Multimedia Tools and Applications (2023) 82:7017–7038
https://doi.org/10.1007/s11042-022-13595-4

Rumour identification on Twitter as a function of novel textual and language-context features

Ghulam Ali[1] · Muhammad Shahid Iqbal Malik[2]

Received: 30 August 2020 / Revised: 12 May 2022 / Accepted: 18 July 2022 /
Published online: 12 August 2022

Abstract

Social microblogs are one of the popular platforms for information spreading. However, with several advantages, these platforms are being used for spreading rumours. At present, the majority of existing approaches identify rumours at the topic level instead of at the tweet/post level. Moreover, prior studies used the sentiment and linguistic features for rumours identification without considering discrete positive and negative emotions and effective part-of-speech features in content-based approaches. Similarly, the majority of prior studies used content-based approaches for feature generation, and recent context-based approaches were not explored. To cope with these challenges, a robust framework for rumour detection at the tweet level is designed in this paper. The model used word2vec embeddings and bidirectional encoder representations from transformers method (BERT) from context-based and discrete emotions, linguistic, and metadata characteristics from content-based approaches. According to our knowledge, we are the first ones who used these features for rumour identification at the tweet/post level. The framework is tested on four real-life twitter microblog datasets. The results show that the detection model is capable of detecting 97%, 86%, 85%, and 80% of rumours on four datasets respectively. In addition, the proposed framework outperformed the three latest state-of-the-art baselines. BERT model presented the best performance among context-based approaches, and linguistic features are best performing among content-based approaches as a stand-alone model. Moreover, the utilization of two-step feature selection further improves the detection model performance.

✉ Muhammad Shahid Iqbal Malik
msi_id@yahoo.com; shahid.iqbal@cust.edu.pk

Ghulam Ali
ghulamali007@yahoo.com

[1] Department of Computer Science, COMSATS University Islamabad, Attock Campus, Islamabad, Pakistan

[2] Department of Computer Science, Capital University of Science and Technology, Kahuta Road, Sihala, 44000 Islamabad, Pakistan

Keywords Classification · Language context · Detection · Twitter · Rumour

1 Introduction

Social networks had played a very significant role in the dissemination of knowledge in the last few years. More precisely, with the facility of short-length messages, microblogs became a more common platform for expressing views and sharing opinions. Twitter is one of the most popular microblog websites that allows users to share their views and opinions in the form of tweets with a length of a maximum of 140 characters [45]. Generally, the dissemination of reliable information is one of the objectives of the social network. Among all the positive aspects of the social network, there are a few downsides, fake news and rumours can easily be spread to millions of people in a short time [36]. Such false/fake information not only causes serious problems for social media websites but also creates disasters for governments and economies. For example, in 2013 fake news was spread about the burst of two bombs in the white house and the US president was wounded. This rumour initially created panic on a large scale and caused a dramatic crash in the stock market. Therefore, the propagation of fake/rumour information on the social network is highly undesirable [10].

Rumour is defined as unverified or unproved information [48]. There are several meanings of rumour, "unverified and instrumentally related facts claim in circulation" is commonly used. The main difference between fake news and rumour is that "fake news is the false information while the rumour is unverified, it is not necessarily false and may turn out to be false or true" [35]. Rumour and fake content identification are similar in technique because they have the majority of features in common. In COVID-19, new rumours according to the situation were circulated that threatened people from one side while people are still being fooled by the others. For instance, smoking alcohol stops COVID-19 or holding the breath for 10 s to check COVID-19 [42, 43]. The dissemination of unverified information causes mistrust on social networks e.g. Facebook was declared a "dust and cloud of nonsense" in 2016 when it was unable to track the spread of rumours about the US presidential election [11]. Unfortunately, such rumours can be found in any area of life. Therefore, it is necessary to identify rumours so that people will not be misled [46].

The early identification of rumours enables us to stop the spread of rumours. People can be prevented or avoided from any threat/panic caused by rumours when we identify it at an early stage. It is difficult to identify the rumours but it is in high demand. Identification of rumour is a challenging task due to three reasons: (1) Demand for real-time detection, (2) The Nature of rumours are confusing, and (3) A lot of work to process huge information. Many studies are conducted on rumour identification in literature. In the majority, text or image content-based methods are presented [26, 27, 40]. Few worked on exploiting the propagation features [23].

More specifically, several types of features have been proposed and various models are developed for rumour identification. For instance, influence potential [36], network characteristics [13, 18, 20, 36, 47], textual features [1, 2, 13], personal interest [36], temporal, semantic and structural features [3, 17, 19, 38] etc. It is necessary to utilize some new characteristics to efficiently identify rumours. Therefore, positive and negative discrete emotions, linguistic and metadata characteristics are utilized from content-based and word2vec [28], and BERT methods are selected from context-based features. According to our knowledge, these features were not used in literature for rumour detection in health and related domains. In addition, the

utilization of a powerful machine learning model is always necessary to build an effective detection/identification system.

The study is organized in the following way: Next section describes research questions and contributions. Section 3 provides related work. Section 4 presents the methodology. Section 5 presents the experimental results and examined their outcomes. Discussion and implications are presented in Section 6 whereas concluding remarks are provided in Section 7.

2 Research questions and contributions

In this case study, three research questions are addressed:

RQ1 Which ML model is more robust in performance to design an effective model for rumour detection.

RQ2 Does any subset of features exist that are most influential and have a strong relationship to rumours?

RQ3 Which type of characteristic (Word2vec, BERT, discrete emotions, metadata, and linguistics) makes the maximum contribution to twitter rumour identification?

The objective of this case study is to examine the impact of proposed discrete emotions, linguistic, metadata, word2vec, and BERT methods for the identification of rumours at the tweet level. The process of feature selection is employed and the best subset is selected using the wrapper method. four well-known twitter datasets, four popular machine learning models, and five evaluation measures are utilized for the experimental setup. The proposed framework is evaluated using an individual type of features as a standalone model and using their hybrid combinations. The findings of this case study provide new insights for Zikavirus, breaking news of Ottawa Shooting and Germanwings Crash events domains in the area of rumour detection. To sum up, the main highlights of the paper are:

1. An effective rumour detection system for Zikavirus, Ottawa Shooting, and Germanwings Crash events is developed using novel features and the random forest model.
2. The BERT model and word2vec embeddings are used to examine the language context of a tweet for rumour identification.
3. A two-step framework for feature selection is employed which enables shortlisting of the best features.
4. The comparison of four ML models reveals that random forest model demonstrated the best performance.
5. The proposed framework operates at the single tweet level rather than at the topic level.
6. The experimental results demonstrate that our model outperforms the three state-of-the-art baselines on all evaluation metrics.
7. The proposed framework is validated on four real-life events twitter datasets and achieved maximum accuracy of approximately 97% on the zikavirus dataset.
8. The findings reveal that URLs, Trust emotion, Verbs, Adjectives, and Propositions are the top-5 textual features to detect rumour at the tweet level.

3 Related work

In recent years, rumour detection has become a hot issue and several approaches have been presented in the literature. Some concentrated on proposing new characteristics while others attempted to apply robust machine learning models. Major approaches are supervised and the most common are content-based.

In 2011, Castillo, et al. presented a method to evaluate information credibility for news articles on the Twitter network [7]. They used the length of distinct words, total words, and sentiments as features. Their system achieved precision and recall in the range of 70–80%. Later in 2016, Ma, et al. developed a method for learning continuous representation of twitter events to deal with rumour detection [22]. In 2017, Kwon, et al. developed a rumour detection model using temporal, structural, and linguistic features [19]. Then a robust model is presented to identify rumours between 3 and 56 days at varying time slots [20]. Four types of features are investigated and varied predictive performance is observed on various time windows. Their results showed that user and linguistic indicators are significant for the short-term whereas temporal and structural features have good performance for the long-term period.

