Chapter 2: Literature Review and Background of Rumour Detection

2.1 Origin of Rumour

Rumours are predominant in our society and it affects our opinions and conducts towards others and usually affects the approach we see in the world. The relative study of rumour began during the Second World War during this period Knapp collects thousand rumours over 1942 period and classified these rumour by systematic study for different purpose and contents of rumour (Knapp, 1944). This study is the foundation for relevant research in the field of rumour detection. The 1960 and 1970 saw another series of concern with numerous prominent publications (Shibutani,1966).

For the study of rumour propagation the first approach that was applied on it was based on mathematical model. The first model applied on it was the D-K model. In assessment of this Maki D and Thomson also applied mathematical models to carry out research on propagation (Kerner, 1968). With the fast growth of mobile, social networks researchers proposed that already created equations are no longer suitable to describe the rumour spreading. So based on diffusion equations a discrete delay model is designed (Zhu et al., 2016).

2.2 Effects of Rumour

Studies within rumour influence focused on the way through which it affects individuals, their opinions and numerous phases in their lives. In an administrative context rumour are problematical for a firm's, parties reputation when it spread negative information.

Rumour related with stock markets such as business announcements, earning expectations, undervalued stocks can affect in major share price variations.

Subsequence paragraph showing some example of rumour that causes indemnities among politics, economy as well as community constancy.

- Through the 2016 U.S. constitutional vote applicants along with followers were enthusiastically used Facebook and Twitter for promotions and express their sentiments. Throughout the event 529 various stories were spread among the media related to the candidates Donald Trump and Hillary Clinton. The stories reached thousands of followers through social media network and swayed the election (Zin et al., 2017).
- In the year 2013 false rumour spread about attack on White House and also about president got wounded. This rumour resolved in the early stages but still it caused numerous social anxieties. The result is stock market decrease instantaneously (Domm, 2013).
- During the period 2014 after the "Malaysia Airlines Flight MH370 lost Contact" 92 various rumours were disseminated among online platform Sina Weibo. It hurt the people emotions and blocks them to understand the actual situation (Jin, 2014).

2.3 Types of Rumour

Several factors are available for categorizing rumour by its type, its value (false, unresolved or true) or related to its degree whether low level or high level. Basically rumours are identified into three parts. Table 2.1define the types of rumour according to knapp.

Table 2.1 Types of rumours (Laverty, 1974)

Wedge-driving Rumour	Wedge-driving rumours are those that disseminate hatred among communities (Aricat, 2018).Wedge-driving rumour derogates other groups.
Pipe dream	Pipe dreams are also known as wishing fulfilment rumour (those reflect wished for outcomes) (Lalehparvaran, 2018).
Bogy Rumour	Bogy rumour spread the fear among the society (Tai and Sun, 2011).

Rumour occurs from various platforms, some rumours are trendy and some are discussed from long period of time among the society. The most useful factors which are used for rumour classification are given below:

2.4 Rumour arising from breaking news

These types of rumour are never observed before to deal with such types of rumour an automatically detection system is required. In such scenario whenever initial recognition and determination is critical posts are essentials to be handled in real phase. Example of such rumour is when during the breaking news an identity of criminal is tested. A classification is used which observed the same cases of criminal but here the names and cases are dissimilar from ever before. In these cases the classifier which is designed contains the appearance of new cases with new terminology ((Zubiaga, 2016).

2.5 Long standing Rumour

Dispersion of rumour for long period of time without its verification known as long standing rumour. Remarkable evidences are used by the association to categorize the event related to such tales (Aker, 2017).

2.6 Paradigms for Automatic Rumour Detection

Numerous labours have been used to tackle online rumours spontaneously by mining the various dataset with various machine learning methods. Rumour recognition approaches are classified among hand-crafted, propagation and neural network based methods.

2.7 Detection Methods for Rumour Detection

With the enhancement of social sites people transfer their contents easily with each other. Despite numerous benefits of social media it also adversely affects the society. Numerous misinformation, rumours disperse among the society that affect the people opinion and reliability of the data transferred among the media. Researchers from various fields shown their concern to discover the rumour at initial stages so that its harmful affect can be decreased. Currently various machine and deep learning methods came into existence for rumour classification problem. The figure 2.1 defines various machine and deep learning methods to detect the rumour.

