False information detection in online content and its role in decision making: a systematic literature review

Article *in* Social Network Analysis and Mining · September 2019

DOI: 10.1007/s13278-019-0595-5

Adil Khan
Bahria University Islamabad Campus

Anam Habib
Gomal University

Aurangzeb Khan
University of Science and Technology Bannu

Social Network Analysis and Mining (2019) 9:50
https://doi.org/10.1007/s13278-019-0595-5

REVIEW ARTICLE

False information detection in online content and its role in decision making: a systematic literature review

Ammara Habib[1] · Muhammad Zubair Asghar[1] · Adil Khan[2,3] · Anam Habib[1] · Aurangzeb Khan[4]

Received: 14 January 2019 / Revised: 26 August 2019 / Accepted: 5 September 2019 / Published online: 17 September 2019

Abstract
This work presents a review of detecting false information and its role in decision making spread across online content. The authenticity of information is an emerging issue that affects society and individuals and has a negative impact on people's decision-making capabilities. The purpose is to understand how different techniques can be used to address the challenge. The approach used for the identification of published articles between 2014 and 2018 is the systematic literature review in which 30 papers were identified and the relevant articles were selected by applying inclusion–exclusion criteria. This review classifies the false information, spreading on social media, into four types. Furthermore, we describe four deep learning and eight machine learning techniques for false information detection. The outcomes of this review will provide the researchers with an insight into the different types of false information, associated detection techniques, and the relationship between false information and decision making. In the field of false information detection, previous studies provided a review of the literature. However, we conducted a systematic literature review by providing specific answers to the proposed research questions. Therefore, our contribution is novel to the field because this type of study is not performed previously.

Keywords False information · Social media · Systematic review · Decision making · Deep learning · Machine learning · Rumor detection

Abbreviations

AUC	Area under the curve
RST–VSM	Rhetorical structure theory and vector space modeling
PKT	Propagation kernel tree
CERT	Cross-topic emerging rumor detection
SRDC	Single-step RDC
TRDC	Two-step RDC
TF-IDF	Term frequency-inverse document frequency
att-RNN	Recurrent neural network with an attention mechanism
LSTM	Long short-term memory
CNN	Convolutional neural network
3HAN	Three-level attention network
MLP	Multilayer perceptron
RNN	Recurrent neural network
CSI	Capture, score and integrate
RST	Rhetorical structure theory

✉ Muhammad Zubair Asghar
zubair@gu.edu.pk

Ammara Habib
ammarahabib10@gmail.com

Adil Khan
adil.adil25@yahoo.com

Anam Habib
anamhabib19@gmail.com

Aurangzeb Khan
aurangzeb.ustb@gmail.com

[1] Institute of Computing and Information Technology, Gomal University, D.I. Khan, KP, Pakistan

[2] Higher Education Department, KPK, Peshawar, Pakistan

[3] School of Computer Science and Technology, SZIC, Computer Science Department, University of Peshawar, Peshawar, Pakistan

[4] Department of Computer Science, University of Science and Technology, Bannu, Pakistan

1 Introduction

Online social networks have become an ideal platform to access information and are continuing to invade our culture (Zhou et al. 2018). To persuade user's decisions and opinions, online content has played a significant role in recent times (Ahmed et al. 2017). Social media serves as a prime source of information for a huge community of online users (Sharma et al. 2019).

Nowadays most of the people spent their time on online social platforms to communicate with world and use social media to engross news and seek out information instead of traditional news media, because information propagation on social media takes very less time and also less expensive rather than traditional news media, e.g., television and newspaper (Granik and Mesyura 2017). For instance, in the USA 49% of the adult population has accessed social media to share information in 2012, while in 2016 over 62% reported grasp news on social media on a daily basis (Wu et al. 2015). However, during the last year, the rapid diffusion of false information raises serious concern because of the fact that social media plays a vital role to influence people daily decisions in political, social and economic domains. Therefore, false information detection in social media is a matter of concern (Sharma et al. 2019).

1.1 The need for false information detection

False information is a global challenge and a global issue. It has gained a lot of recognition from multiple stakeholders in which journalist, non-governmental organizations, civil society, politicians and researchers are involved. Moreover, cyberbullying, abusive language and hate speech are also carrying out online. False information has real-life consequences such that it may alter people decisions, opinions and beliefs. Therefore, false information detection is of the utmost importance (Granik and Mesyura 2017).

Usually, false information is associated with time-critical and newly emerging events which are not verified accurately by existing studies. On the other hand, the data that are produced through false information are unstructured, incomplete and noisy, so efficient methods are required to identify non-credible information (Granik and Mesyura 2017).

Research focusing on false information detection is still in its early stages, so to minimize the impact of false information, further research needs to be carried out (Ahmed et al. 2017). In this review, we attempt to provide an overview of different types of false information, detection techniques and the role of false information detection in decision making.

1.2 Research motivation

The key motivation of this review is on the following grounds.

1. False information detection is a challenging issue, and it is still in its early stages of development which urges further investigation. So it is essential to explore further research directions to enrich existing false information detection techniques.
2. Rapid research advancements made in false information detection have motivated us to conduct the systematic review by searching, identifying, summarizing and evaluating relevant studies.

1.3 Our contributions

Our major contribution made in this review is as follows.

1. Classify the false information into rumors, fake news, misinformation and hoax.
2. Discuss the role of different types of false information: rumors, fake news, misinformation and hoax in decision making.
3. An overview of existing machine learning methods used for the detection of false information detection is presented.
4. Explore different types of deep learning techniques used for false information detection.

1.4 Relation to the previous work

In this section, we present an overview of the related surveys conducted on false information detection.

The objective of a survey conducted by Shu et al. (2017) is to detect fake news in social media content. Fake news problem is explored in two phases, i.e., characterization and detection. The basic principles and concepts of both social media and traditional media are introduced in the characterization phase. For the detection phase, existing approaches are reviewed from the data mining perspective. Various evaluation metrics are used to check the performance of the classifier in which area under the curve (AUC) is more statistically consistent than accuracy. There is a serious negative impact on the individual and society due to the extensive spread of false information and fake news such as to obtain political and financial gain, but this survey did not address his issue.

Zubiaga et al. (2018) conducted a survey on the detection and resolution of rumors in social media by taking into account two types of rumors. First one is the long-standing rumors that disseminate for a long period of time and the second one is that which spread rapidly like breaking news, also called newly emerging rumors. However, they addressed the approaches which deal with the rumors, whereas there is a need to carry investigate other types such as misinformation.

In their work on rumor diffusion in Twitter (Serrano et al. 2015), rumor dissemination is a review along four dimensions, namely: exploratory data analysis, rumor detection, epidemiological modeling and multiagent-based social simulation. Further analysis is required to upgrade and assimilate research presented in this area as the research is persistently growing in the development of rumor/false information verification due to the proliferation of social media.

The remainder survey is assembled in the following sections. Section 2 provides a detailed taxonomy of the survey conducted. Section 3 presents a discussion on the comparative results, and finally, Sect. 4 presents the overall conclusions for this paper.

2 Survey methodology

The methodology followed in this survey is presented as follows.

2.1 Survey protocol

In this work, we used different electronic repositories to search related articles. In the next step, inclusion and exclusion criteria are applied to filter a number of acquired articles. Finally, on the basis of research questions, relevant works are selected, and after detailed analysis, results are reported.

2.2 Research questions

The following research questions are addressed.

RQ1: What are different types of false information and what is their role in decision making as identified in the published literature?
RQ2: What are the different machine learning techniques used for the detection and classification of false information as reported in the literature?
RQ3: Why deep learning is important for false information detection and what are the available techniques to perform false information detection and classification with deep learning perspective?

2.3 Data sources

To identify relevant research articles, we used different digital libraries, such as ACM Digital Library (www.acm.org/dl), IEEE Xplore (https://ieeexplore.ieee.org), Google Scholar (https://scholar.google.com.pk), Springer Link (https://link.springer.com), and Science Direct (www.sciencedirect.com).

In the next step, inclusion and exclusion criteria are applied for selecting the most relevant articles.

2.4 Inclusion–exclusion principle

To retrieve relevant research articles, a systematic keyword-based search has been carried out by posing different search queries, as shown in Table 1. The inclusion and exclusion principle is applied to determine whether a study should be included or otherwise. Inclusion principle works as: the search queries are relevant to the study, the articles are published in journals and conferences proceedings and the articles are written in English. The exclusion principle includes articles not focused on the false information in social media and articles not following the inclusion criteria.

3 Survey classification

This section presents a detailed summary of the survey conducted on false information detection and associated techniques which will help to find the research gaps and proposing solutions for false information detection. This review is carried out in the following directions (Fig. 1): rumors, fake news, misinformation, and hoax.

3.1 Types of false information and what is their role in decision making

There are different types of false information, namely (1) rumors, (2) fake news, (3) misinformation and (4) hoax.

