinese medicine and diabetes, and the original variables include release time, number of plays (or views), number of bullet screen comments, number of likes, number of coins, number of added-to-favorites and number of reviews. Two medical experts, one data expert, and one sociologist were invited to label the misleading videos manually. Specifically, the medical experts identified the falsehood of the video first, and then the data expert and sociologist identified the misleadingness of the video. Further, we constructed a balanced dataset with half misleading videos and half legitimate videos. As a result, the dataset used in the experiment consists of 200 misleading videos and 200 legitimate videos. The statistics of contextual features used in this study are shown in Table 1.

Table 1. Statistics of contextual features

Contextual features	**Mean**	**Standard Deviation**	**Minimum**	**Maximum**
SR	0.09	0.06	0	0.36
SA	0.77	0.09	0.46	0.98
WF	299.53	993.03	0.31	11350.70
IL	77.58	129.59	2.84	1123.85
CR	13.48	38.25	0.47	456.89
FaR	2.91	5.03	0.08	53.87
FoR	15.25	44.82	0.07	411.20
BRR	0.43	0.71	0	6.57
WRR	405.63	892.37	3.25	10501.00

4.2. **Experiment Setting.** We randomly selected 200 videos (half positive and half negative) as the training set, and the remaining 200 as the test set. The proposed naïve Bayesian model was compared with other baseline models including SVM, k-NN, decision tree, and random forest. As neural networks need to be trained on massive amounts of data, they were not used in this study. All experiments were performed on Python 3.6.12. The parameters of the benchmark models are as follows.

(1) SVM. The kernel function of the SVM used in the present work is the Gaussian radial basis function. The penalty parameter C was set at 10, and the kernel function parameter was set at 10.

(2) k-NN. For the k-NN model used here, the k value was set to 5, the number for concurrent operations for neighbor search was set to 1, and the size of leaves was set to 60.

(3) Decision tree. In the decision tree selected for comparison in the present work, no weights were assigned to the sample data of each class. The information increment was used as the criterion to measure the quality of node splitting, which is the criterion for the C4.5 algorithm. For the decision tree model, the maximum depth was set at 4, the minimum sample size on the leaf node was 6; the "best" strategy was used for node splitting, which means the node division was based on the "best split".

(4) Random forest. For the random forest model used here, the number of trees in the forest was set at 100, the information increment was used as the measure for the quality of node splitting, the maximum depth of the tree was 3, and the minimum number of samples at a leaf node was set at 5.

4.3. **Experiment Results.** In the experiments, four measures were used to assess the models' performance—accuracy, precision, recall, and F1-score. TP represents the number of correctly classified misleading videos; FN represents the number of misleading videos misclassified as legitimate videos by the model; FP means the number of legitimate videos misclassified as misleading videos; and TN represents the number of correctly classified legitimate videos. Accuracy describes the percentage of correct classification of fake and legitimate videos:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{23}$$

Precision describes the percentage of real misleading videos in the detected misleading videos:

$$Precision = \frac{TP}{TP + FP} \tag{24}$$

Recall describes the probability of the labeled misleading video being detected by the model:

$$Recall = \frac{TP}{TP + FN} \tag{25}$$

F1 score is a comprehensive measure that combines accuracy and recall:

$$F1 = \frac{2 \times Accuracy \times Recall}{Accuracy + Recall} \tag{26}$$

The proposed nine features are divided into four groups. The first group is mainly about similarity, including SR and SA. The second group is mainly about the video recommendation degree, including WF and IL. The third group is mainly about costly feedbacks, including CR, FaR and FoR. The fourth group is about reviews, including BRR and WRR. The effects of four feature groups are shown in Table 2. As Table 2 shows, feature

Table 2. The effects of four feature combinations

Feature combinations		Accuracy	Precision	Recall	F1-score
1	SR+WF+CR+BRR	0.705	0.697	0.783	0.738
2	SR+WF+FaR+BRR	0.725	0.707	0.821	0.760
3	SR+WF+FoR+BRR	0.715	0.702	0.802	0.749
4	SR+IL+CR+BRR	0.705	0.697	0.783	0.738
5	SR+IL+FaR+BRR	0.725	0.707	0.821	0.760
6	SR+IL+FoR+BRR	0.715	0.702	0.802	0.749
7	SA+WF+FaR+BRR	0.710	0.697	0.802	0.746
8	SA+WF+FoR+BRR	0.715	0.699	0.811	0.751
9	SA+IL+CR+BRR	0.685	0.684	0.755	0.717
10	SA+IL+FaR+BRR	0.710	0.698	0.802	0.746
11	SA+IL+FoR+BRR	0.715	0.699	0.811	0.751

Note: Feature combinations with Accuracy below 0.7 and Precision, Recall, and F1-score below 0.6 were removed from the table.

combinations 2 and 5 perform the best, reaching the highest F1 score 0.760. The combination of SR, WF/IL, FaR, BRR achieves the best performance. The results in Table 2 also provide some information about the effectiveness of the features. For example, BRR is more effective than WRR in the fourth group (review related features).