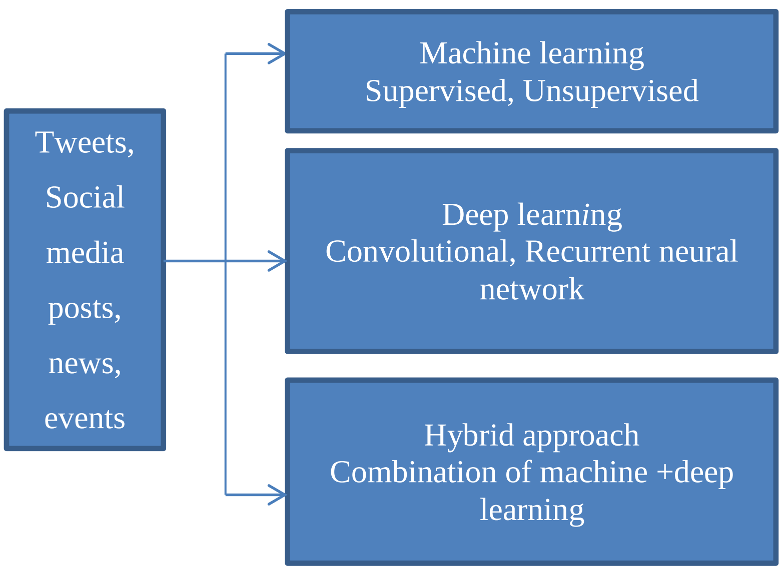

Fig. 2.1 Methods of rumour detection (Bondielli et al., 2019)

2.7.1 Machine Learning Rumour Detection Techniques

Machine learning method is a procedure to train machines to test the data more proficiently and acquire result more accurately. Machine learning algorithms understand the data pattern and extract material from the accessible dataset. The main concern of these algorithms to enhance the computer programs that access the information and then use the same information for their learning process.

These algorithms learn from previous data to enhance the future results. The main concern of machine learning is to develop automatic learning techniques without any intervention of human being. In machine learning the algorithm provide its own solution based on the training dataset given to it in the initial stage.

(Castillo et al., 2011) developed method identify the rumour from content, user and social features with various classifiers. J48 classifier provides 89% accuracy on handcrafted feature. Twitter monitor is used to detect the events.

(Qazvinian et al., 2011) proposed a framework to detect long standing rumour. This framework not detect any rumour from breaking news where the prior knowledge not known.

(Alsaeedi and Al-Sarem, 2020) proposed J48 classifier to train the model on WEKA platform. The work provides 82% f-score. The efficiency of preprocessing work is not good because the limitation of WEKA tool.

(Zhao et al., 2015) create enquiry phrase-based method to detect the rumour. The proposed method assemble similar phrase in one cluster. The work provides 52% precision score. The main limitation of the work is slow response and expressions are created manually.

(Zubiaga et al., 2017) developed conditional random fields (CRF) for rumour detection task on content and social features. PHEME dataset is used for the proposed work and the work achieves 60.7% f1-score.

(Vijeev et al., 2018) apply NLP techniques for rumour detection purpose on user and content features. Chi-square method is used to rank the best features. Various classifiers such as support vector machine, Naive Bayes, Random forest applied on the PHEME dataset. The greatest accuracy provided through the RF classifier was 74.6%.

(Sivasangari et al., 2018) discussed the death case of Jayalalithaa former Chief Minister of Tamil Nadu in the year 2016. Dataset is collected with the help of hashtag from Twitter. (Endsley et al., 2014) proposed work measure user credibility in social media sites. Users with similar behavior are identified and then grouped into a cluster. Weights are assigned to each cluster depends upon its size. US senate voting history dataset is used.

(AlMansour et al., 2014) proposed work analyzes the credibility of content, author and reader. Arabic contents are considered for experimental purpose.

(Chen et al., 2018) identifies rumour from Weibo based on user behavior by using unsupervised learning model. Seven time slots are used in the work. (Pradeepa et al., 2018) proposed work identifies rumour based on user id of twitter user. Dataset about follower's id is collected and then clustering is performed to identify which cluster is the main cause of rumour dispersion and how it spread.

(Olivieri et al., 2019) proposed work select news statement from Google search engine via its API to detect fake news. The work is dependence on Google's Custom-Search API. If the Google search engine changes its schema the proposed technique is not efficient for data extraction and results.