3.1.1 Rumors

Rumor is a piece of information whose truthfulness is in doubt and source is unreliable and probably produces under the emergency situation which creates panic in public, diminishing the government credibility, disturbs the social order and even threatens the national security (Liang et al. 2015). Rumors can lead to emotions like fear and anxiety so in such cases people cannot make decisions as rationally

Table 1 List of keywords for searching relevant articles

Machine learning + rumor	Machine learning + fake news	Machine learning + misinformation	Machine learning + hoax	Deep learning + rumor	Deep learning + fake news	Deep learning + misinformation	Deep learning + hoax
Rumor detection	Online fake news detection	Misinformation identification	Hoax detection in social networks	Identification of rumor	Detection of fake content	Online misinformation	Hoax identification in social media

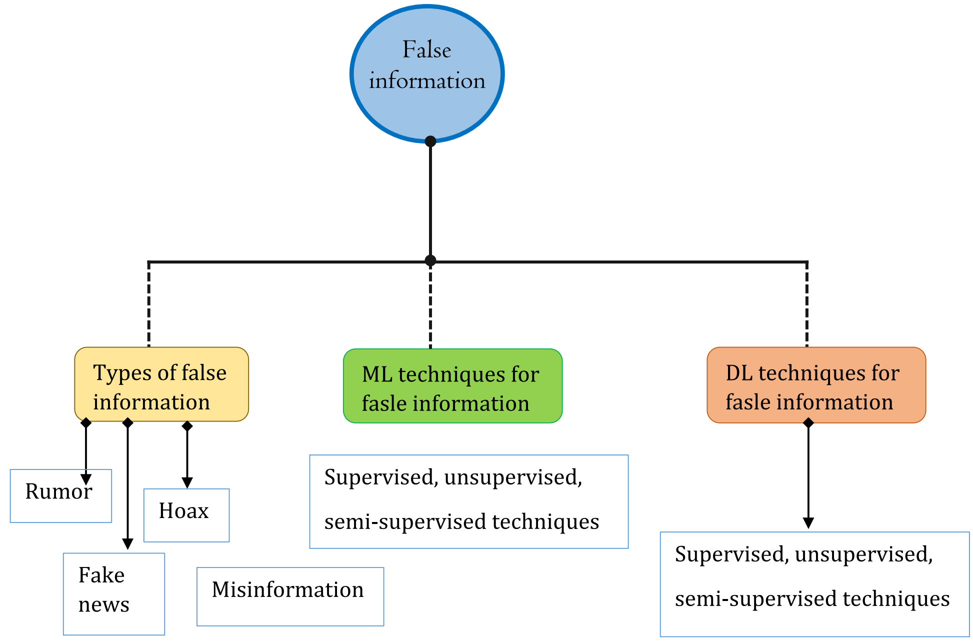


Fig. 1 Survey classification diagram

as they otherwise do (Jin et al. 2017). For example, after a powerful earthquake in Japan, on March 12, 2011, a rumor had been widespread across China via Sina Weibo and other microblog networks that iodized salt was capable of protecting people from nuclear radiation. So to buy salt, people rushed to store, supermarket and dispensaries. During that time, the iodized salt price had increased 5–10 times. Therefore, it is an essential task to detect and report rumors from microblogging sites before they widespread (Stantic 2017).

3.1.2 Fake news

Fake news is written and published with the intent to mislead its readers in order to gain the reader's attention with mala fide financial or political intentions, executed by sensationalist, exaggerated, or patently false headlines (Allcott and Gentzkow 2017).

On March 11, 2011, one of the worst effects of fake news in China was an incidence after the Fukushima Daiichi nuclear disaster. In a microblog, fake news was posted claiming that salt can prevent radiation, but the nuclear disaster makes salt contaminated. So people in China stockpiled table salt due to the fake news, which resulted in a huge demand–supply disorder in the market (Chen et al. 2016).

3.1.3 Misinformation

It is false or inaccurate information, communicated intentionally or unintentionally, with an attempt to present it as being true.

In the USA, both politicians and media accomplish some of the misinformation campaigns to convince the public. For example, in the 9/11 attack, through misinformation propagation, Saddam Hussein was made involved some way. Such misinformation has an adverse impact on society, because believing in such misinformation often leads to wrong decisions (http://yris.yira.org/acheson-prize/2473).

3.1.4 Hoax

Electronic messages with evil intention to misguide recipients consist of audio, text and multimedia content. A hoax news spread from fake news website ABCNEWS.com, claimed that in schools president Obama does not allow the recital of Pledge of Allegiance.

Hoaxes are spread with political or some other special interests, and due to the lack of media literacy, people are unable to distinguish between real news and hoaxes spread on social media (https://www.factcheck.org/2016/09/obama-did-not-ban-the-pledge/).

3.1.5 Differentiation between rumor, fake news, misinformation and hoax

The differentiation of the rumor, fake news, misinformation, and hoax is described as follows:

Once the authoritative sources confirmed a statement as false or fabricated, then it is labeled as a *rumor* (Wu et al. 2015). On the other hand, *fake news* involves the information which contradicts the ground-truth information or makes it harder for the reader to distinguish true from false information, and make them believe in the false version of the information (Pomerantsev and Weiss 2014). *Misinformation* spread without an intent to deceive the reader. Thus, the causes of misinformation involve lack of understanding, attention or misrepresentation of an original piece of true information which is spread via tweets, blogs and articles (Fallis 2009). And *hoax* is a false story to deceive the truth, meaning "to cheat" usually spread through the news outlet to gain politically and financial benefits (Hunt 2016).

3.1.6 Literature for the role of false information in decision making

To detect deceptive news from online content, discourse analysis is performed using the rhetorical structure theory and vector space modeling (RST–VSM) which is proposed by (Rubin et al. 2015). The experiment shows that the proposed model achieved 63% accuracy. Further improvements are required to overcome the limitations toward news verification system.

To identify fake news, tensor modeling is proposed to automatically identify false information by categorizing it into different types. To refine and consolidate the results of tensor decomposition further, ensemble method is proposed. The proposed method outperformed the existing state-of-the-art methods with an accuracy of up to 80%. By increasing the number of a top news article to 25 or larger, the number of an outlier is also increasing (Sharma et al. 2019).

Dalal et al. (2015) in their work on analyzing the effect of rumors on hiring decisions proposed Mechanical Turk human intelligence system. Experimental results show that hiring decisions are affected by rumors. However, the study was not ecologically valid due to the use of rumor manipulation.

To address the problem of identification of veracity of rumors in social media, a number of classifiers are used: (1) random forest, (2) decision tree, (3) support vector machine, (4) logistic regression and (5) Naïve Bayes. Results conclude that decision tree produced the best performance with an accuracy of 96.6% and precision 1.0. One of the future areas of this work is to build another classifier that automatically classifies the tweets into rumor and annotates them. As the proposed system manually categorized the tweets into rumors, it is a time-consuming task and requires much time and effort (Giasemidis et al. 2016).

Theoretical analysis of rumors during social crises on Twitter data is carried out by Oh et al. 2013. For this purpose, a logistic regression model is proposed. Three social incidents are taken into account: (1) Seattle cafe shooting, (2) Mumbai terrorist attack and (3) Toyota recalls. Experiments show a good model fit for Seattle shooting effect with χ^2 value of 292.69, Mumbai terrorist attack with χ^2 value of 680.21 and Toyota recall data with χ^2 value of 260.94.

3.1.7 Impact of false information on decision making

We make decisions in our both personal and professional lives. With the rise of false information, there is a risk that a number of wrong decisions will reach to unprecedented levels; this is because false information spreads faster, wider, further and deeper than correct information (https://www.vlerick.com/en/about-vlerick/news/are-your-decisions-influenced-by-fake-news). There is a 70% greater chance for false information to disseminate than the truth according to the researchers. False information can also affect the reputation of the business. For instance, false information spread about Starbucks that it would be giving discounts to undocumented immigrants (https://www.businessinsider.com/fake-news-starbucks-free-coffee-to-undocumented-immigrants). Furthermore, false information on the platforms such as social media and the web affects the stock market (Bollen et al. 2011) and had slowed responses during disasters (Gupta et al. 2013) and the terrorist attack (Fisher 2016), thereby influencing people opinion and decisions. According to the World Economic Forum (Howard and Kollanyi 2016), the impact of false information makes it one of the modern dangers to society. Therefore, to detect and prevent false information on the web and social media, there is a need for an effective detection method.

3.2 Machine learning techniques for detection and classification of false information

3.2.1 Introduction

Machine learning is an artificial intelligence discipline that gives the system the capability to automatically learn and improve from experience without being explicitly programmed. Some of the machine learning methods are: (1) supervised learning method: prediction is made on the basis of labeled examples; (2) unsupervised learning method: it uses unlabeled information for classification; (3) semi-supervised learning algorithm: it uses both labeled and unlabeled data for training, i.e., it falls somewhere in between supervised and unsupervised learning (Habib et al. 2019).

Social network sites are in trouble as these are considered an ideal platform for the spreading misinformation. Therefore, the detection of false news needs immediate attention. For this purpose, researchers get benefits to form machine learning algorithms that give an accurate identification of false news and credible news (Lipton 2018).