Next in 2018, Sicilia, et al. proposed a rumour detection model at tweet level aiming by exploiting influence potential measures, personal interest, and network characteristics [36]. Their method achieved an accuracy of 89%. Then, a framework for identifying users spreading rumours is developed by Ruchansky, et al. [34]. The model used a recurrent neural network and it outperformed four standard baselines. Similarly, a system is developed by Vijeev, et al. [39] in the same year to identify rumours on Twitter microblog. Content-based and user-based features are used, and three machine learning models are tested. random forest classifier outperformed. s.

Recently in 2019, a detection model is proposed to detect rumours early in time [37]. The method reduces the time span of prediction by 85%, which is better than the state-of-the-art baseline. Next, Hamidian, et al. [13] derived a two-step model to address rumour detection problem and then classification aiming in exploiting the network-specific, n-grams and pragmatic features. Similarly, text-based fusion neural network model by Chen, et al. [8], graph convolutional networks based method by Huang, et al. [16], and rumour veracity detection by Kumar, et al. [18]. Then a novel method for detecting rumours in the Arabic language using "semi-supervised expectation maximization." is presented [1]. User and content level features are used. Wang, et al. presented a method for the detection of rumours [41], aiming to exploit the structures for dynamic propagation and content characteristics in combination. Their method is very effective for capturing the dynamic structure. The details of the literature are also presented in tabular form as shown in Table 1.

More recently in 2020, a probabilistic model is developed [47] aiming to use not only retweeting behavior but also intent. The proposed system is effective in the detection of malicious users. Then Bai, et al. proposed a stochastic attention convolutional neural network-based system to detect rumour by using fine-grained and coarse-grained features [2]. Similarly, the identification of retweeting behavior for rumours is presented by Tian, et al. [38]. They used reaction time, retweeting frequency and TF-IDF features for model construction and their system achieved an accuracy of 88%. Bian et al. developed a propagation and dispersion-based bi-directional graph convolutional network method to detect rumour [3]. According to the authors, their method is more effective than the state-of-the-art baseline. Then Huang et, al. proposed a heterogeneous graph attention network framework to identify rumour [17]. They developed the tweet-word-user graph using semantic features on Twitter network.

Table 1 Features and ML models used in the literature

Category	Feature description	Model	Refs
User and content type	Potential measures, personal interest and network characteristics	Random forest	[36]
User and content type (semantic, sentiment)	Network specific, n-grams, pragmatic features	C4.5 classifier	[13]
User and content type (structure, syntactic)	User, structural, linguistic, and temporal features	Random forest	[20]
User and content type (sentiment, syntactic)	Sentiment, mentions, hashtags, time span	Semi-supervised expectation–maximization	[1]
Content type (structure, syntactic)	Temporal, structural, and linguistic	Logistic regression, random forest	[19]
User and content type (syntactic, sentiment)	Lexical, syntactical, negation, pragmatic and network specific	Decision tree	[18]
User Type	User intention and story veracity	Expectation Maximization model	[47]
Content type	Course-grained features, fine-grained features	Neural Network model	[2]
Content type (Semantic & metadata)	Reaction time, Tweeting frequency, TF-IDF	Neural Network model	[38]
Content type (Semantic & structure)	Structure features, casual features	Bidirectional graph based model	[3]
Content type (Semantic)	Point-wise mutual influence, TF-IDF	Tweet-word subgraph based model	[17]
Graph type	Reply tree and user graph	Graph convolutional network	[21]
Content and user type	Sentiment, content relevance, user attention rate	Lifelong machine learning	[14]

In 2021, the graph convolutional network-based rumour detection model is developed by lotfi, et al. [21]. Reply tree and user graph are extracted for each conversation and they claimed that their model outperformed the baseline but the time and space complexity of their model is very high. The spatiotemporal graph and attention-based neural networks are also used in citywide crowd flows prediction problems [5, 12, 15, 33]. Later in 2022, HE, et al. [14] proposed another model for propaganda detection using lifelong machine learning technique. They used sentiment, content relevance and user attention rate features but the time and space complexity of their model are very high.

Most of the aforementioned literature belongs to supervised learning. The review presented so far depicts that majority of approaches detect rumour at the topic/conversation level. Rumour identification at the post/tweet level needs more attention. Next, different topics do not have the same structure of sentences and semantics of words, methods based on such features as well as the use of characteristics at the topic level maybe not be directly relevant to detect rumour for a specific topic level. Therefore, available solutions cannot be directly applicable at the tweet level. In addition, prior contributions at tweet-level used influence potential [36], network characteristics [13, 18, 20, 36, 47], textual features [1, 2, 13], personal interest [36], temporal, semantic and structural features [3, 17, 19, 38] etc. for rumour detection. To the best of our knowledge, no one used discrete emotions, tweet-related metadata, word2vec, and BERT embedding techniques as characteristics for rumour identification at the tweet/post level. Inspired by these ideas, we propose a novel framework that exploits linguistic, metadata, discrete emotions, word2vec and BERT techniques to deal rumour detection at tweet-level. We hope to detect rumour more accurately.

4 Methodology

The components of our proposed framework are presented here. The pipeline of the rumour detection framework is shown in Fig. 1. First, four real-life publicly available twitter datasets are collected. In addition, to cope with datasets, the more required information is crawled from Twitter social microblog. The datasets are further considered for pre-processing (cleaning and removal of irrelevant information). Then it leads to the extraction of five types of features (discrete emotion, linguistic metadata, word2vec, and BERT). Feature normalization (min-max normalization) and feature selection are applied to representative features. Four popular machine learning (ML) models, five evaluation measures, and 20-fold cross-validation are used in experiments. As an outcome, the system classifies tweets into rumour or not-rumour class.

4.1 Problem formulation

Let $\chi_{m,n}$ be a feature matrix, having m rows and n columns, where m represents the number of tweets in and n denotes the number of features. is the collection of tweets $= \{\tau_1, \tau_2, \tau_3, \ldots, \tau_m\}$ in the dataset and χ_i is the feature vector of the tweet τ_i such that $\chi_i \in R^n$. Every tweet τ_i is an instance/sample, consisting of the following components {D, L, M, W, B, C}. Where D represents discrete emotions, L represents linguistic, M represents metadata features, W represents word2vec, and B represents BERT embeddings related to tweets whereas C is the target class label i.e. rumour or not-rumor.

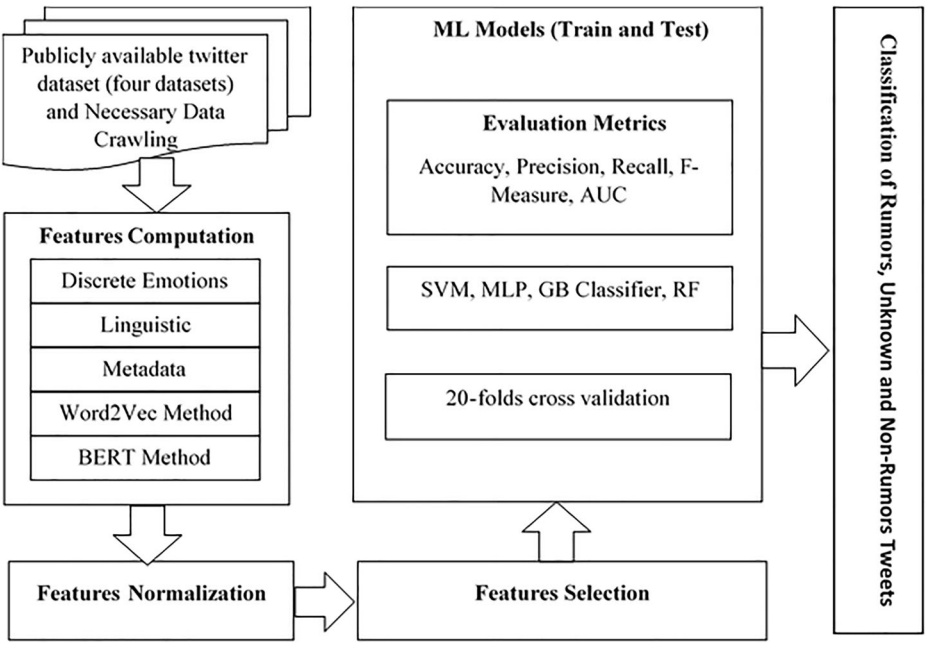

Fig. 1 Flow of steps in research methodology

Let Y be the vector of predicted class labels for all tweets and y_i represents the predicted class label for τ_1 (i.e. rumour or non-rumour). To classify whether a tweet is a rumour or not-rumour, we define the following predictive function.

$$y_i = F(\tau_i/\chi_i) \tag{1}$$

Where

$$F(\tau_i/\chi_i) = \begin{bmatrix} \geq 0 & \text{if} & y_i = +1, & \text{rumour} \\ < 0 & \text{if} & y_i = -1, & \text{not rumour} \end{bmatrix} \tag{2}$$

Our aim here is to develop a predictive model that will minimize the predictive error of y_i given χ_i.