(Khan et al., 2016) analyzed news events from Twitter and then check its credibility by applying SVM ranking algorithm. The main limitation of proposed work is they need human annotator to obtain ground truth of each event. The model works on predefined rumour only.

(Qazvinian et al., 2011) identify the rumour in the microblog and check the effectiveness of rumour identification in three categories: content-based, network based and microblog specific.

(Ahmed et al., 2018) proposed a classifier that contains set of predefined rumours. This method is beneficial for long standing rumours and it is not appropriate for breaking news.Various machine learning techniques are defined in Table 2.2.

Table 2.2 Machine learning techniques (Bondielli et al., 2019)

Platform	Algorithm/ Technique	Features	Results	Drawbacks
Twitter (Castillo et al., 2011)	SVM, J48, Naive Bayes	Content, user and social features	89% accuracy with J48	Feature extraction is time consuming, labeling done by humans
Twitter (Qazvinian, et al., 2011)	Naive bayes	Content features	Content and twitter based features provides highest recall and precision	Human needs for labeling
Twitter (Hamidian et al., 2015)	J48, DT	Content, network, Twitter specific features	WEKA tool is used for preprocessing	Work is restricted for implementation only on WEKA
Twitter (Zhao et al., 2015)	Clustering method	Content- features	Proposed work detect rumour at early stages	Response time is low
PHEME Vijeev et al., 2018)	RF, SVM, NB	Content, user features	74.6% accuracy	Few features are used

2.7.2 Deep Learning Methods

Deep learning methods are a part of machine learning and these techniques are developed by the motivation of human mind functionality. Deep learning is widely used investigation topic nowadays because it provides lot of prominent result in many fields including natural language processing and mining of text. As compared with traditional methods it provides lot of advantages. Traditional methods depend upon manually examination, feature extraction task is time overwhelming and many results are biased result. DL mine hidden characteristics from text and images. The two most usually applied standards including RNN and CNN.

The RNN model linked nodes chronologically with one other and create a graph. The RNN networks mainly efficient for forming consecutive records, like human language and getting appropriate characteristics from various means of information (Hochreiter et al., 1997) (Ruchansky et al., 2017).
(Ma et al., 2018) developed a RNN based model named tanh-RNN, GRU and LSTM. The beginning post provides as an input for the model and every node of the tree might be reactive post.

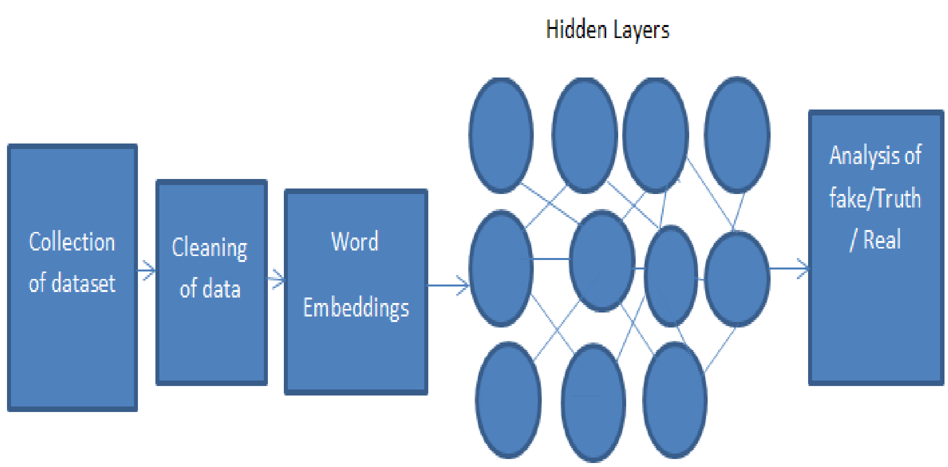

Fig. 2.2 Deep learning architecture (Asghar et al., 2019)

(Yu et al., 2017) deliberated the drawbacks of RNN developed by (Ma et al., 2018), they developed another technique based upon CNN. The work provides 93.3% accuracy on Sina Weibo dataset. (Nguyen et al., 2017) combined CNN and LSTM techniques. Firstly CNN model is used to extract the expression. In next phase LSTM is used for tweet representation. Snopes and Urban Legends data is used for experimental purpose. The work achieved 81.9% accuracy on training dataset. (Alkhodair et al., 2010) identify the rumours from breaking news on Twitter platform. The proposed model combines Word2Vec and LSTM-RNN techniques. PHEME dataset is used for experimental purpose and achieved 79.5% accuracy on synthetic and social features.