3.2.2 Machine learning-based classification of false information

To classify information into false and credible, different machine learning techniques (e.g., support vector machine, Naïve Bayes, K-nearest neighbor and logistic regression can be applied (Zubiaga et al. 2018). In this section, we present an overview of how a machine learning classifier can be applied for the classification of online content into credible and false information. Following are the different steps to be applied for the classification of false information: (1) *Preprocessing* This step aims at reducing noise in the text by performing stemming, lemmatization, lowercase, and remove stop words to improve the performance of classifier (Khan et al. 2019; Asghar et al. 2019), (2) *Feature Extraction* To obtain the relevant and useful information for classifying the online content, feature extraction technique is applied and the extracted information is made input to the next stage, (3)*Train–test split* The dataset is divided into two parts: training and testing, where training is used for learning the algorithm and testing is used for the evaluation of the performance of the model, and (4) *Applying machine learning classifier* The ML model estimates the features belonging to a specific class (credible information or false information) by using train data to predict the class of unseen data. Each of the ML classifiers has its own mathematical/statistical model to predict whether the input information is either credible or false as shown in Fig. 2.

3.2.3 Machine learning techniques for rumor detection

In the following paragraphs, we represent the review of machine learning techniques for rumor detection, grouped together on the basis of supervised learning method.

Supervised learning method A novel model, called propagation kernel tree (PKT), is proposed for rumor detection using SVM classifier in a supervised learning framework (Ma et al. 2017). Experiments on Twitter dataset show that for both general and early rumor detection, the proposed system outperforms state-of-the-art baseline. Rumor detection task can be improved by exploring network learning representation framework. Health-related rumors on Twitter are captured using a pipeline of health professionals, crowdsourcing and machine learning techniques including Naïve Bayes, random forest and random decision tree (Ghenai and Mejova 2017). Results show that random tree classifier is the best, achieving a precision of 0.946 and recall 0.944. The main limitation of this work is that only a handful of rumors are considered due to the limited resources. To perform early detection of emerging rumors in social media, Wu et al. (2017) proposed a cross-topic emerging rumor detection (CERT) framework-based sparse representation model. Experimental results depict that CERT detects emerging rumors faster than existing approaches. The system can be enhanced by incorporating a cross-modal information. In their work on misinformation propagation, Hamidian et al. (2015) worked on rumor detection and rumor classification using a supervised method, namely J48 classifier, implemented in WEKA platform under a single-step RDC (SRDC) and two-step RDC (TRDC). Results show that for the MIX data set, TRDC attains an F measure of 82.9%, which is better than that of comparing SRDC. The limitation of the method is that preprocessing does not give benefit to the model due to the weakness of the tools used for processing Twitter content.

Yang et al. (2015) proposed an automatic rumor detection system based on hot topic detection and a set of features.

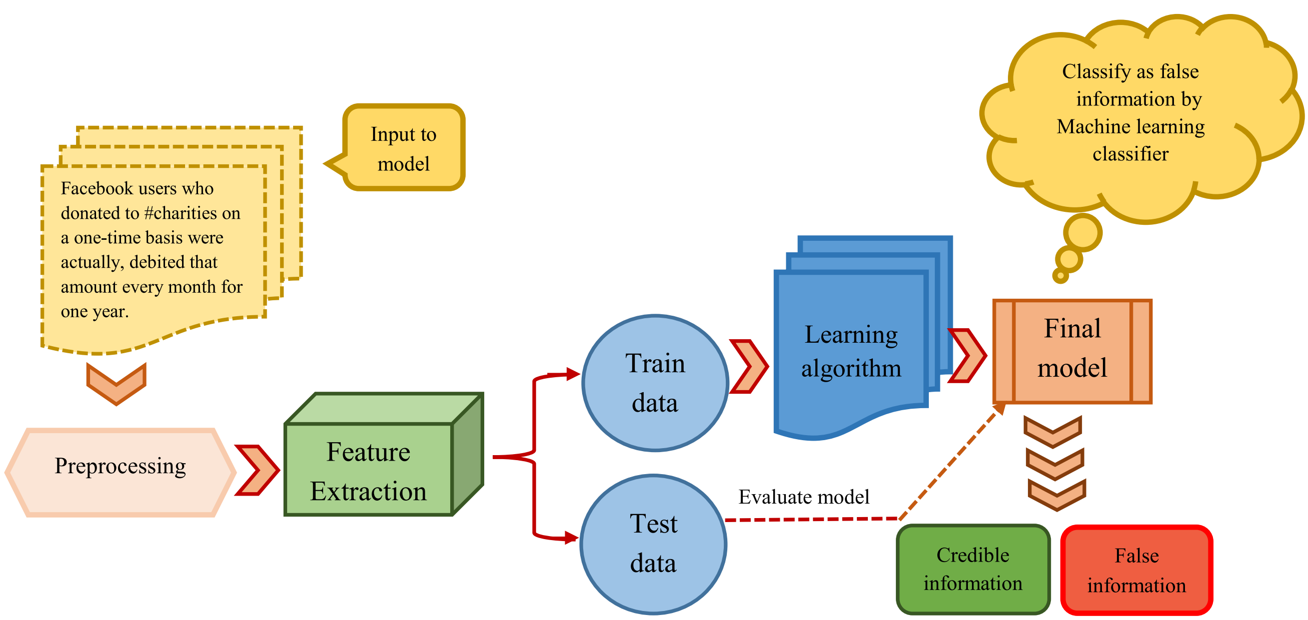


Fig. 2 Machine learning diagram

For this purpose, they compared the performance of Naive Bayes, logistic regression and random forest on rumor classification task by conducting an extensive set of experiments. The experimental results show that the random forest achieves the best result. To improve the result, more sophisticated probabilistic models with more features are investigated.

3.2.4 Machine learning techniques for fake news

For fake news, the machine learning techniques are grouped according to supervised, unsupervised and semi-supervised methods.

Supervised learning method Ahmad et al. (2017) proposed an online fake news detection system based on *n* gram and machine learning techniques. The results reported show that an accuracy of 92% is achieved with term frequency-inverse document frequency (TF-IDF) as feature extraction and linear SVM as a classifier. The accuracy of the algorithm decreases with the increase in *n* gram. In another work Granik and Mesyura (2017) used Naïve Bayesian classifier for the fake news detection and achieved an accuracy of 74%. Due to the skewness of dataset, the classification accuracy for fake news is slightly worse. For fake news detection, Fairbanks et al. (2018) proposed a novel framework by addressing two problems, namely (1) bias detection and (2) credibility assessment. For bias detection content model is established, and for credibility assessment structural model is developed. Experimental results show using TF-IDF, matrix logistic regression outperforms the random forest. For automatic identification of fake news, a classification model is proposed comprising of lexical, syntactic and semantic information, with the additional support of textual features (Pérez-Rosas et al. 2017). The model achieved the best performance as compared to the human ability to spot fake news. There is a significant loss in accuracy in the cross-domain analysis. To discriminate between deceptive and truth text statement, logistic regression model is proposed by Ball and Elworthy (2014). An accuracy of 71.67% is obtained. However, the composition of features from diverse domains along with a cross-corpus analysis of classification models is required. Shen et al. (2014) took into account image content for automatic news verification in microblogs. For this purpose, a novel visual and statistical image feature is proposed. Experiments show that out of four classifiers, random forest achieved the best accuracy of 83.6%.

Semi-supervised method To detect inaccurate/misleading headlines, a two-step method is proposed by Wei and Wan (2017), which consists of (1) ambiguous headline detection and (2) misleading headline detection. A class sequential rule is proposed for the detection of ambiguous headlines and for misleading headline detection, consistency between the headlines and their associated body texts is measured by evaluating different features. The effectiveness of the proposed model over baseline methods is verified through experiments.

Unsupervised method Tensor modeling-based detection of fake news is proposed by (Hosseinimotlagh and Papalexakis 2018) to automatically identify false information by categorizing it into different types. To refine and consolidate the results of tensor decomposition, the ensemble method is proposed. The proposed method outperforms the existing state-of-the-art methods with an accuracy value of up to 80%. By increasing the number of a top news article to 25 or larger, the number of outliers has also increased.

3.2.5 Machine learning techniques for misinformation

The machine learning techniques for misinformation identification are grouped on the basis of supervised learning method.

Supervised learning method The aim of this work (Kim et al. 2018) is the detection of fake news and misinformation by leveraging the crowd wisdom. An efficient online algorithm, namely CURB, is proposed to detect the spread of misinformation. Experimental results show that the proposed algorithm can effectively reduce the spread of misinformation performed on two real-world datasets. However, further, improvement can be made by designing an adversarial behavior from the part of the crowd. Agrawal et al. (2017) proposed a system for the identification of fake content and misinformation using support vector machine and pairwise ranking algorithms. Results show that the proposed model outperforms the stand-alone classification system. In order to obtain a decision from ranked scores, more classification and ranking schemes should be tested. Misinformation detection in social media during natural disasters is the problem addressed by Rajdev and Lee (2015). Flat and hierarchical classification approaches are proposed for misinformation detection by proposing different features. The result shows that for the identification of misinformation the proposed model outperforms the comparing methods with an accuracy of 96.43% and 0.961 *F* measure. However, the performance of the system for the identification of fake tweet was not efficient.

3.2.6 Machine learning techniques for hoax

The review of studies for hoax detection is grouped on the basis of supervised machine learning method.