4.2 Datasets

In this case study, four real-life twitter data sets are used. The first dataset (DS1) is publicly available and is built by extracting tweets from 111 events on Twitter [20]. In this dataset, every tweet is annotated as either rumour or not-rumour. In the beginning, we have 111 events and there are several tweets in each event. We selected 12 health-related events of which 4 are non-rumour events and the remaining are rumour events. After preprocessing, we have 653 instances in total of which 359 instances are rumour (positive) whereas 294 are non-rumour. The second dataset (DS2) is designed by crawling tweets related to the health-domain. Zika virus is the only topic and related tweets are considered. In other words, using #Zikavirus and Zika microcephaly [36], the tweets are selected. After preprocessing, we have 693 instances as shown in Table 2, in which 58% belong to the rumour class and 42% are related to the non-rumour class. The third dataset (DS3) is also publicly available and contains tweets collected from breaking news of the Ottawa Shooting event. The total number of instances is 890 among which 470 are rumours (52.8%) and 420 are non-rumours (47.2%). The fourth dataset (DS4) consists of tweets related to breaking news of the Germanwings Crash event and is publicly available. It contains 469 instances among which 238 are rumours (50.7%) and 231 are non-rumours (49.3%) respectively.

4.3 Machine learning models and evaluation metrics

Four machine learning models are used in experiments to classify tweets into rumours or non-rumours. The ML models are: Gradient Boosting Classifier (GBC) [12], Multilayer Perceptron

Table 2 Description of datasets

Name	# Tweets	Description
DS1	653	Tweets from 12 different events related to the health domain
DS2	693	Tweets crawled using #Zikavirus and zika microcephaly.
DS3	890	Tweets collected from the breaking news of Ottawa shooting
DS4	469	Tweets collected from the breaking news of Germanwings Crash

(MLP) [15], Support Vector Machine (SVM) [33], and Random Forest (RF) [5]. Furthermore, the 20-fold cross-validation method and five evaluation measures are used to evaluate the performance. The measures are precision, accuracy, recall, f1-score, and area under the curve (AUC). Python programming language is used to code the models [32].

4.4 Feature extraction

In this case study, five types of features are extracted. The objective is to find the set of influential features that can detect rumour or non-rumour accurately at the tweet level. The features are (1) Word2vec Embedding, (2) BERT model, (3) Discrete emotions, (4) Linguistic and (5) Metadata type. A detailed description of these features is provided next.

4.4.1 Word2Vec embedding model

To capture the semantic of a word, word-embedding is one of the most popular representations of text. Word2vec is one of the methods to generate word embeddings. It can be utilized to get insights for rumour detection from tweet data. Word2Vec is based on an unsupervised shallow two-layer neural network, that can be trained for generating high quality, distributed, and continuous dense vector representation of words [28]. It can capture contextual and semantic similarity and consists of two learning algorithms, i.e. continuous bag-of-words (CBOWs) and continuous skip-gram. The architectures of both algorithms are shown in Fig. 2.

In the continuous bag-of-words model, the target word is predicted given the context words, whereas the skip-gram model predicts the context words given the target word. We used the skip-gram model to generate context words up to 100 dimensions using DS1 and DS2 Twitter datasets. Each context word in a dimension contains information about one aspect of a particular work. The objective of using word2vec is to capture the context words to accurately identify rumours in the tweet text.

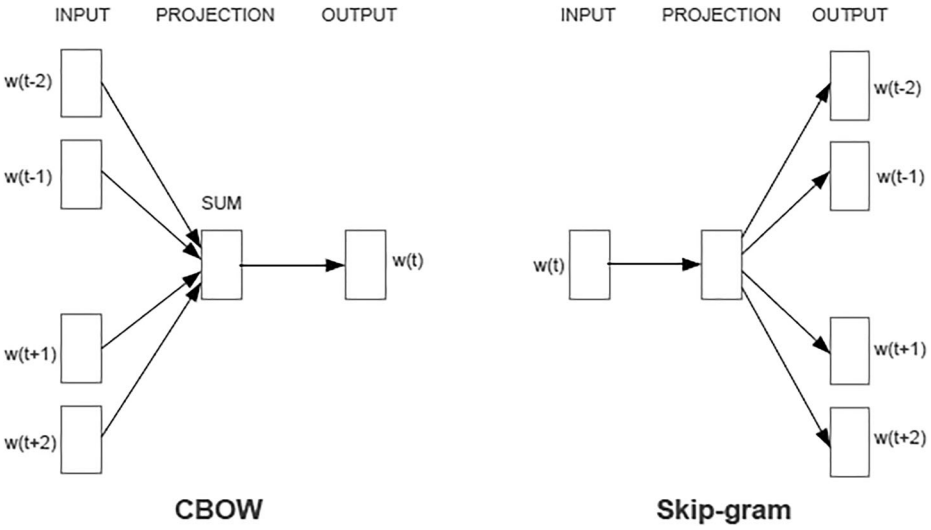

Fig. 2 The CBOW and Skip-gram architecture of word2vec [29]

4.4.2 BERT model

BERT is a transformer-based ML approach, designed by Jacon Devin and his colleagues in 2018 [9]. It is developed for learning tasks in natural language processing. BERT model can be employed for various language tasks such as sentiment analysis, next sentence classification, question answering, named entity recognition, etc. Also, Google has been using BERT for understanding users' searches since 2019 [31]. The BERT has two models (1) BERT-base and (2) BERT-large. Both models are pre-trained. The BERT-base uses 12-encoders with 12-bidirectional self-attention heads whereas BERT-large uses 24-encoders with 24-bidirectional heads. The architecture of BERT for natural language processing is presented in Fig. 3.

It utilizes an attention mechanism (transformer) that learns contextual relations among sub-words/words in a text. The transformer consists of two modules; the first is an encoder that takes text input and the second is a decoder that predicts the desired output. It is bidirectional or non-directional because the directional models read input sequentially whereas the encoder reads the entire input sequence at once. We are the first to use the BERT model for rumour detection at the tweet level. As our task resembles NLP, therefore utilization of BERT will be more beneficial.