(Asghar et al., 2019) proposed model combine BiLSTM and CNN approach. Benchmark PHEME dataset is used for experimental purpose and achieve 86.12% accuracy. χ^2 statistical test is used to check the effectiveness of the work. (Lie et al., 2019) developed BiGRU-CNN model for classification of Chinese tweets. The performance of the proposed model can be enhanced by using different set of features. Table 2.3 describes the various deep learning techniques.

Table 2.3 Deep learning techniques (Zhang et al., 2015)

Platform	Algorithm/ Technique	Methodology	Results	Drawbacks
PHEME (Alkhodair et al., 2020)	LSTM-RNN	Word2Vec	Word2Vec model with LSTM-RNN help the model to learn the best features	Accuracy of the model not so good
PHEME (Asghar et al., 2019)	BiLSTM-CNN	Word2Vector and FastText	86.12% accuracy achieved	Accuracy needs improvement

PHEME and Twitter (Liu et al., 2019)	Stacked LSTM + Stacked Attention	Twitter features	0.817 and 0.804, F1-score and accuracy	Computational cost is high
Twitter and Weibo (Ma et al., 2016)	RNN	TF*IDF	88.1% accuracy for Twitter and 90.1% for Weibo dataset	Need improvement in accuracy
PHEME (Ajao et al., 2018)	LSTM -CNN	Word2Vector	LSTM-CNN model achieved 82% accuracy	Performance need improvement

2.8 Bibliometric analysis of rumour detection

During the literature review we have done bibliometric analysis of rumour detection to identify main journals, most cited articles, most productive country, prominent authors and institutions. The present study uses VOSviewer software to implement the bibliometric analysis of 2935 records related to rumour dissemination by collecting the data from the Web of Science for the period of 1989-2021. The bibliometric results have shown publications trends, main journals, most cited articles, most productive country, prominent authors and institutions. Further net map analysis illustrates the growth of rumour detection in past, present and future as well.

2.8.1 Methodology

The techniques used for the research contained two phases, 1) collection of data 2) bibliometric analysis of the data. In the first phase, scientific data was collected from the database. In the second part, bibliometric analysis was done to describe the various dimensions. The existing literature was studied for data collection and to identify salient findings, potential research gaps to highlight the boundaries of information

(Tranfield et al., 2003). Organised literature reviews were accomplished via an iterative sequence describing proper search keywords, examining the literature, and completing the investigation (Saunders et al., 2009). As the number of articles increased, various visualization tools were also developed to conduct the bibliometric analysis (Hou et al., 2018). Figure 2.3 describes the various steps used for this assessment.

An online bibliometric "Web of Science" database was used for research. Web of Science is the most extensive database currently managed by Clarivate Analytics, it contains numerous search databases to provide technological and academic exploration. The database covers 21,100 peer-reviewed, 205,000 conferences proceedings and 104,000 selected books. It is more comprehensive than other databases and captures the most reputed international journals.

Clarivate Analytics's Web of Science (WoS) is the world's leading scientific citation search and analytical information platform. It is used as both a research tool supporting a broad array of scientific tasks across diverse knowledge domains as well as a dataset for large-scale data-intensive studies. WoS has been used in thousands of published academic studies over the past 20 years.

All types of scientific papers collected from the database were chosen for the study and analysed through VOSviewer. VOSviewer is one of the best tools for bibliometric analysis freely available on www.VOSviewer.com developed by Nees Jan van Eck and Ludo Waltman(Van et al., 2007). Different from other computer programs, VOSviewer provides an enhanced graphical demonstration of the data. The viewing proficiency of VOSviewer is suitable to map massive content (Van et al., 2010).