Supervised machine learning method To detect hoax in an Indonesian news site, text classification is performed using multidimensional matrix support vector machine and stochastic gradient descent. The accuracy and precision of the support vector machine are increased by 4 to 20% using stochastic gradient descent. To gain a better understanding of hoax context, news has to be considered at the document level (Prasetijo et al. 2017). For the automatic detection of a hoax in social networks, Tacchini et al. (2017) proposed two classification techniques: (1) logistic regression and (2) Boolean crowdsourcing. Results show that the proposed techniques are robust producing an accuracy of 99%. To use the algorithm for the entire social network, there is a need for manual classification. Hoax news detection in the Indonesian language is accomplished using Naïve Bayes classifier to classify news as a hoax and valid. Experiments show the system achieved an accuracy of 78% for hoax detection with a precision of 76% and a recall of 89% (Pratiwi et al. 2017) (Table 2).

3.3 Deep learning techniques for detection and classification of false information

3.3.1 Introduction

Deep learning, also known as deep machine learning, deep structured learning, and hierarchical learning, is a set of algorithms originated by the function and structure of the brain, and it is a subfield of machine learning. In a simple case, there are two sets of neurons, one receives an input signal and other sends the output signals. Deep learning models can be used in a variety of tasks such as speech recognition, computer vision and natural language processing and automatic handwriting generation (Shridhar et al. 2017).

A deep neural network is a good choice in dynamic and complicated social media scenarios. The deep neural network can automatically extract local–global significant features. Therefore, for massive social media information, the deep neural network can be considered as an effective tool for the identification of misinformation (Khan et al. 2019).

Deep learning-based neural network models have shown their effectiveness for efficient detection of false news over the state-of-the-art machine learning methods. This is due to the effective learning capacity of deep neural network models (Ma et al. 2016).

3.3.2 Classification of online content into false and credible information using deep learning

Different deep learning techniques like RNN, LSTM, CNN can be applied to classify the information as false or credible (Girgis et al. 2018). In this section, we present an overview of how a deep leaning technique, namely CNN, can be applied for the classification of online content into credible and false information. It works in two phases, presented as follows: (1) *Feature representations* In the first phase, the target information is encoded into the feature representation vectors (Potdar et al. 2017); (2) *Classification Layers* The representation vectors constructed by the first step are made an input to the classification part. Each deep learning model has its own way of encoding target information into a feature when given an input. The CNN layer captures local dependency between neighbor words, using convolutional filters. The convolutional layer extracts local features and then pooled to a smaller dimension through maxpooling layer to avoid overfitting.

When the input is given to our deep learning model (CNN), the embedding layer generates the word embedding representations. The output of the first layer (embedding representation) is made an input to the CNN layer to generate the feature vector. Finally, the output of CNN layer is passed to the dense layer using a sigmoid activation function for labeling the text as credible or false information (Fig. 3).

3.3.3 Deep learning techniques for rumor detection

Supervised learning For effective rumor detection in microblogs, a novel recurrent neural network with an attention mechanism (att-RNN) is proposed to fuse multimodal features (Jin et al. 2017). Experiments show that with respect to existing feature-based methods, the att-RNN achieves the best performance on both datasets. However, in the attention model, there is no mechanism that explicitly guarantees the relation of images with text/social context. A content representation scheme is proposed by Ma et al. (2017) for rumor detection microblogs using a bag of words and a neural network with a focus on generating text vectors from rumor contents. Experimental results show that the bag-of-words model has achieved 90% accuracy, whereas the neural network model received an accuracy of 60%. However, the performance of the proposed model can be enhanced further by investigating other options and features.

3.3.4 Deep learning techniques for fake news

Supervised leaning method Long et al. (2017) proposed long short-term memory (LSTM)-based classifier for classifying the news as fake and genuine. To form a hybrid model, the speaker information and attention model information are added into the LSTM. Their model achieved an accuracy of 14.5% higher than the state-of-the-art method using a benchmark dataset. However, without a speaker profile, LSTM does not perform better than the baseline methods. In their work on fake news detection, Wu et al. (2018) proposed a novel model, namely Trace Miner, based on LSTM–RNNs and embedding technique. Experimental results show that the proposed model outperforms the state-of-the-art approaches on real-world dataset. The major limitation associated with their approach is that only network information is utilized, ignoring content information. To detect fake news, Singhania et al. (2017) proposed a three-level attention

Table 2 Relevant studies to false information machine learning techniques

Study	Objective	Methods	Technique	Data source and results	Limitation and future work
Ma et al. (2017)	To identify rumors	Kernel method [PTK, cPTK] SVM classifier	Supervised learning framework	Twitter (https://www.statista.com/topics/3251/fake-news) Experiment depicts; for both general and early rumor detection, the proposed system outperforms state-of-the-art baseline	In limitation, rumor detection task can be improved by exploring network learning representation framework. In the future, massive unlabeled rumor data from social media are investigated for unsupervised models
Ghenai and Mejova (2017)	To track health-related rumors on Twitter	Three algorithms are used [1. Naïve Bayes, 2. random forest, and 3. random tree classifier] Supervised method	Supervised method	Twitter Result shows; random tree classifier achieves good precision of 0.946 and recall 0.944	Only a handful of rumors are considered due to the limited resources of this study Future considerations: 1. Multi-lingual nature of data-set is addressed 2. To smooth interaction between a data scientist and communication professionals, a user-friendly interface is developed
Wu et al. (2017)	Early detection of rumor in social media	Cross-Training For Emerging Rumor Detection (CERT) [1. structure learning, feature selection, 3. classifier learning]	Supervised method	Twitter Experiment shows; CERT not only finds emerging rumors faster than existing approaches but also yields effective classifiers	Considers only text data, excluding content such as short videos and images Future directions To further facilitate the detection of rumors, cross-modal information is a requirement
Hamidian and Diab (2015)	Rumor detection and classification (RDC)	Two main experiments [1. Single-step RDC (SRDC), 2. Two-step RDC (TRDC)]	Supervised learning	Twitter Observation shows TRDC attains an F measure of 82.9% compared to 74% in SRDC	Limitation of the method: Preprocessing does not give benefit to the model The proposed methodology is planned to expand to streaming tweets
Yang et al. (2015)	Detect the rumor for social media with hot topic detection	Novel methodology Based on hot topic detection Rumor identification with a various set of features	Supervised approach	News and Twitter Experiment shows; random forest achieves the best result	To improve the result, more sophisticated probabilistic models and more features should be introduced In the future, more vectors will be introduced into multidimensional sentence model

Table 2 (continued)

Study	Objective	Methods	Technique	Data source and results	Limitation and future work
Ahmed et al. (2017)	Online fake news detection	Proposed approach 1. n gram model 2. Data preprocessing 3. Feature extraction 4. Machine learning algorithm (SGD, SVM, LSVM, KNN, DT) Supervised	Supervised technique	1. News article (https://www.cs.ucsb.edu/~william/data/liar_dataset.zip) 2. Hornedataset (https://www.uvic.ca/engineering/ece/isot/datasets/index.php) Experimental results; Accuracy is 92% (TF-IDF) as feature extraction and linear SVM as a classifier	The accuracy of the algorithm decreases with the increase in the *n* gram The model will be tested on a more publically available dataset
Fairbanks et al. (2018)	Address two problems; Bias detection Credibility assessment	The content model for bias detection Structural method for credibility assessment	Supervised learning	News article (global database of events languages and tones) Experiments shows; using TF-IDF matrix logistic regression outperform random forest for both the problems	Limitations: Decay in the performance of bias detection to an accuracy of 0.63 and AUC to 0.51 Future directions: to generalize the work the focus is on developing more robust and novel features
Pérez-Rosas et al. (2017)	Automatic identification of fake news	Proposed model 1. Linguistic feature 2. Readability features 3. SVM classifier 4. Learning curves	Supervised learning	Crowdsourcing covering six news domains Web covering celebrities The model achieved the best performance as compared to the human ability to spot fake news.	There is a significant loss in accuracy in cross-domain analyses as compared to within domain results Further research is required to consolidate the findings
Ball and Elworthy (2014)	Online deceptive text detection	Feature extraction Logistic regression	Supervised approach	Twitter Result depicts; 71.67% of accuracy is obtained.	The preeminence should be on the amalgam of features from different domains and investigation on cross-corpus of classification models Further analysis is required in order to get the benefits of reducing the size of the input feature vector through PCA and evaluation of logistic regression feature parameter
Shen et al. (2014)	Fake followers detection	Proposed a binary classifier to distinguish fake followers from legitimate users	Supervised learning	Sina Weibo Experiment shows 83.6% accuracy is achieved	There is still room to improve the performance of the proposed system

Table 2 (continued)