We also report the results with all the nine features considered and compare our model with other baselines (shown in Table 3). As shown in Table 3, our proposed model achieves a higher precision than all the other baselines except the random forest model. Our model combined with all nine features achieved the highest F1-socre 0.81, which is significantly higher than other baselines such as SVM (0.63), k-NN (0.60), decision tree (0.52) and random forest (0.66). The model that considers all the nine features also outperforms the model combining the best four features identified in Table 2, indicating all the proposed nine features are useful in the misleading video detection task, and also the method of applying contextual features to identify misleading video is proved to be efficient.

Table 3. The comparison of different models (with all 9 features)

Model	Accuracy	Precision	Recall	F1-score
SVM	0.69	0.83	0.51	0.63
k-NN	0.61	0.65	0.57	0.60
Decision tree	0.62	0.76	0.40	0.52
Random forest	0.70	0.81	056	0.66
Our model	0.77	0.71	0.93	0.81

5. **Conclusion.** Nowadays, most detection work applies deep learning method which has poor interpretability. In order to solve the problem, this paper uses Bayesian model to replace deep learning model, so that ensures the efficiency and interpretability at the same time. In this study, we put forward nine contextual features and proposed a naïve Bayesian classifier-based model for misleading video detection. The results suggest that all the nine contextual features are in the misleading video detection task and our model outperforms other baselines including SVM, k-NN, decision tree and random forest in terms of accuracy (0.77), precision (0.71), recall (0.93) and F1-score (0.81).

Future research can be performed from the following three aspects. First, more contextual features can be proposed and evaluated. Second, the detection model can be optimized to improve its robustness and increase its generalizability. Third, the dataset for misleading video detection could be expanded. Based on a richer dataset, the deep learning approach can be applied in the future.

Funding. This research was supported by the National Natural Science Foundation of China under grants 72271093.

REFERENCES

[1] S. Loeb, S. Sengupta, M. Butaney, J. N. Macaluso, S. W. Czarniecki, R. Robbins, R. S. Braithwaite, L.-S. Gao, N. Byrne, D. Walter, A. Langford, "Dissemination of misinformative and biased information about prostate cancer on YouTube," *European Urology*, vol. 75, no. 4, pp. 564-567, 2019.

[2] K. C. Madathil, A. J. Rivera-Rodriguez, J. S. Greenstein, A. K. Gramopadhye, "Healthcare information on YouTube: A systematic review," *Health Informatics Journal*, vol. 21, no. 3, pp. 173-194, 2015.

[3] L.-W. Zhang, K.-L. Chen, H. Jiang, J. Zhao, "How the Health Rumor Misleads People's Perception in a Public Health Emergency: Lessons from a Purchase Craze during the COVID-19 Outbreak in China," *International Journal of Environmental Research and Public Health*, vol. 17, no. 19, pp. 1-15, 2020.

[4] E.-K. Wang, X. Zhang, F. Wang, T.-Y. Wu, C.-M. Chen, "Multilayer Dense Attention Model for Image Caption," *IEEE Access*, vol. 7, pp. 66358-66368, 2019.

[5] N. Hussain, H. T. Mirza, I. Hussain, "Detecting spam review through spammer's behavior analysis," *Advances in Distributed Computing and Artificial Intelligence Journal*, vol. 8, no. 2, pp. 61-71, 2019.

[6] S. Syed-Abdul, L. Fernandez-Luque, W.-S. Jian, Y.-C. Li, S. Crain, M. H. Hsu, Y.-C. Wang, D. Khandregzen, E. Chuluunbaatar, P. A. Nguyen, D. M. Liou, "Misleading health-related information promoted through video-based social media: anorexia on YouTube," *Journal of Medical Internet Research*, vol. 15, no. 2, pp. 1-13, 2013.

[7] R. Hou, V. Perez-Rosas, S. Loeb, R. Mihalcea, "Towards automatic detection of misinformation in online medical videos," in the 2019 International Conference on Multimodal Interaction. ACM, 2019, pp. 235-243.

[8] A. Ghenai , Y. Mejova, "Fake cures: user-centric modeling of health misinformation in social media," in the ACM on Human-Computer Interaction. CSCW, 2018, pp. 58-78.