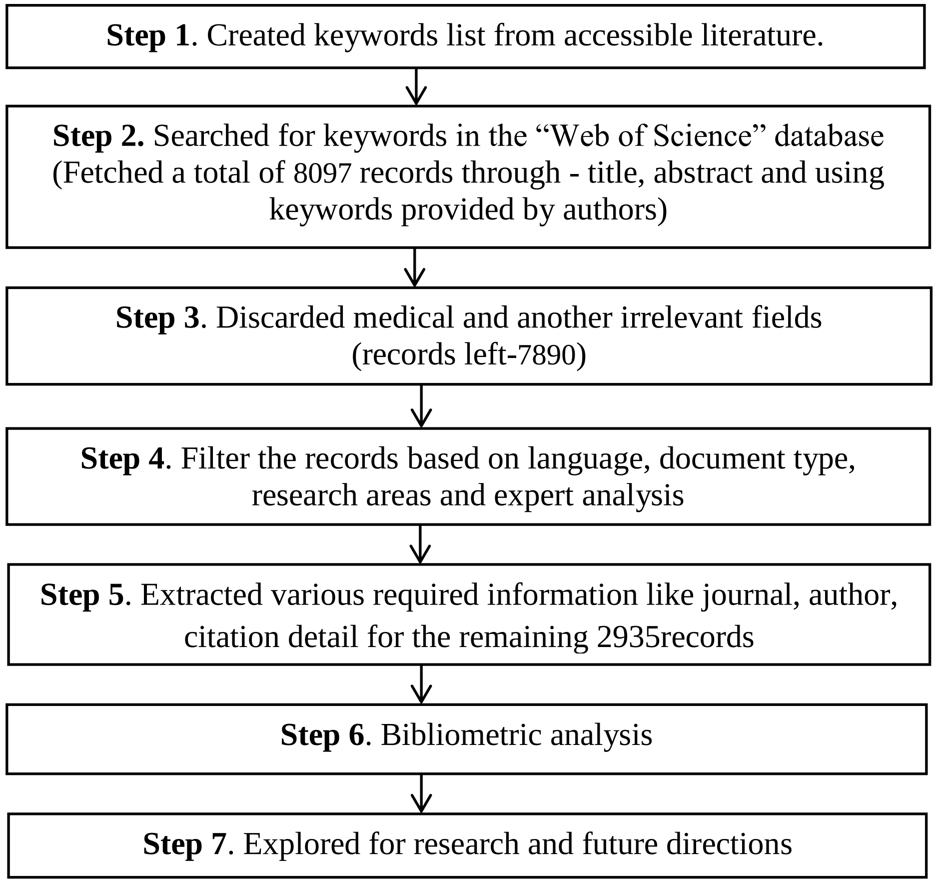

Fig. 2.3 Data refining and analysis procedure

2.8.2 Describing the suitable search terms

The keywords used for the collection of data are Rumors, framework, Techniques, social websites, social network, Approaches, algorithms. Combination of these keywords are including Rumors and framework, Rumors and Techniques, Rumor and "Technique", Rumors and "social websites", Rumors and "social sites", Rumors and social sites, Rumors and social network, Rumors and "social networks", Rumors and "social media", Rumours, "Fake news" and Social networks, "Fake news" and social media," Fake news", Rumours and Approaches, Rumours and social websites, Rumours and social network, Rumours and "social media", Rumors and "Approach", Rumors and Approaches, Rumors and "algorithms", Rumors and "framework". The keyword phrase-wise numbers of records accessible from the Web of Science database are presented in Table 2.4.

Table 2.4 Keywords associated to rumour related research

Keyword	Output through Web of Science 2021	Keyword	Output through Web of Science 2021	Keyword	Output through Web of Science 2021
Rumors and framework	200	Rumors and “social networks”	392	"Fake news"	1767
Rumors and Techniques	350	Rumors and “social media”	341	Fake news	1854
Rumors and “Techniques”	204	Rumors and social media	445	Rumours and Approaches	624
Rumors and “Technique”	173	Rumors	6346	Rumours and Social websites	25
Rumors and social websites	25	Rumour	6346	Rumours and Social sites	65
Rumors and “social websites”	1	“Rumours”	2348	Rumours and “social network”	224
Rumors and “social sites”	0	“Rumors”	2305	Rumours and “social media”	341

Rumors and social sites	65	“Fake news” and social networks	327	Rumours and social media	445
Rumors and social networks	642	“Fake news” and social media	818	Rumors and “Approach”	478
Rumors and “Approaches”	182	Rumors and “algorithms”	135	Rumors and “framework”	188
Rumors and Approach	624	Rumors and algorithms	244	Rumors framework	200

In the next phase, redundant records were discarded, and the remaining 1907 records were subjected to bibliometric analysis.

2.8.3 Preliminary search outcomes

The abstract of several journal articles, conference papers, books, and the chapter of books from the database were stored using the various keywords and title. The constraint and data selection setting in VOSviewer is shown in Table 2.5.