4.4.3 Discrete emotions

Discrete emotions are the type of textual features. According to theory, discrete emotions are biologically determined emotional responses whose recognition are the same for all persons regardless of cultural differences [44]. Eight discrete emotions are classified as discrete positive and discrete negative. Anticipation, joy, surprise, and trust are discrete positive

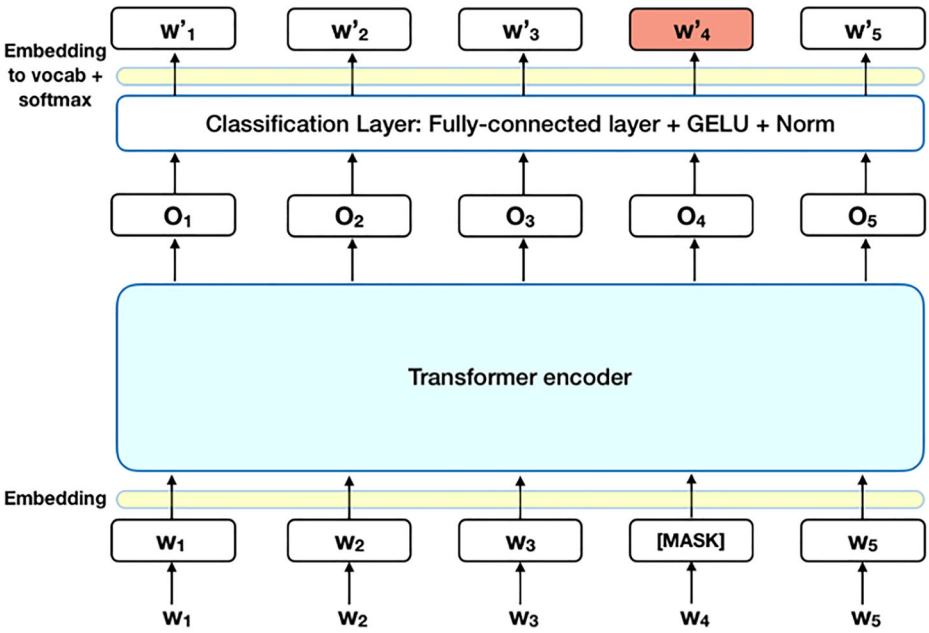

Fig. 3 Architecture of the BERT model for natural language processing [9]

whereas anxiety, sadness, anger, and disgust are discrete negative emotions [25]. These emotions can be extracted using the NRC lexicon [30] provided by National Research Council Canada. The lexicon contains 8265 words. The mathematical formula to compute each discrete emotion is the same. E.g. for trust emotion:

$$\text{Trust-emotion} = (\#\text{trust-related words} * 100)/\text{total-words in a tweet} \quad (3)$$

The details of the NRC emotion lexicon, the list of emotion dimensions, and the number of words related to each emotion dimension are described in Table 3. The aim is to investigate the influence of discrete emotions embedded in tweet text on rumour identification. According to our knowledge, we were the first to use these emotions for rumour detection at the tweet level. The utilization of discrete emotions will uncover the significance of each positive and negative emotion.

4.4.4 Linguistic features

Linguistic features of tweet text are the important predictors that can influence rumour identification [25]. Part-of-speech is a type of linguistic characteristic. They are the list of words that have similar grammatical properties and follow the linguistic rules [6]. Thirty-five part-of-speech tags are available by Natural Language Tool Kit (NLTK). These features can be easily extracted using NLTK part-of-speech tagger [4] and then their percentage can be computed from the tweet text. These tags-based characteristics may have a significant role in detecting rumours at the tweet level. The list of all extracted linguistic features is presented in Table 4.

4.4.5 Metadata features

Prior studies demonstrated that metadata characteristics play a significant role in natural language processing tasks such as helpfulness prediction and rumour detection [24, 36]. These features can cause an effect in an intangible or indirect way. They consist of all the properties related to the Twitter account of a user such as followings, number of followers, age of user's account, presence of questions marks and URLs in user's tweet, etc. The set of proposed metadata features was not used by prior studies for rumour detection. The utilization of these features will improve the classification performance of the rumour detection model. The list of extracted metadata features is presented in Table 4.

Table 3 Details of emotion lexicon provided by NRC [30]

Emotion type	Number of words
Trust	1231
Anticipation	839
Joy	689
Surprise	534
Anxiety	1476
Anger	1247
Sadness	1191
Disgust	1058
Total	8265

Table 4 List of extracted discrete emotions, linguistic and metadata features

Feature type	Feature name	Feature type	Feature name
Discrete Emotions	Anticipation	Linguistic	Prepositions
	Joy		Adjectives
	Surprise		Comparative Adj
	Trust		Superlative Adj
	Anxiety		Plural nouns
	Sadness		Singular noun
	Anger		Plural noun
	Disgust		Proper singular noun
Metadata	# Full Stop		Proper plural noun
	URLs		Personal pronoun
	# Hashes		Possessive pronoun
	Colon		Adverbs
	Number of followers		Comparative Adverbs
	Number of followings		Superlative Adverbs
	Number of Statuses		Particle
	Followers_Following_ratio		Interjection
	Difference_in_minutes		Verbs
	Number of tweets within last 90 days		Past participle verbs
	Number of tweets within last 365 days		3rd-person singular present tense verbs
	Average tweets per day in last 60-90 days		Non 3rd-person singular present tense verbs
	Average tweets per day before 90 days		Coordinating conjunction
	Average tweets per day in last 365 days		Determiner

4.4.6 Feature selection

The selection of the extracted features is an important task in the feature engineering process. The objective of this section is to examine which feature combination is most significant and to test whether the proposed ones are significant or not for the classification task. We designed a two-step strategy to select the most significant combination of proposed features so that an effective rumour detection model can be designed. Various candidate sets of features are compared using the random forest classifier and every feature is evaluated using the accuracy metric. In beginning, we have 44 features in total. In the first step of the double-round strategy, every feature performance is evaluated using the accuracy metric, and features are ranked in descending order. We selected the top-23 and their performances are shown in Fig. 4.

In the second step, a customized elimination method is applied to all features selected in step 1. The impact of each feature is evaluated by eliminating it from the feature set and then measuring performance with the rest of the features. Random forest is used as the classifier. The steps of the elimination method are; At first, by combining 23 features, the accuracy measure is computed and denoted by the $\text{Accuracy}_{\text{base feature set}}$. After that, every feature is removed one by one, and performance is computed using the rest of the features, denoted by $\text{Accuracy}_{\text{drop f from base set}}$. Each feature's impact is calculated by taking the difference between $\text{Accuracy}_{\text{drop f from base set}}$ and $\text{Accuracy}_{\text{base feature set}}$ as described by Eq. (4). If I(f) is zero or above, then it reveals that elimination of that particular feature is useful. Thus, we can eliminate that feature without any loss. If I(f) is negative, then accuracy will decrease by

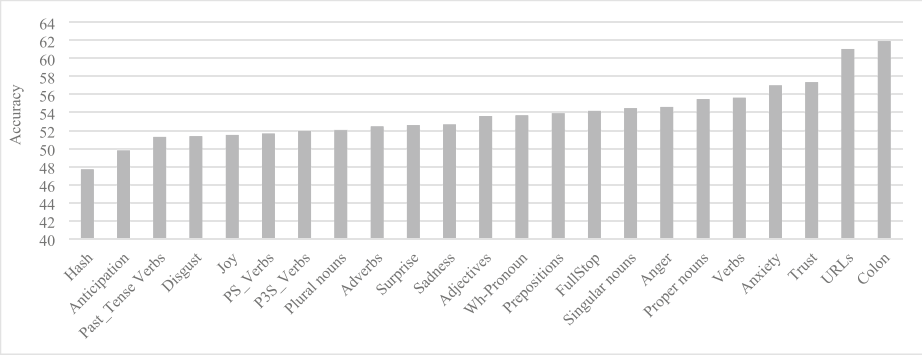

Fig. 4 Top-23 features performance using accuracy metric (step 1)

eliminating that feature. We dropped those features which have I(f) ≥ 0 and the remaining features are selected. Finally, we got the 15 best features and their I(f) values are presented in Fig. 5. The list of selected features is presented in Table 5.

$$\mathrm{I(f)} = \mathrm{Accuracy}_{\text{drop f from base set}} - \mathrm{Accuracy}_{\text{base feature set}} \tag{4}$$

4.5 Baselines

For comparisons, we selected three prior studies. The reason why we have chosen these is that these approaches also used the Twitter platform for the dataset construction.