Study	Objective	Methods	Technique	Data source and results	Limitation and future work
Wei and Wan (2017)	Inaccurate headline detection	Two problem tasks Identifying ambiguous Headline Identifying misleading Headline CO-training approach	Semi-supervised	Chinese news-site The effectiveness of the model is shown over baseline methods through experiments	There is a higher rate of inaccuracy in society and the entertainment news domain To avoid ambiguity in dataset especially in society and entertainment news, more rigorous rule will be investigated
Hosseinimotlagh and Papalexakis (2018)	The fake news identification problem	Proposed method Tier-Spatial relation Extraction CO/PARAFAC decomposition Tensor ensemble co-clustering	Unsupervised learning	Fake news dataset by Kaggle. (http://www.fakenewschallenge.org/) The proposed method outperformed the existing state-of-the-art method	By increasing the number of a top news article to 25 or larger, the number of an outlier is also increasing In the future; more fake news category will be investigated
Kim et al. (2018)	Detection and reduction of fake news and misinformation	Online curb algorithm is proposed	Supervised approach	Twitter and Weibo Experimental results; the proposed algorithm may be able to effectively reduce the spread of misinformation performed on two real-world datasets	The more robust algorithm should be designed to adversarial behavior from the part of the crowd The quadratic loss of the posterior estimate of the misinformation rate is optimized by the algorithm; however, for other losses it is useful to derive fact-checking intensities
Agrawal et al. (2017)	Detection of fake content and misinformation in social media	Features [Image-based features, Twitter content, user-based features, and tweet-based feature Support vector machine (SVM) Scheme (B) Ranking scheme	Supervised learning	Tweets Results show; The proposed model outperforms the stand-alone classification system	In order to obtain a decision from ranked scores, more classification and ranking schemes should be tested For further research, the proposed method will be explored for other multimedia types, e.g., videos to detect fake content or misinformation
Rajdev and Lee (2015)	Misinformation detection in social media	Features [User features, tweet features] Flat classification [FT classifier, NB tree, RF]	Supervised learning	Twitter The proposed model outperforms with an accuracy of 96.43% and 0.961 F measure	The identification of a fake tweet was not good Further research will be extended for the real-time dataset

Table 2 (continued)

Study	Objective	Methods	Technique	Data source and results	Limitation and future work
Prasetijo et al. (2017)	Hoax detection	Methodology Text classification, preprocessing, term weighting method, SVM, SGD	Supervised machine learning	News website The accuracy and precision of SVM are increased by 4–20% using SGD	To gain a better understanding of hoax context, the overall sentences in a news site must be used To overcome many patterns of hoax news, the combination of sentiment analysis, text feature and also voting based on a search engine can be done
Tacchini et al. (2017)	Automatic fake news Hoax detection in social networks	Logistic regression Boolean label crowdsourcing (BLC)	Supervised learning	Facebook Results show; an accuracy exceeding 99% for LR and 99.4% for the harmonic algorithm	In order to use the algorithm for the entire social network, there is a need for manual classification To extend the application in the real-world scenario, the proposed technique may be sufficiently robust
Pratiwiet al. (2017)	Hoax news identification	Naïve Bayes classifier	Supervised learning	New articles Results depict accuracy of 78%	Future research will be to apply dataset on other classification methods

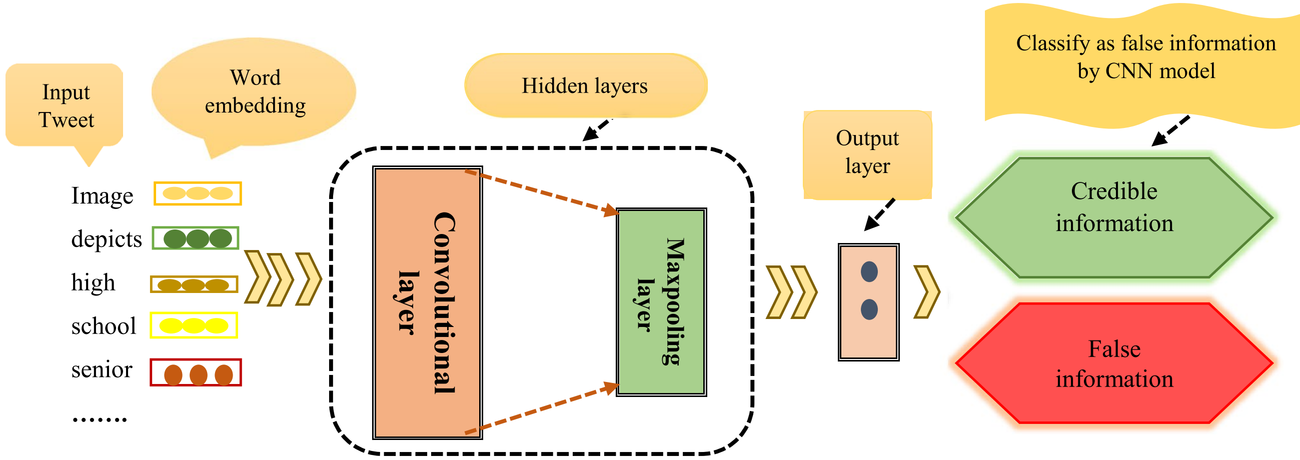


Fig. 3 Deep learning model

network (3HAN) using deep neural network concepts. A three-level news vector (3HAN) is constructed consisting of words, sentences and headlines. An accuracy of 96.77% is obtained on a large real-world dataset. However, a web-based interface is required to detect fake news.

Semi-supervised method Shu et al. (2017) proposed a novel framework called TriFN using tri-relationship for detecting fake news. The effectiveness of the proposed model has demonstrated through experiments, and in the early stage of news dissemination, it achieved good detection performance. As fake news usually evolves very fast so there is a need to explore effective feature for early fake news detection.

3.3.5 Deep learning techniques for misinformation

Semi-supervised learning A novel approach for misinformation identification is proposed by Yu et al. (2017) that is based on the convolutional neural network called CAMI. The effectiveness of the proposed model is demonstrated by performing extensive experiments on two large datasets. However, effective temporal features are not incorporated.

Unsupervised learning A challenge of stance detection in misinformation identification is addressed in Riedel et al. (2017) by proposing a multilayer perceptron (MLP) with one hidden layer. The proposed model exhibits satisfactory performance and correctly classifies instances into related and unrelated. However, the model performance on the stance detection of disagree label is quite poor.

3.3.6 Deep learning techniques for hoax

Supervised learning Automatic classification of suspicious and verified news is carried out by Ruchansky et al. (2017) using linguistically infused neural network model. Experiments show that the proposed model outperforms the logistic regression baselines. By fusing grammar and syntax features, the performance of the model decreases by 0.02 F1 for multiclass, and an accuracy of 0.02 is achieved for the binary classification task. The problem of hoax detection is investigated by (https://www.uvic.ca/engineering/ece/isot/datasets/index.php) CSI model, consisting of three modules: capture, score and integrate. The first module uses RNN that is based on the response and text, second is based on the behavior of the user to learn source characteristics, and third is used to classify an article as fake or otherwise. The experiment shows that the CSI model classifies fake news accurately as compared to the existing systems. With many open challenges, fake news detection still remains a challenging problem despite encouraging results (Table 3).

4 Results and discussion

4.1 Answers to posed research questions

In a systematic literature review, we analyzed 30 studies on false information detection. It is identified that all the articles in this review employed machine learning and deep learning approaches. Moreover, it is observed that for conducting experiments, most of the data sources are chosen from Twitter and Sina Weibo. However, in order to extend the applicability of the models, it is essential to select data from other social networks like YouTube.

In response to RQ1, we have explored four types of false information (rumor, fake news, misinformation and hoax). Although the role of false information in decision making has been investigated in the recent past, false information plays a significant role in manipulating the daily decision making. So, further investigation is required in order to cope

Table 3 Relevant studies to false information deep learning techniques

Study	Objective	Methods	Techniques	Data source and results	Limitation and future work
Jin et al. (2017)	Detect rumors in microblog	Features[text, image, social context] att-RNN., LSTM, deep CNN	Supervised learning method	Twitter and Sina Weibo Experiments show that with existing feature-based methods the att-RNN achieves the best performance on both datasets	In the attention model, there is no mechanism that explicitly guarantees the relation of images with text/social context With additional data, the proposed visual sub-network could be fine-tuned
Ma et al. (2017)	Microblog rumor detection	Two text representations [BOW, TF-IDF] Neural network model [CBOW, Skip-gram] Seven Algorithms [NB, Logic SMO, MLP, Knn, RF, RT]	Supervised learning method	Sina Weibo Experiments show; Classification accuracy for BOW model is 90% as compared to the neural network language model	To promote the performance of rumor detection further study is required Future work; how to integrate the context information is left
Long et al. (2017)	Fake news detection	Hybrid model LSTM Attention information Speaker profile	Supervised approach	Liar Dataset [News] The proposed model outperforms the state-of-the-art model with an accuracy of 14.5%	Without speaker, profile LSTM does not perform better than baseline Further, the model performance will be evaluated with more dataset
Wu and Liu (2018)	Fake news detection Standard news classification	Propose model Embedding method LSTM-RNN	Supervised learning	Twitter news data Experiment shows; The proposed model outperforms the state-of-the-art model	Only network information is utilized. There is a loss of dependency between users For future direction, it will be investigated whether the proposed model facilitates other network mining task, i.e., link prediction and recommendation
Singhania et al. (2017)	Fake news detection	Model design Sequence encoder using GRU Supervised pre-training using the headline[3HAN-PT]	Supervised method	News article The effectiveness of the model with an accuracy of 96.77% is observed in the large real-world dataset	To detect fake news, a web application based on 3HAN should be established To enable further manual fact-checking, the output can be visualized through a heat map
Shu et al. (2017)	Fake news detection	Novel framework TriFN to model tri-relationship is proposed	Semi-supervised learning	BuzzFeed (https://github.com/BuzzFeedNews/2016-10-facebook-fact-check/tree/master/data) Polifact (https://github.com/bs-detector/bs-detector) The proposed model achieved good detection performance	There is a need to explore effective feature for early fake news detection Future directions; the malicious or low-quality user is important