[9] P. Palod, A. Patwari, S. Bahety, S. Bagchi, P. Goyal, "Misleading metadata detection on YouTube," *Advances in Information Retrieval*, vol. 11438, pp. 140-147, 2019.

[10] X. Liu , B. Zhang , A. Susarla , R. Padman, "YouTube for patient education: a deep learning approach for understanding medical knowledge from user-generated videos," in the KDD Workshop on Machine Learning for Medicine and Healthcare. KDD, 2018, pp. 1-3.

[11] A. Alsaeedi , M. Al-Sarem, "Detecting rumors on social media based on a CNN deep learning technique," *Arabian Journal for Science and Engineering*, vol. 45, pp. 10813-10844, 2020.

[12] A. Mahabub, "A robust technique of fake news detection using ensemble voting classifier and comparison with other classifiers," *SN Applied Sciences*, vol. 2, no. 4, pp. 525-534, 2020.

[13] P. Shah, Z. Kobti, "Multimodal fake news detection using a cultural algorithm with situational and normative knowledge," in the 2020 IEEE Congress on Evolutionary Computation (CEC). IEEE, 2020, pp. 1-7.

[14] I. Ahmad, M. Yousaf, S. Yousaf, M. O. Ahmad, "Fake news detection using machine learning ensemble methods," *Complexity*, pp. 1-11, 2020.

[15] J. A. Nasir, O. S. Khan, I. Varlamis, "Fake news detection: a hybrid CNN-RNN based deep learning approach," *International Journal of Information Management Data Insights*, vol. 1, no. 1, pp. 1-13, 2021.

[16] O. Ajao, D. Bhowmik, S. Zargari, "Fake news identification on Twitter with hybrid CNN and RNN models," in the 9th International Conference on Social Media and Society. ACM, 2018, pp. 226-230.

[17] A. Agarwal, M. Mittal, A. Pathak, L. M. Goyal, "Fake news detection using a blend of neural networks: an application of deep learning," *SN Computer Science*, vol. 1, no. 3, pp. 143-152, 2020.

[18] N. Ashar, H. Bhatt, S. Mehta, "A framework for detection of video spam on YouTube," *International Journal of Computer Science and Information Technologies (IJCSIT)*, vol. 6, no. 6, pp. 5068-5072, 2015.

[19] S. Kanodia, R. Sasheendran, V. Pathari, "A novel approach for YouTube video spam detection using Markov decision process," in the 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI). ICACCI, 2018, pp. 60-66.

[20] N. Aggarwal, S. Agrawal, A. Sureka, "Mining YouTube metadata for detecting privacy invading harassment and misdemeanor videos," in the 12th International Conference on Privacy. IEEE, 2014, pp. 84-93.

[21] S. Seth, S. Biswas, "Multimodal spam classification using deep learning techniques," in the 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). IEEE, 2017, pp. 346-349.

[22] A. Araujo, B. Negrevergne, Y. Chevaleyre, J. Atif, "Training compact deep learning models for video classification using circulant matrices," in the European Conference on Computer Vision Workshops. Munich(DE). ECCV, 2018, pp. 271-286.

[23] H.-M. He, T. Watson, C. Maple, J. Mehnen, A. Tiwari, "A new semantic attribute deep learning with a linguistic attribute hierarchy for spam detection," in the International Joint Conference on Neural Networks (IJCNN). IEEE, 2017, pp. 3862-3869.

[24] Abdullah-All-Tanvir, E. M. Mahir, S. Akhter, M. R. Huq, "Detecting fake news using machine learning and deep learning algorithms," in the 2019 7th International Conference on Smart Computing & Communications (ICSCC). IEEE, 2019, pp. 1-5.

[25] M. Granik, V. Mesyura, "Fake news detection using naive Bayes classifier," in the IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON). IEEE, 2017, pp. 900-903.

[26] T. S. Reshmi, SDM. Raja, J. Priya, "Fake news detection using source information and Bayes classifier," *IOP Conference Series Materials Science and Engineering*, vol. 1084, no. 1, pp. 1-6, 2021.

[27] F. Isada, Y. Isada, "An empirical study of the influence of trust and empathy in word-of-mouth communication on the Internet on consumer behaviour," in the 5th International Conference on Applied Economics, Business and Development (AEBD 13). AEBD, 2013, pp. 194-202.

[28] M. Y. Day, C.-C. Wang, C.-C. Chen, S.-C. Yang, "Exploring review spammers by review similarity: a case of fake review in Taiwan," in the Third International Conference on Electronics and Software Science (ICESS2017). ICESS, 2017, pp. 166-170.