1. Sicilia, et al. [36] used influence potential measures, personal interest, and network characteristics.
2. Kumar, et al. [18] used content-based, pragmatic, and network-specific features for rumour detection on Twitter.
3. Huang, et al. [17] used a heterogeneous graph attention network framework for rumour detection.

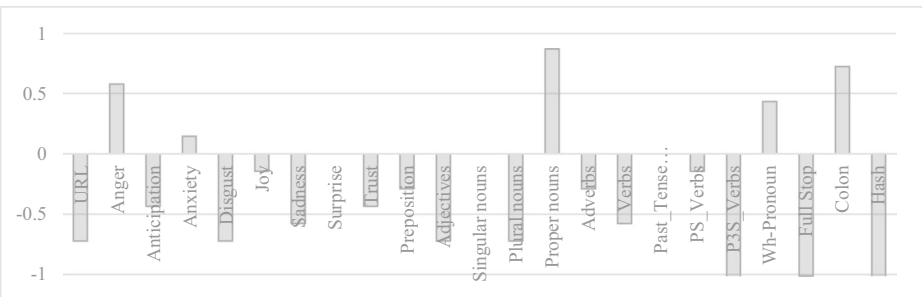

Fig. 5 Performance of top-15 features (round 2)

Table 5 List of selected features

Feature Type	Feature Name	Description
Discrete Emotions	Anticipation	No of Anticipation lexicon words in the tweet Text
	Disgust	No. of Disgust lexicon words in the tweet Text
	Joy	No. of Joy lexicon words in the tweet Text
	Sadness	No. of Sad lexicon words in the tweet Text
	Trust	No. of Trust lexicon words in the tweet Text
Linguistic	Prepositions	No. of prepositions in the tweet Text
	Adjectives	No. of adjectives in the tweet Text
	Plural nouns	No. of plural Nouns in the tweet Text
	Adverbs	No. of adverbs in the tweet Text
	Verbs	No. of verbs in the tweet Text
	PS_Verbs	No. of singular verbs which are in present tense in tweet Text
	P3S_Verbs	No. of 3rd person singular verbs which are in present tense in tweet Text
Metadata	# Full Stop	No, of Full stops in the tweet Text
	URLs	Whether the tweet text contains any URL?
	# Hashes	No. of Hashes in the tweet Text

5 Results and analysis

In this section, three types of experiments are conducted to evaluate the effectiveness of the proposed framework for rumour detection in specific health events, Ottawa shooting, and Germanwings crash events. We used python language for feature extraction, training, and testing of ML models. In addition, the Weka tool is used for feature selection and feature normalization tasks.

5.1 Prediction performance

In this section, we look for the best-performing ML classifier using the proposed set of features for rumour detection on Twitter. For this purpose, a hybrid combination of discrete emotions, metadata, and linguistic features is used to compare the performances of four classifiers. Four popular machine learning models (Section 4.2) with 20-fold cross-validation are implemented in the python programming language [32]. As a result, four rumour detection models are built and evaluated using five evaluation measures. i.e. precision, accuracy, recall, f1-score, and area under the curve are measured. The aforementioned mechanism is employed for all datasets (DS1, DS2, DS3, and DS4), and the results are shown in Tables 6, 7, 8, and 9. The evaluation of classifiers is presented in ascending order in all tables. For all datasets, the random forest has outperformed three other classifiers against five evaluation metrics. This established the efficacy of the random forest classifier as compared to three other classifiers on Twitter datasets. In

Table 6 Comparison of classifiers using Dataset 1

Classifiers	Accuracy	Precision	Recall	F1-Score	AUC
GBC	73.83	73.5	74.5	74.0	79.72
MLP	73.84	74.0	73.5	73.5	79.3
SVM	74.12	72.5	74.0	73.95	74.0
RF	77.38	77.5	77.0	77.0	84.48

Table 7 Comparison of classifiers using Dataset 2

Classifiers	Accuracy	Precision	Recall	F1-Score	AUC
MLP	79.05	79.26	85.98	81.85	88.79
SVM	81.10	80.13	86.23	82.90	91.92
GBC	81.69	81.20	87.45	84.01	92.87
RF	82.90	82.32	88.60	85.09	93.96

addition, random forest, as well as SVM, have been the most used models in the literature for rumour detection [7, 36].

On the other end, the significance of the hybrid combination of discrete emotion, linguistic, and metadata for rumour identification is also tested and we obtain the best values of all performance indexes (accuracy, precision, recall, f1-score, AUC) with random forest classifier as shown in Tables 6, 7, 8 and 9. In addition, as compared to DS1, DS2 and DS3, we obtain better performance indexes with DS2. The AUC measures (93.96% and 86.88%) are very effective on DS2 and DS3 respectively. We obtain 83% accuracy on DS2 which demonstrates the significance of the rumour detection model with random forest.

In addition, 82.32% precision, 88.60% recall, and 85.09% f1-score with DS2 are also effective indexes. On the other hand, on DS2, DS3, and DS4, the GBC classifier presented the second-best performance, whereas SVM and MLP are in the third and fourth rank. In contrast, the performances of classifiers on DS1 are ranked as RF, SVM, MLP, and GBC classifiers. Thus, we conclude that with random forest on DS1, we obtain at least 77% performance, on DS2, at least 82.32% performance, on DS3, at least 80.18% performance, and on DS4, at least 73.19% performance.

5.2 Feature-wise performance comparison

Exhaustive experiments are conducted to evaluate the significance of all proposed features as a stand-alone model, and comparisons with three state-of-the-art latest baselines for rumour detection on Twitter. The random forest classifier is selected because it outperformed others in prior experiments. For the experimental setup, five evaluation measures, 20-fold cross-validation, and four datasets are used. From Table 10, it is evident that the BERT model outperformed the other features as a standalone model on dataset 1. We obtain 96.7% accuracy and 99.5% AUC indexes that are very effective and validate the significance of bidirectional encode for rumour detection. In addition, we can note that all performance indexes are very promising with the BERT model. Likewise, the word2vec embedding model also demonstrated better performance as compared to linguistic, discrete emotion, and metadata features as a standalone model. Moreover, its performance is comparable with the BERT model. Thus both contextual models outperform the three textual models.

Table 8 Comparison of classifiers using Dataset 3

Classifiers	Accuracy	Precision	Recall	F1-Score	AUC
MLP	76.95	77.05	77.59	77.19	82.41
SVM	78.89	78.19	78.36	78.03	84.34
GBC	79.42	79.23	79.10	79.39	85.76
RF	80.18	80.32	80.19	80.45	86.88

Table 9 Comparison of classifiers using Dataset 4

Classifiers	Accuracy	Precision	Recall	F1-Score	AUC
MLP	70.11	73.01	70.12	70.18	74.98
SVM	71.39	74.04	71.29	71.34	76.45
GBC	72.24	75.19	72.32	72.43	77.81
RF	73.19	76.09	73.36	73.52	78.83

Among textual features, the linguistic model presented better performance as compared to the discrete emotion and metadata model. We can summarize that among textual models, linguistic features outperformed. The performance of three state-of-the-art latest baselines is also added as shown in Table 10. Huang, et al. method demonstrated better performance than two other baselines. It is also observed that BERT and Word2vec models presented much better performances as compared to three baselines as a standalone model. In hybrid combination, textual models (metadata + discrete + linguistic) also presented better performance as compared to the three baselines. In addition, a hybrid combination of Word2vec + BERT and all proposed features demonstrated much better performance indexes as compared to three standard baselines. This proves the significance of proposed BERT, Word2vec, and linguistic features as a stand-alone model and as a hybrid model using dataset 1. Hence, both textual and contextual features-based rumour detection models are robust.