Table 3 (continued)

Study	Objective	Methods	Techniques	Data source and results	Limitation and future work
Yu et al. (2017)	Misinformation identification Early detection	Proposed CAMI Model Group correlative microblog post Each group learning representation via paragraph vector High-level modeling by CNN(K max pooling)	Unsupervised approach	Twitter and Weibo Effectiveness; the proposed model; outperforming both in misinformation and early detection task	Effective temporal features are not incorporated To better achieve the task of misinformation and early detection inherent properties of misinformation in social media will be demonstrated
Riedel et al. (2017)	Misinformation fake news stance detection	Proposed model. 1. Features (BOW, TF, TF-IDF) 2. Classifier (MLP)	Supervised learning	News (Headline and body) The proposed model performs satisfactorily and perfectly classify instances into related and unrelated	The model performance on the stance detection of disagree label is quite poor The future goal is the in-depth analysis of the model
Ruchansky et al. (2017)	To classify fake news accurately To identify a group of suspicious users	CSI Proposed model 3 modules (Capture, Source, Integrate)	Supervised learning method	Twitter and Sina Weibo Result depicts; CSI model classifies fake news accurately than the existing system	The available labeled examples of fake and true news may be limited To get more accurate and timely predictions, human in the learning process could be included

with the issue. However, certain limitations of the previous work are: (1) the rhetorical structure theory (RST) for news discourse analysis is performed manually, which is time-consuming (Rubin et al. 2015), (2) accuracy of the proposed logistic regression model is 56%; however, it is not significantly better than the Chi-square test, i.e., p value = 0.854 and 1(df) = 0.0339 (Rubin et al. 2015); (3)there is a significant increase in the outlier when the top news article number reaches to 25 or become larger than this number (Sharma et al. 2019); (4) annotation and categorization of tweets are performed manually, which is time-consuming (Wu and Liu 2018). The solutions to the aforementioned shortcomings are: (1) By automating the RST framework, its robustness can be improved. (2) Dimensionality reduction techniques can be applied for improving the accuracy. (3) The outliers can be reduced by using z-score, proximity-based models and probabilistic modeling. (4) For automatic annotation and categorization of tweets, a classifier could be used.

We have identified different machine learning techniques for false information detection and classification in response to RQ2. The existing machine learning techniques have several deficiencies: (1) The query formalization is performed manually with limited coverage of medical-related terms (Ghenai and Mejova 2017); (2) improvement in performance with respect to rumor identification model is essentially required (Hamidian and Diab 2015); (3) the early rumor detection model only uses the text content of social media (Wu et al. 2017); (4) there is a decay in the performance of bias detection to an accuracy of 0.63 and AUC to 0.51 (Granik and Mesyura 2017); (5) existing datasets are limited for achieving improved performance with respect to detection of fake news (Granik and Mesyura 2017); (6) fake followers affect the credibility of microblog system so it is necessary to keep track of the evolution of such followers (Shen et al. 2014); and (7) the accuracy of the machine learning algorithm decreases with the increase in n gram features, such as tri-gram and four-gram (Ahmed et al. 2017). To overcome these deficiencies, certain solutions have been proposed: (1) there is a need to add extended set of medical terms synonym; (2) for achieving improved results more classification methods along with an additional set of features should be investigated; (3) to further facilitate the rumor detection model, other social media content such as short videos and images could be included;(4) more robust and novel features should be developed to avoid performance decay; (5) to progress in the fake news detection, the researchers can collaborate in building a shared dataset; (6) fake followers could be tracked by constantly adding new features; and (7) as deep learning automatically extract latent features, it can be used to maintain the accuracy of algorithm.

In response to RQ3, we have identified different deep learning techniques and highlighted their importance for false information detection. The identified challenges are: (1) the stance detection model accuracy for the disagree label is quite poor (Long et al. 2017); (2) only social network information is utilized for fake news detection (Wu and Liu 2018); (3) to train the deep neural network the number of manual fact check articles is too less (Singhania et al. 2017); and (4) fake news detection is still an open challenge despite encouraging results (Allcott and Gentzkow 2017). The solutions to these challenges are: (1) the accuracy of model on the disagree label can be improved by increasing the number of instances; (2) the content and social network information should be used together because the content information is easily available; (3) a web application based on 3HAN should be developed that manually check new articles in real time; (4) to cope with the challenge and to get accurate and timely predictions, a model that integrates both crowdsourcing and reinforcement learning concepts can be investigated.

4.2 Comparison of false information detection techniques

For research work and practical applications, there is a need to select the best-performing approach. However, due to some factors, the direct comparison between those systems is difficult. Firstly, the dataset used by the original authors is different due to which the comparison between the reported and implemented results is not fair. More importantly, the reported results are not reproducible, some methods have shown excellent results and some have much lower performance than the reported one, because the original author describes their system with varying degrees of detail and accuracy. Ahmed et al. (Ahmed et al. 2017) investigated two feature extraction techniques, namely (1) TF-IDF and (2) TF. There results depict that TF-IDF outperformed the other methods and achieved the best results than the TF. In order to improve the accuracy, Ball et al. (Ball and Elworthy 2014) reduced the dimensionality of features using the feature extraction technique, namely principal component analysis (PCA). As compared to other studies, the results depict better performance. Furthermore, Jin et al. (2017) explored a novel visual feature for microblog news verification such as clarity score, coherence score and diversity score. These features significantly improve the verification accuracy and performance. Similar to Ahmed et al.'s (2017) work, Prasetijo et al. (2017) also incorporated feature representation technique known as TF-IDF technique, and the results depict that the accuracy of the proposed model increases.

We applied a dataset, namely News_dataset (https://www.uvic.ca/engineering/ece/isot/datasets/index.php), on the methods reported in the papers. In our implementation, we did our best to follow the description in the respective paper as exactly as possible. However, in some cases, the original paper gave too little clue or lack of explanation, so we had to guess what the authors meant. In our experiments, we used Anaconda-based Python environment as the original author often did not specify the tools they have used. With this, we conducted a quantitative comparison among the methods on a consistent dataset.

4.2.1 Quantitative comparison on the same dataset

To evaluate the performance of the existing approaches on a consistent dataset, we have used a public dataset a news domain dataset which contains 10,000 news labeled as fake and 10,000 labeled as real. For the results of our evaluation, the selected approaches are shown in Table 4, showing both the reported accuracy in the respected papers and our implemented accuracy. The difference in the accuracies is due to the lack of detail in original publications, which did not allow for exact reproduction of the techniques in our implementation.

As we have used different experiment setting, tools and data, in some cases, the reported results are not comparable with our results. For example, Ahmed et al. (2017) reported 92% accuracy, but we obtained 74.37%. Similarly, Long et al. (2017) reported 41.5% accuracy and we obtained 94%.

5 Trends in false information

In this section, the following trends are identified relevant to the false information.

5.1 Year-wise article publication

In this trend, we identify the percentage of an article published based on year. The percentage of articles which can be observed in a pie chart together with the year they were published is shown in Fig. 4. We observed that the majority of the articles are on the fake news, moderate on rumor and misinformation, and very little on hoax (https://www.businessinsider.com/fake-news-starbucks-free-coffee-to-undocumented-immigrants).

5.2 Methods for classification

In this trend, we distinguish the research work into machine learning and deep learning perspective.

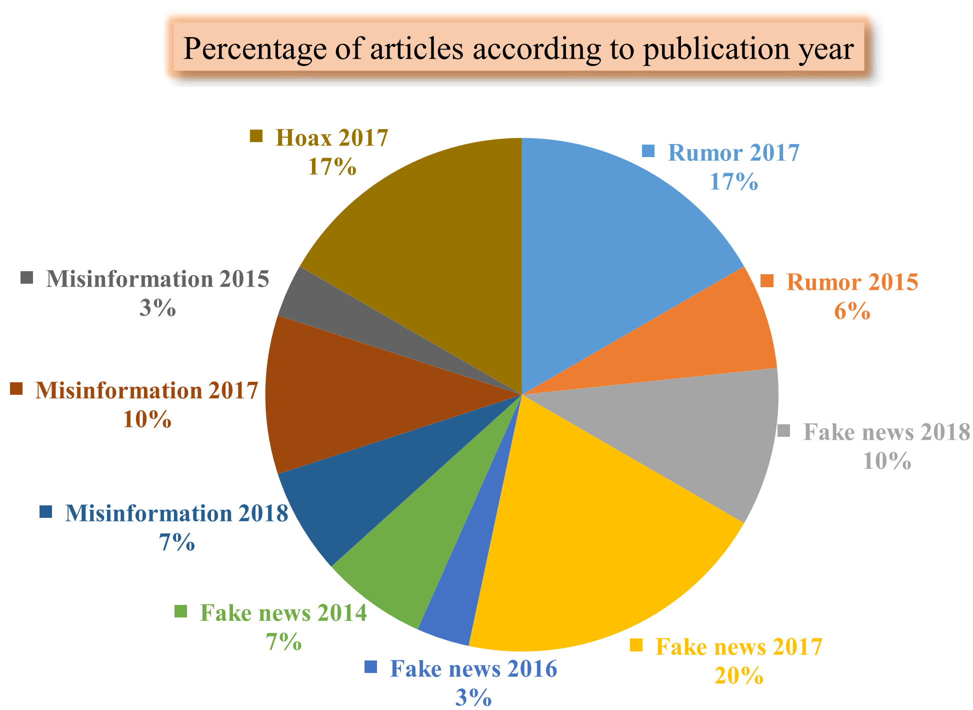
Figure 5 depicts that mostly the work has been done in machine learning using SVM (30%), Naïve Bayes (18%), random forest (18%), random tree classifier (4%), logistic regression (15%), stochastic gradient descent (7%), *k*-nearest neighbor (4%), decision tree (4%) and deep learning using RNN (30%), LSTM (30%), convolutional neural network (20%), MLP (20%) (Ma et al. 2017).