Table 2.5 The parametric constraint in VOSviewer and selection of data

Constraint Setting	
Duration	1989-2021
Terminology	Heading, abstract, keywords ,combination of keywords
Node type	Country, authors, cited journal, keywords, references

Retrieval approaches	
Search technique	Subject-oriented retrieval
Search interval	1989 -2021
Search term	Rumours, Framework , Social Sites, Approaches, Fake news
Database	Web of Science
Search result	2935 records

The data collection related to rumour transmission was primarily fetched from the Web of Science that included the most significant and prominent journals (Song et al., 2016). As a result 8097 unique documents were collected from 1989 to 2021. By applying filters as shown in Figure. 2.3, we left with 2935 relevant records only for this research work.

2.8.4 Bibliometric Analysis Results

The purpose of the bibliometric analysis is to examine publication trends, prominent journals, authors, institutions, and countries by using MS- excel tool (Broad, 1987). Also, predicting the emerging model and techniques can be applied in the rumour detection over social media domain while using the VOSviewer.

2.8.5 Publications of Rumour detection

The research journey in rumour detection and its related fields started 30 years back in 1989, starting with 23 publications during 1989, 28 publications during 1992, 31 publications during 2005, 47 articles during 2007 and 593 during 2021.

Table 2.6 Publication years

Sr. No.	Publication Years	Record Count	% of 2,935	Sr. No.	Publication Years	Record Count	% of 2,935
1	1989	23	0.78%	17	2005	31	1.06%
2	1990	8	0.27%	18	2006	38	1.30%
3	1991	13	0.44%	19	2007	47	1.60%
4	1992	28	0.95%	20	2008	34	1.16%
5	1993	11	0.38%	21	2009	31	1.06%
6	1994	18	0.61%	22	2010	47	1.60%
7	1995	28	0.95%	23	2011	60	2.04%
8	1996	27	0.92%	24	2012	61	2.08%
9	1997	26	0.89%	25	2013	75	2.56%
10	1998	21	0.72%	26	2014	87	2.96%
11	1999	27	0.92%	27	2015	88	3.00%
12	2000	25	0.85%	28	2016	98	3.34%
13	2001	22	0.75%	29	2017	150	5.11%
14	2002	23	0.78%	30	2018	239	8.14%
15	2003	30	1.02%	31	2019	343	11.69%
16	2004	30	1.02%	32	2020	553	18.84%
				33	2021	593	20.20%

The table 2.6 depicts the publication from 1989-2021 along with their citation. The initial period did not contribute much to publication. From the research, it is found that the last 15 year from 2006-2021 are the most productive. As shown in Figure 2.4 there has been an exponential evolution in the publication related to rumour detection since 2006.