Using dataset 2, once again BERT model outperformed all other features as a standalone model. But performance indexes on DS2 are less than performance indexes on DS1 when the BERT model is used. In addition, the Word2vec model presented the second-best performance as shown in Table 11. Hence, again contextual features outperformed the textual features. In textual features, linguistic features again outperformed the discrete emotions and metadata features. Therefore, we can summarize that the outstanding performance of linguistic features is consistent on datasets 1 and 2 in textual characteristics. Moreover, the prominent performance of the BERT model is also consistent upon both datasets in contextual features.

With dataset 2, It is observed that hybrid textual features also outperformed the three latest state-of-the-art baseline approaches as shown in Table 11. The performance of word2vec and BERT model in combination again demonstrated better than hybrid textual features. The best performance is achieved using hybrid combination of textual and contexual features. This

Table 10 Feature-wise performance using dataset 1

Features Type	Accuracy	Precision	Recall	F1-Score	AUC
Metadata	63.44	63.00	62.50	62.50	62.68
Discrete Emotions	66.73	68.50	68.00	66.50	67.37
Linguistic	71.57	71.50	71.50	71.50	79.21
Word2vec	96.6	97.00	97.0	96.50	99.67
BERT	**96.77**	**97.00**	**97.00**	**96.50**	**99.51**
Baseline 1 (Sicilia, et al. method)	73.58	74.10	73.12	72.46	81.56
Baseline 2 (Kumar, et al. method)	75.81	75.45	74.85	74.52	82.78
Baseline 3 (Huang, et al. method)	76.19	76.62	75.60	75.39	82.95
Metadata + Discrete + Linguistic	**77.38**	**77.5**	**77.0**	**77.0**	**84.48**
Word2vec + BERT	**96.18**	**96.0**	**96.50**	**96.50**	**99.62**
Metadata + Discrete + Linguistic + Word2vec + BERT	**96.65**	**97.0**	**96.59**	**96.64**	**99.63**

These are the best values obtained against each type of experiment

Table 11 Feature-wise performance using dataset 2

Features Type	Accuracy	Precision	Recall	F1-Score	AUC
Metadata	57.92	55.00	53.67	53.67	77.84
Discrete Emotions	67.48	74.33	60.33	65.67	76.68
Linguistic	79.86	77.00	76.67	77.00	89.98
Word2vec	81.60	81.50	87.00	84.50	93.69
BERT	**83.34**	**83.50**	**87.00**	**85.50**	**94.13**
Baseline 1 (Sicilia, et al. method)	80.13	78.65	77.73	78.88	90.65
Baseline 2 (Kumar, et al. method)	81.09	79.90	79.12	79.90	91.56
Baseline 3 (Huang, et al. method)	81.91	81.12	81.23	80.86	92.98
Metadata + Discrete + Linguistic	**82.90**	**82.32**	**88.60**	**85.09**	**93.96**
Word2vec + BERT	**84.74**	**84.67**	**89.0**	**86.05**	**94.42**
Metadata + Discrete + Linguistic + Word2vec + BERT	**85.15**	**85.10**	**91.67**	**86.93**	**95.47**

These are the best values obtained against each type of experiment

proves the significance of proposed textual and contextual features for rumour detection on Twitter using five performance indexes on DS1 and DS2.

On dataset 3 and dataset 4, same outstanding performances are observed as on dataset 1 and dataset 2 (Tables 12 and 13). The BERT model presented best performance as a standalone model but we get comparatively low threshold on dataset 4 as compared to first three datasets. Similarly, word2vec presented second best performance as a standalone model and outperformed the three standard baselines. Hence contextual features are more significant in identification of rumours at the tweet level as compared to textual features (evident from results on four datasets). The hybrid combination of linguistic, discrete emotions and metadata presented better performance than three baseline. In addition, the best performance is observed by using hybrid combination of contextual and textual features. This proves the significance of contextual and textual features for identification of rumours.

5.3 Feature importance

The importance of individual textual features for rumour detection is evaluated in this section. For the experimental setup, the random forest classifier runs a 20-fold cross-validation with an accuracy performance measure, and four datasets (DS1, DS2, DS3, and DS4) are used. Fifteen

Table 12 Feature-wise performance using dataset 3

Features Type	Accuracy	Precision	Recall	F1-Score	AUC
Metadata	56.75	57.00	57.00	56.00	56.99
Discrete Emotions	62.01	62.00	62.00	61.00	57.68
Linguistic	66.05	68.00	67.00	66.00	68.56
Word2vec	82.23	82.00	82.00	82.00	89.61
BERT	**83.59**	**84.00**	**84.00**	**84.0**	**91.70**
Baseline 1 (Sicilia, et al. method)	68.51	69.49	69.12	68.46	69.78
Baseline 2 (Kumar, et al. method)	70.74	71.23	71.19	70.61	71.85
Baseline 3 (Huang, et al. method)	71.84	72.92	72.79	71.93	72.98
Metadata + Discrete + Linguistic	**80.18**	**80.32**	**80.19**	**80.45**	**86.88**
Word2vec + BERT	**84.14**	**85.00**	**83.00**	**84.70**	**92.52**
Metadata + Discrete + Linguistic + Word2vec + BERT	**84.93**	**86.00**	**85.00**	**85.00**	**92.89**

These are the best values obtained against each type of experiment

Table 13 Feature-wise performance using dataset 4

Features Type	Accuracy	Precision	Recall	F1-Score	AUC
Metadata	54.15	54.02	54.10	53.02	55.04
Discrete Emotions	54.84	55.09	54.87	54.19	55.87
Linguistic	60.38	60.00	60.00	60.00	59.22
Word2vec	74.41	78.10	74.07	74.05	79.50
BERT	**78.10**	**81.00**	**78.05**	**78.14**	**84.73**
Baseline 1 (Sicilia, et al. method)	65.58	66.10	65.12	64.96	69.56
Baseline 2 (Kumar, et al. method)	67.81	67.45	67.85	67.52	70.78
Baseline 3 (Huang, et al. method)	69.19	69.62	70.60	69.39	72.45
Metadata + Discrete + Linguistic	**73.19**	**76.09**	**73.36**	**73.52**	**78.83**
Word2vec + BERT	**78.71**	**79.50**	**79.40**	**79.41**	**85.37**
Metadata + Discrete + Linguistic + Word2vec + BERT	**79.76**	**80.00**	**80.50**	**80.76**	**86.62**

These are the best values obtained against each type of experiment

textual features are evaluated individually and their performance is presented in Figs. 6, 7, 8, and 9 respectively.

For dataset 1, 'URLs' is observed to be the most effective feature for detecting rumour at the tweet level. 'Trust emotion' is the second best, and 'Verbs' is the third-best feature. The prominence of the 'URLs' feature reveals that rumour tweets contain more URLs than non-rumour tweets. In the same context, rumour tweets comparatively use more trust-related emotional words. Next 'Adjectives' and 'Prepositions' are the fourth and fifth-best features. It uncovers that tweets embedded with more 'adjectives and prepositions' have the maximum probability to be rumours. The '#Full stop and Sadness emotion' are the next best characteristics for rumour identification using dataset 1.

Using DS2, we again find 'URLs' to be the best feature for identifying a tweet as a rumour as shown in Fig. 7. Moreover, this feature also demonstrated the best performance with DS1. Thus the effectiveness of the 'URLs' feature is consistent in both datasets. The second best feature is 'Trust' whereas 'Verbs and Adjectives' are the third and fourth-best features. It is being observed that top-4 features have a consistent performance on dataset 1 and dataset 2 (Figs. 6 and 7). While differences are also being observed like '#Full Stop' is at position 5 on DS1 but is shifted to 6th position on DS2, instead, 'Sadness' comes at position 5 on DS2. In addition, '#hashes and Anticipation' features switch their positions on DS1 and DS2. The overall performances of the fifteen features are consistent on both datasets.