Table 4 Quantitative comparison of false information detection approaches

Paper	Approach	Machine learning techniques/deep learning techniques	Reported accuracy (%)	Accuracy in our dataset (%)
Ahmed et al. (2017)	Features extraction techniques like term frequency-inverse document frequency (TF-IDF) and TF	SGD, SVM, LSVM, KNN and DT	92	74.37
Yu et al. (2017)	Convolutional approach	DT-Rank, SVM-RBF, DTC, RFC, SVM-TSGRU-2, CAMI	77.7	74
Ball and Elworthy (2014)	Feature extraction such as principal component analysis	LR	71.67	66.3
Jin et al. (2017)	Explore novel visual and statistical image features such as clarity score, coherence score, and diversity score	SVM, LR, Kstar, RF	83.6	78.7
Prasetijo et al. (2017)	Supervised learning technique. Used TF-IDF technique for feature representation	SVM, SGD	86	84.57
Jin et al. (2017)	Multimodal fusion	attn.-RNN	68.2	75
Long et al. (2017)	Hybrid attention-based LSTM model	att-LSTM	41.5	94
Wu and Liu. (2018)	TraceMiner	SVM, XGBoost, TM (Deepwalk) TM (LINE),TM (LSTM-RNN)	91.24	86
Singhania et al. (2017)	Three-level hierarchical attention network (3HAN)	SVM, GRU, Attn.-BiGRU	96.77	**95**

Bold value means best accuracy

Fig. 4 The percentage of articles (pie chart) with respect to four dimensions: rumor, fake News, misinformation, and hoax



6 Open issues and future directions

In this section, we examine the open research problems and future directions.

Lack of work in rumors, misinformation, and hoax In recent past, research in false information has focused largely on the first components of the pipeline, namely fake news. The latter has not been addressed significantly (Wu et al. 2017). So we

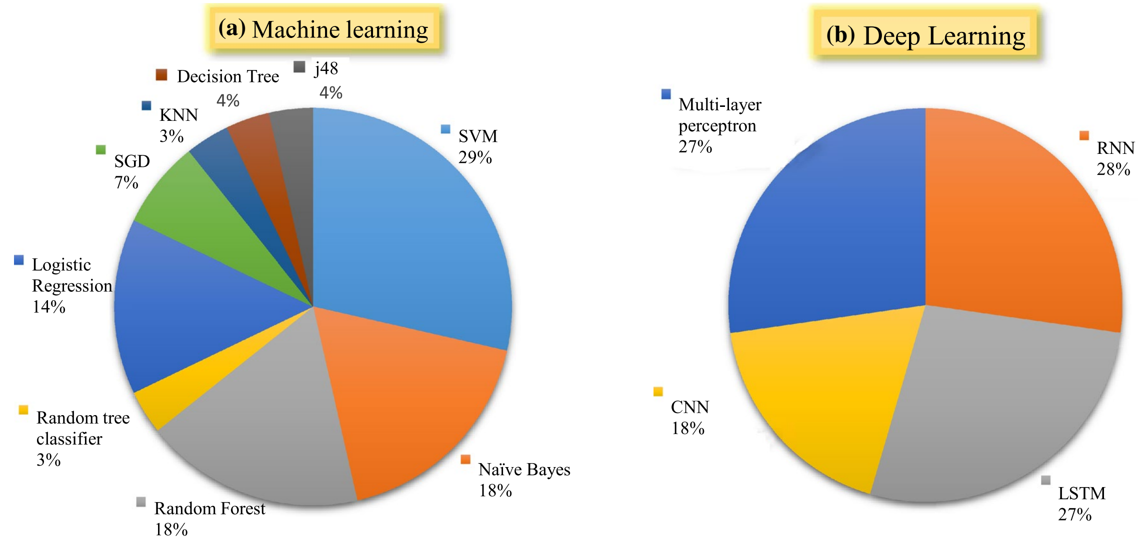


Fig. 5 Percentage of article using **a** machine learning technique and **b** deep learning technique

argue that in order to deal with false information detection, rumor, misinformation, and hoax should be the future research directions (Habib et al. 2019).

The dearth of the shared dataset One of the limitations is the lack of publically available data toward the development of false information detection (Ma et al. 2017). So in order to promote further research over the different dataset, we encourage researchers to collaborate in building a shared dataset.

The want for early false information detection One of the provocative challenges is the early false information detection whose objective is to provide early alerts of false information during the diffusion process. As early detection of false information will hinder the further propagation of false information on different social networks (Jin et al. 2017), emphasizing on early false detection is another research area of interest.

Should emphasize on deep learning In going through existing literature, we found that most of the previous work attempted to detect false information through machine learning algorithms in which features are extracted manually, which is a labor-intensive task and takes time and the alternative is to use deep learning. As deep neural models (Khan et al. 2019) can capture the hidden representations effectively, we can achieve better performance as compared to the machine learning algorithms (Ma et al. 2016). Therefore, we encourage researchers to focus in this direction.

7 Conclusion

False information dissemination is faster than the authentic information on the Internet (Study: On Twitter, False News Travels Faster than True Stories. MIT News, http://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308). The easy access, low cost and large-scale application of social media provide an ample place for false information diffusion. Therefore, false information detection is an emerging issue. In this survey, we explored an emanate problem of false information and its role in decision making and also discussed different types of false information, namely (1) rumor, (2) fake news, (3) misinformation and (4) hoax by reviewing existing articles. We find different machine learning and deep learning techniques for false information detection. On the basis of limitations mentioned in Sect. 4, we propose the following future directions: (1) further extension in false information types is essential to assist people to make rational decisions, (2) to lubricate the false information detection the unlabeled data should be investigated for unsupervised machine learning models as labeling of data is labor-intensive, (3) to tackle with false information problem, further exploration of deep learning models is favorable, (4) in order to flourish the research in the field, a mixed dataset that covers different domains like health, politics and education should be developed, and finally, (4) hoax detection is the least address area in false information detection, which requires further attention.

Author's contribution AH and MZA conceived and designed the experiments; AH, MZA and AK performed the experiments; and AH, AH, MZA, AK and AK analyzed the data.

Availability of data and material This work deals with the review of relevant studies, and the required data are present inside the article.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

References

Agrawal T, Gupta R, Narayanan S (2017) Multimodal detection of fake social media use through a fusion of classification and pairwise ranking systems. In: 2017 25th European conference on signal processing (EUSIPCO), IEEE, pp 1045–1049

Ahmed H, Traore I, Saad S (2017) Detection of online fake news using N-gram analysis and machine learning techniques. In: International conference on intelligent, secure, and dependable systems in distributed and cloud environments, Springer, Cham, pp 127–138

Allcott H, Gentzkow M (2017) Social media and fake news in the 2016 election. J Econ Perspect 31(2):211–236

Asghar MZ, Rahman F, Kundi FM, Ahmad S (2019) Development of stock market trend prediction system using multiple regression. Comput Math Organ Theor 25(3):1–31

Ball L, Elworthy J (2014) Fake or real? The computational detection of online deceptive text. J Market Anal 2:187. https://doi.org/10.1057/jma.2014.15

Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. J Comput Sci 2(1):1–8

Chen H, Sun M, Tu C, Lin Y, Liu Z (2016) Neural sentiment classification with user and product attention. EMNLP

Dalal DK, Diab DL, Tindale RS (2015) I heard that…: Do rumors affect hiring decisions? Int J Sel Assess 23(3):224–236

Fairbanks J, Fitch N, Knauf N, Briscoe E (2018) Credibility assessment in the news: do we need to read. In: Proceddings of the MIS2 held in conjuction with 11th international conference on web search and data mining, pp 799–800

Fallis D (2009) A conceptual analysis of disinformation. In: Proceedings of in conference

Fisher M, Cox JW, Hermann P (2016) Pizzagate: from rumor, to hashtag, to gunfire in DC, Washington Post

Ghenai A, Mejova Y (2017) Catching Zika fever: application of crowdsourcing and machine learning for tracking health misinformation on Twitter. arXiv preprint arXiv:1707.03778