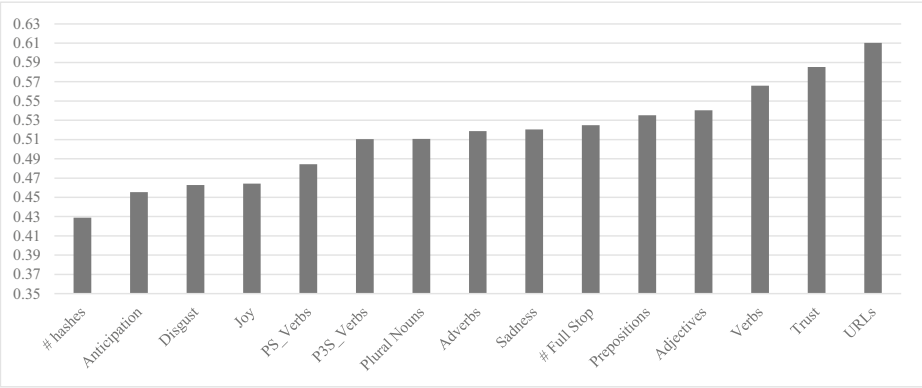

Fig. 6 Importance of fifteen content features using accuracy measure (dataset 1)

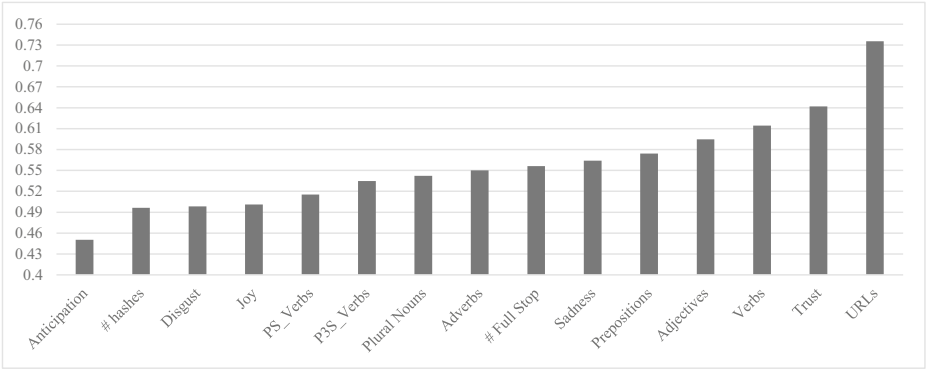

Fig. 7 Importance of fifteen content features using accuracy measure (dataset 2)

Using DS3 and DS4, we can observe the similar performances of individual features as we observed on DS1 and DS2. The performance of the top-5 features is 100% consistent. However, 'Joy and Disgust emotions' features switch each other at 12th and 13th positions. In addition, the '#Full Stop' feature is at position 8 on DS3 and DS4 whereas it is at position 7 on DS2 and at position 6 on DS1. The 'anticipation and #hashes' features switch their positions on DS3 and DS4 as compared to DS2. Thus, after experiments on four datasets, we can conclude that 'URLs, Trust, Verbs, Adjectives, Prepositions, Sadness, Adverbs, and #Full Stop' are the top-8 textual features to identify rumour at the tweet level and their performance is almost consistent on four datasets.

6 Discussions and implications

This research has improved the accuracy of the detection model for rumours in specific events of health-domain, Ottawa shooting and Germanwings crash events on social microblog platforms and presented a robust detection model with 96.7% accuracy. Mainly, five types of features: BERT model, word2vec embeddings, discrete emotions, linguistics, and metadata are investigated. The textual features are further considered for feature selection and the two-step method is adapted to identify the most important features. The rumour detection model based on the random forest classifier is finally designed, which outperformed the three latest

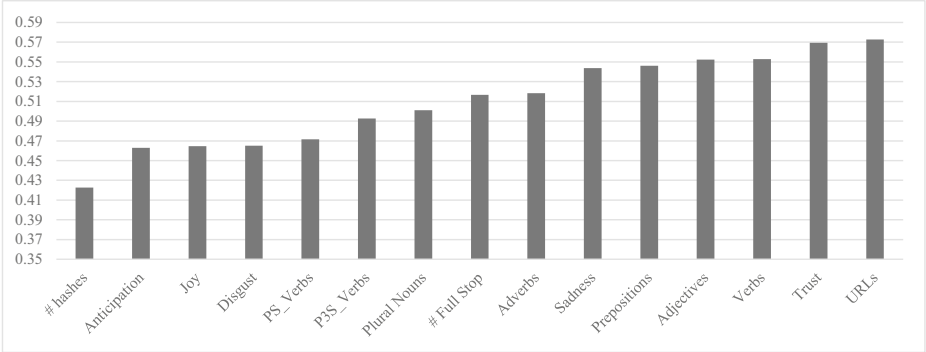

Fig. 8 Importance of fifteen content features using accuracy measure (dataset 3)

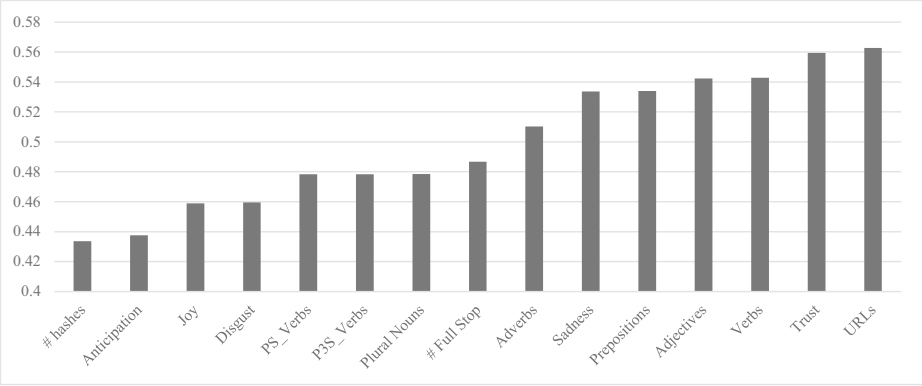

Fig. 9 Importance of fifteen content features using accuracy measure (dataset 4)

standard baselines. Like few prior studies, our research uses a single observation window to generate the results. However, there are, some studies in literature, which are inspired by rumour identification, that changes over time. Therefore, using a single observation window is one of the limitations of our research, and findings cannot be generalized for all cases.

From a theoretical perspective, this research has reduced the training time and complexity of the rumour detection model as compared to prior models. In addition, our research used those features that enhance the accuracy of the model and are stable at the tweet level. All these objectives are achieved using an influential set of features and a robust ML model. If we consider network features for rumour identification, they are comparatively difficult to extract as well as dynamic in nature (changing over time and extracting a user network graph is more complex). In contrast, our methodology provides more optimal solution in less time.

If we look practically at this research, it is more applicable and relevant to social media platforms where everyone can share their opinions freely which could cause rumours to spread. These platforms should have a system for detecting fake/rumour at the post or tweet level. Also, news agencies often use popular social media platforms to gather information and need such a system to detect rumour/fake information at an early stage. Most of the time, this rumour/fake news not only targets news agencies but also makes losses at the national and worldwide levels. This research delivers a solid mechanism to detect such rumours efficiently at the tweet/post level.

7 Conclusions

This case study developed a framework for rumour detection that works at the tweet level in the specific health, Ottawa shooting and Germanwings crash events. The framework is different from other literature approaches in the sense that it did not incorporate the use of topic information as a feature and thus avoids any prior domain-related assumptions. Two types of contextual and three types of textual characteristics are proposed to investigate their impact as a standalone model and as a hybrid model on the detection of rumours. The performance of four classifiers is compared by running on 20-fold cross-validation and four

real-life datasets. Our model presented 97% accuracy on dataset 1, 85% accuracy on dataset 2, 85% accuracy on dataset 3, and 80% accuracy on dataset 4, which is far better than the three latest state-of-the-art baselines. BERT model presented the best performance among applied contextual features and linguistic features presented the best performance among applied textual features. Moreover, the best textual features are selected using the two-step feature selection method. Generally, the BERT model presented the best performance as a standalone model. The findings indicate that 'URLs, Trust emotion, Verbs, Adjectives, and Propositions' are the five best textual features for rumour detection.

In the future, some extensions can be made. First, these experiments are restricted to four specific datasets. The framework can be utilized for other domain datasets. Second, the proposed system can be applied in other domains, such as fraud detection and security, etc. Third, new social and semantic characteristics could be incorporated to improve the detection model accuracy. Fourth, evolutionary algorithms or ensemble models can be applied to build a robust detection model.

Data availability The dataset is publicly available.

Code availability The code of the experiment is not available.

Declarations

Conflicts of interest/Competing interests There is no conflict of interest.

References

1. Alzanin S, Azmi A (2019) Rumor detection in Arabic tweets using semi-supervised and unsupervised expectation–maximization. Knowledge-Based System 185: 104945
2. Bai N, Wang Z, Meng F (2020) A stochastic attention CNN model for rumor stance classification. IEEE Access 8:80771–80778
3. Bian T et al (2020) Rumor detection on social media with bi-directional graph convolutional networks. In: Proceedings of the AAAI conference on artificial intelligence 34 (1), pp 549–556
4. Bird S, Klein E, Loper E (2009) Natural language processing with Python: analyzing text with the natural language toolkit. O'Reilly Media, Inc.
5. Breiman L (2001) Random forests. Mach Learn 45(1):5–32
6. Brill E (1992) A simple rule-based part of speech tagger. Pennsylvania Univ Philadelphia Dept of Computer and Information Science
7. Castillo C, Mendoza M, Poblete B (2011) Information credibility on twitter. In: Proceedings of the 20th international conference on World wide web, pp 675–684
8. Chen Y, Hu L, Sui J (2019) Text-based fusion neural network for rumor detection. In: International Conference on Knowledge Science, Engineering and Management. Springer, Berlin, pp 105-109
9. Devlin J, Chang M-W, Lee K, Toutanova K (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805
10. Domm P (2013) False rumor of explosion at White House causes stocks to briefly plunge. AP confirms its Twitter feed was hacked. CNBC. COM, vol 23
11. Figueira Á, Oliveira L (2017) The current state of fake news: challenges and opportunities. Procedia Comput Sci 121:817–825
12. Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Annals of statistics: 1189–1232
13. Hamidian S, Diab MT (2019) Rumor detection and classification for twitter data. arXiv preprint arXiv: 1912.08926
14. He X, Tuerhong G, Wushouer M, Xin D (2022) Rumors detection based on lifelong machine learning. IEEE Access 10:25605–25620

15. Hinton GE (1990) Connectionist learning procedures. Artificial Intelligence, 40 1–3: 185 234, 1989. Reprinted in J. Carbonell, editor. Machine Learning: Paradigms and Methods. MIT Press
16. Huang Q, Zhou C, Wu J, Wang M, Wang B (2019) Deep structure learning for rumor detection on Twitter. In: International Joint Conference on Neural Networks (IJCNN). IEEE, pp 1–8
17. Huang Q, Yu J, Wu J, Wang B (2020) Heterogeneous graph attention networks for early detection of rumors on Twitter. In: International Joint Conference on Neural Networks (IJCNN). IEEE, pp 1–8
18. Kumar A, Sangwan SR, Nayyar A (2019) Rumour veracity detection on twitter using particle swarm optimized shallow classifiers,. Multimed Tools Appl 78(17):24083–24101
19. Kwon S, Cha M, Jung K, Chen W, Wang Y (2013) Prominent features of rumor propagation in online social media. In: IEEE 13th International Conference on Data Mining. IEEE, pp 1103–1108
20. Kwon S, Cha M, Jung K (2017) Rumor detection over varying time windows,. PLoS ONE 12(1):e0168344
21. Lotfi S, Mirzarezaee M, Hosseinzadeh M, Seydi V (2021) Detection of rumor conversations in Twitter using graph convolutional networks. Appl Intell 51(7):4774–4787
22. Ma J et al (2016) Detecting rumors from microblogs with recurrent neural networks, In: Proceedings of the 25th International Joint Conference on Artificial Intelligence, p 3818–3824
23. Ma J, Gao W, Wong K-F (2018) Rumor detection on twitter with tree-structured recursive neural networks. Association for Computational Linguistics
24. Malik MSI (2020) 'Predicting users' review helpfulness: the role of significant review and reviewer characteristics. Soft Computing, pp 1–16
25. Malik M, Hussain A (2017) Helpfulness of product reviews as a function of discrete positive and negative emotions,. Comput Hum Behav 73:290–302
26. 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 Systems with Applications, 153: 112986
27. Mihalcea R, Strapparava C (2009) The lie detector: Explorations in the automatic recognition of deceptive language. In: Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, pp 309–312
28. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J (2013) Distributed representations of words and phrases and their compositionality. arXiv preprint arXiv:1310.4546
29. Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781
30. Mohammad SM, Turney PD (2013) Nrc emotion lexicon,. National Research Council, Canada, vol 2
31. Nayak P (2019) Understanding searches better than ever before. The Keyword, vol 295
32. Pedregosa F et al (2011) Scikit-learn: Machine learning in Python. J Mach Learn Res 12:2825–2830
33. Platt J (1999) Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Adv Large Margin Classifiers 10(3):61–74
34. Ruchansky N, Seo S, Liu Y (2017) Csi: A hybrid deep model for fake news detection. In: Proceedings of the ACM on Conference on Information and Knowledge Management, pp 797–806
35. Sharma K, Qian F, Jiang H, Ruchansky N, Zhang M, Liu Y (2019) Combating fake news: A survey on identification and mitigation techniques,. ACM Trans Intell Syst Technol (TIST) 10(3):1–42
36. Sicilia R, Lo Giudice S, Pei Y, Pechenizkiy M, Soda P (2018) Twitter rumour detection in the health domain,. Expert Syst Appl 110:33–40
37. Song C, Yang C, Chen H, Tu C, Liu Z, Sun M (2019) CED: credible early detection of social media rumors. IEEE Trans Knowl Data Eng
38. Tian Y, Fan R, Ding X, Zhang X, Gan T (2020) Predicting rumor retweeting behavior of social media users in public emergencies,. IEEE Access 8:87121–87132
39. Vijeev A, Mahapatra A, Shyamkrishna A, Murthy S (2018). A hybrid approach to rumour detection in microblogging platforms. In: 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI) IEEE, pp 337–342
40. Wang WY (2017) "Liar, liar pants on fire": A new benchmark dataset for fake news detection. arXiv preprint arXiv:1705.00648
41. Wang S, Kong Q, Wang Y, Wang L (2019) Enhancing rumor detection in social media using dynamic propagation structures. In: IEEE International Conference on Intelligence and Security Informatics (ISI). IEEE, pp 41–46
42. WHO (2020) FACT: Drinking alcohol DOES NOT protect you against #COVID19 and can be dangerous. [Scial microblog]
43. WHO (2020) FACT: Being able to hold your breath for 10 seconds or more without coughing or feeling discomfort DOES NOT mean you are free from the #coronavirus disease or any other lung disease. [Scial microblog]
44. Wikipedia (2021) https://en.wikipedia.org/wiki/Discrete_emotion_theory. Access dates 22 March 2022

45. Xu R, Xia Y, Wong K-F, Li W (2008) Opinion annotation in on-line Chinese product reviews. In: LREC, vol 8, pp 26–30
46. Yang S, Yao J, Qazi A (2020) Does the review deserve more helpfulness when its title resembles the content? Locating helpful reviews by text mining,. Inf Process Manag 57(2):102179
47. Zhang Y, Hara T (2020) A probabilistic model for malicious user and rumor detection on social media, pp 1–10
48. Zubiaga A, Liakata M, Procter R, Wong SH, Tolmie P (2016) Analysing how people orient to and spread rumours in social media by looking at conversational threads. PLoS ONE 11(3):e0150989