Giasemidis G, Singleton C, Agrafiotis I, Nurse JR, Pilgrim A, Willis C, Greetham DV (2016) Determining the veracity of rumors on Twitter. In: International conference on social informatics, Springer, Cham, pp 185–205

Girgis S, Amer E, Gadallah M (2018) Deep learning algorithms for detecting fake news in online text. In: 2018 13th international conference on computer engineering and systems (ICCES), IEEE, pp 93–97

Granik M, Mesyura V (2017) Fake news detection using naive Bayes classifier. In: 2017 IEEE first Ukraine conference on electrical and computer engineering (UKRCON), IEEE, pp 900–903

Gupta A, Lamba H, Kumaraguru P, Joshi A (2013) Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. In: Proceedings of the 22nd international conference on World Wide Web, ACM

Habib A, Akbar S, Asghar MZ, Khattak AM, Ali R, Batool U (2019) Rumor detection in business reviews using supervised machine learning. In: 2018 5th international conference on behavioral, economic, and socio-cultural computing (BESC), IEEE, pp. 233–237

Hamidian S, Diab M (2015) Rumor detection and classification for twitter data. In: The fifth international conference on social media technologies, communication, and informatics, SOTICS, IARIA, pp 71–77

Hosseinimotlagh S, Papalexakis EE (2018) Unsupervised content-based identification of fake news articles with tensor decomposition ensembles. MIS2, Marina Del Rey

Howard PN, Kollanyi B (2016) Bots, #strongerin, and #brexit: computational propaganda during the uk-eu referendum. SSRN Electron J. https://doi.org/10.2139/ssrn.2798311

Hunt E (2016) What is fake news? How to spot it and what you can do to stop it. Guardian. https://www.theguardian.com/media/2016/dec/18/what-is-fake-news-pizzagate. Accessed 21 Mar 2019

Jin Z, Cao J, Zhang Y, Zhou J, Tian Q (2017a) Novel visual and statistical image features for microblogs news verification. IEEE Trans Multimed 19(3):598–608

Jin Z, Cao J, Guo H, Zhang Y, Luo J (2017) Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In: Proceedings of the 2017 ACM on multimedia conference, ACM, pp 795–816

Khan A, Younis U, Kundi AS, Asghar MZ, Ullah I, Aslam N, Ahmed I (2019) Sentiment classification of user reviews using supervised learning techniques with comparative opinion mining perspective. In: Science and information conference, Springer, Cham, pp 23–29

Khan A, Feng J, Liu S, Asghar MZ (2019b) Optimal skipping rates: training agents with fine-grained control using deep reinforcement learning. J Robot. https://doi.org/10.1155/2019/2970408

Kim J, Tabibian B, Oh A, Schölkopf B, Gomez-Rodriguez M (2018) Leveraging the crowd to detect and reduce the spread of fake news and misinformation. In: Proceedings of the eleventh ACM international conference on web search and data mining, ACM, pp 324–332

Liang G, He W, Xu C, Chen L, Zeng J (2015) Rumor identification in microblogging systems based on users' behavior. IEEE Trans Comput Soc Syst 2(3):99–108

Lipton ZC (2018) Is fake news a machine learning problem? http://approximatelycorrect.com/2017/01/23/is-fake-news-a-machine-learning-problem/. Accessed 8 Oct 2018

Long Y, Lu Q, Xiang R, Li M, Huang CR (2017) Fake news detection through multi-perspective speaker profiles. In: Proceedings of the eighth international joint conference on natural language processing, vol 2: Short Papers, vol 2, pp 252–256

Ma J, Gao W, Mitra P, Kwon S, Jansen BJ, Wong KF, Cha M (2016) Detecting rumors from microblogs with recurrent neural networks. In: IJCAI, pp 3818–3824

Ma J, Gao W, Wong KF (2017) Detect rumors in microblog posts using propagation structure via kernel learning. In: Proceedings of the 55th annual meeting of the association for computational linguistics, vol 1: long papers, vol 1, pp 708–717

Ma B, Lin D, Cao D (2017) Content representation for microblog rumor detection. In: Advances in computational intelligence systems. Springer, New York, pp 245–251

Oh O, Agrawal M, Rao HR (2013) Community intelligence and social media services: a rumor theoretic analysis of tweets during social crises. Mis Q 37(2):407–426

Pérez-Rosas V, Kleinberg B, Lefevre A, Mihalcea R (2017) Automatic detection of fake news. arXiv preprint arXiv:1708.07104

Pomerantsev P, Weiss M (2014) The menace of unreality: how the Kremlin weaponizes information, culture and money. Institute of Modern Russia New York

Potdar K, Pardawala TS, Pai CD (2017) A comparative study of categorical variable encoding techniques for neural network classifiers. Int J Comput Appl 175(4):7–9

Prasetijo AB, Isnanto RR, Eridani D, Soetrisno YAA, Arfan M, Sofwan A (2017) Hoax detection system on Indonesian news sites based on text classification using SVM and SGD. In: 2017 4th international conference on information technology, computer, and electrical engineering (ICITACEE), IEEE, pp 45–49

Pratiwi IYR, Asmara RA, Rahutomo F (2017) Study of hoax news detection using naïve Bayes classifier in Indonesian language. In: 2017 11th international conference on information and communication technology and system (ICTS), IEEE, pp 73–78
Rajdev M, Lee K (2015) Fake and spam messages: detecting misinformation during natural disasters on social media. In: 2015 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology (WI-IAT), vol 1, IEEE, pp 17–20
Riedel B, Augenstein I, Spithourakis GP, Riedel S (2017) A simple but tough-to-beat baseline for the Fake News Challenge stance detection task. arXiv preprint arXiv:1707.03264
Rubin VL, Conroy NJ, Chen Y (2015) Towards news verification: deception detection methods for news discourse. In: Proceedings of the Hawaii international conference on system sciences (HICSS48) symposium on rapid screening technologies, deception detection and credibility assessment symposium, pp 5–8
Ruchansky N, Seo S, Liu Y (2017) CSI: a hybrid deep model for fake news detection. In: Proceedings of the 2017 ACM on conference on information and knowledge management, ACM, pp 797–806
Serrano E, Iglesias CA, Garijo M (2015) A survey of Twitter rumor spreading simulations. In: Computational collective intelligence, Springer, Cham, pp 113–122
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):21
Shen Y, Yu J, Dong K, Nan K (2014) Automatic fake followers detection in Chinese micro-blogging system. In: Tseng VS, Ho TB, Zhou ZH, Chen ALP, Kao HY (eds) Advances in knowledge discovery and data mining, PAKDD 2014. Lecture notes in computer science, vol 8444. Springer, Cham
Shridhar K (2017) A beginners guide to deep learning—Kumar Shridhar—medium. Medium. https://medium.com/@shridhar743/a-beginners-guide-to-deep-learning-5ee814cf7706. 26 May 2017
Shu K, Sliva A, Wang S, Tang J, Liu H (2017a) Fake news detection on social media: a data mining perspective. ACM SIGKDD Explor Newsl 19(1):22–36
Shu K, Wang S, Liu H (2017) Exploiting tri-relationship for fake news detection. arXiv preprint arXiv:1712.07709
Singhania S, Fernandez N, Rao S (2017) 3HAN: a deep neural network for fake news detection. In: Liu D, Xie S, Li Y, Zhao D, El-Alfy ES (eds) Neural information processing ICONIP 2017. Lecture notes in computer science, vol 10635, Springer, Cham
Stantic B (2017) Provenance-based rumor detection. In: 28th Australasian database conference on databases theory and applications: ADC 2017, Brisbane, QLD, Australia, vol 10538, Springer, Cham, p 125
Tacchini E, Ballarin G, Della Vedova ML, Moret S, de Alfaro L (2017) Some like it hoax: automated fake news detection in social networks. arXiv preprint arXiv:1704.07506
Wei W, Wan X (2017) Learning to identify ambiguous and misleading news headlines. arXiv preprint arXiv:1705.06031
Wu L, Liu H (2018) Tracing fake-news footprints: characterizing social media messages by how they propagate. In: Proceedings of the eleventh ACM international conference on web search and data mining. ACM, pp 637–645
Wu K, Yang S, Zhu KQ (2015) False rumors detection on Sina Weibo by propagation structures. In: IEEE international conference on data engineering, ICDE
Wu L, Li J, Hu X, Liu H (2017) Gleaning wisdom from the past: early detection of emerging rumors in social media. In: Proceedings of the 2017 SIAM international conference on data mining. Society for Industrial and Applied Mathematics, pp 99–107
Yang Z, Wang C, Zhang F, Zhang Y, Zhang H (2015) Emerging rumor identification for social media with hot topic detection. In: Web information system and application conference (WISA), IEEE, pp 53–58
Yu F, Liu Q, Wu S, Wang L, Tan T (2017) A convolutional approach for misinformation identification. In: Proceedings of the twenty-sixth international joint conference on artificial intelligence (IJCAI-17), pp 3901–3907
Zhou X, Zafarani R (2018) Fake news: a survey of research, detection methods, and opportunities. arXiv preprint arXiv:1812.00315
Zubiaga A, Aker A, Bontcheva K, Liakata M, Procter R (2018) Detection and resolution of rumours in social media: a survey. ACM Comput Surv (CSUR) 51(2):32