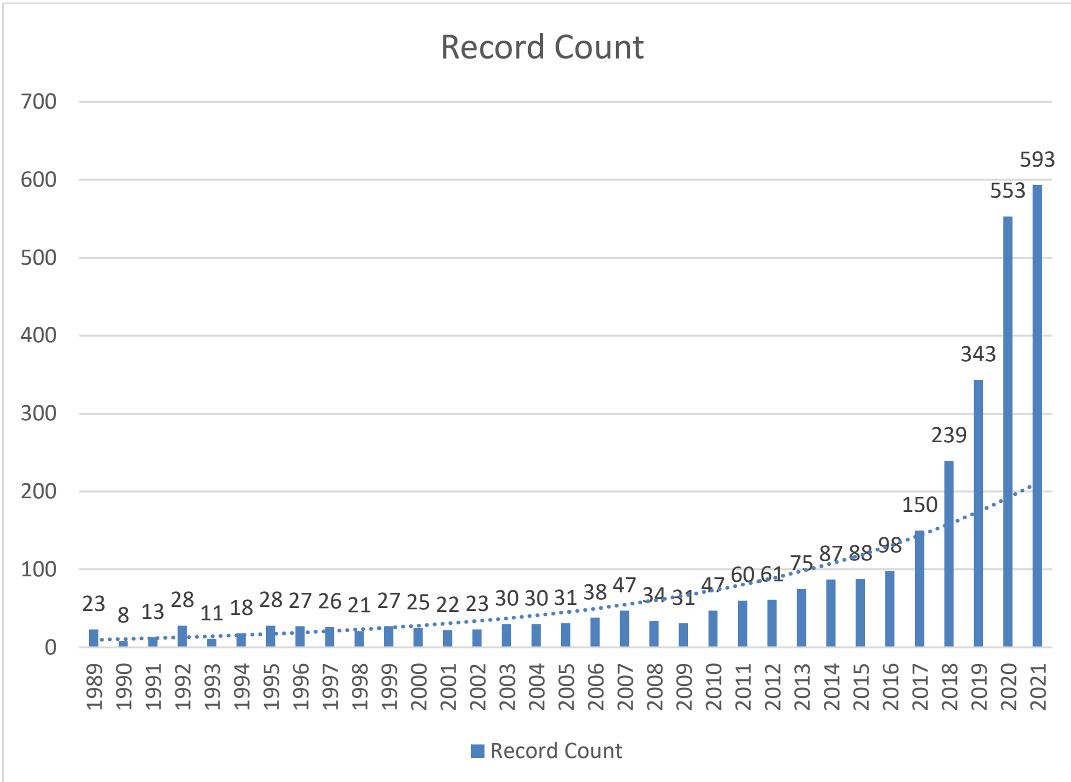

Fig. 2.4 The publishing trend in the field of Rumour Detection

2.8.6 Citation analysis

VOSviewer permits scholars to enhance the knowledge and emergent trend by co-citation networks (Xie, 2015). Citation and co-citation are different measurements for conducting bibliometric analysis (Pilkington and Meredith, 2009). Heavily cited articles greatly influence the subject compared to those articles less cited by people. Such articles are an indication of greater importance in the field (Garfield and Merton, 1979). The citation analysis defines all reference used in the article without being restricted by the investigator, so it is possible to collect a large sample and another proper perception from the articles, journals and authors (Garfield, 1983).

2.8.7 Productive Articles

Citations measure the significance of publications in the specified area of investigation (Garfield, 1972). Citation specifies the research accomplishments by researchers in a specific field (Smith, 2008). However, the citations changed with time are the citation life cycle (Galiani and Gálvez , 2017). A paper is cited by other articles known as a local citation, and the overall Scopus citations for the paper termed as the global citation. Table 2.7 illustrates the top ten maximum times cited articles in rumour detection.

Table 2.7 Most prominent articles in rumour detection

Paper	TC	ACPY
Allcott H, 2017, J Econ Perspect	1034	172.3333
Lewandowsky S, 2012, Psychol Sci Public Interest	915	83.1818
Lazer Dmj, 2018, Science	900	180
Del Vicario M, 2016, Proc Natl Acad Sci U S A	602	86
Brockmann D, 2013, Science	584	58.4
Archer J, 2005, Pers Soc Psychol Rev	546	30.3333
Funk S, 2009, Proc Natl Acad Sci U S A	511	36.5
Tandoc Ec, 2018, Digit Journal	500	100
Chakrabarti D, 2008, Acm Trans Inf Syst Secur	396	26.4
Lewandowsky S, 2017, J Appl Res Mem Cogn	326	54.3333

(TC- Total Citations, ACPY- Average Citation per Year)

Table 4 indicates that the articles written by (Allcott et al.,2017) are mostly cited with a total of 1034 citations. The second most cited article authored by (Lewandowsky et al.,2012) (TC= 915, ACPY=83.1818) followed by articles authored by (Lazer et al.,2018) (TC=900, ACPY=180), (Del Vicario et al.,2016) (TC=602, ACPY=86), (Brockmann et al.,2013) (TC=584, ACPY=58.4), (Archer et al.,2005) (TC=546, ACPY=30.3333), (Funk et al.,2009) (TC= 511, ACPY=36.5) , (Tandoc et al.,2018) (TC=500, ACPY=100), (Chakrabarti et al.,2008) (TC=396,ACPY=26.4), (Lewandowsky et al., 2017) (TC=326, ACPY=54.33) in that order.

2.8.8 Most Productive country in rumour detection research

One thousand nine hundred seven publications from 87 countries were selected for analysis. The top ten countries with leading publications are presented in Table 2.8.

Table 2.8 Productive countries in Rumour detection

Countries/Regions	**TP**	**% TP**
USA	1,039	35.40%
Peoples R China	324	11.04%
England	290	9.88%
Canada	157	5.35%
Australia	144	4.91%
Germany	138	4.70%
Spain	117	3.99%
France	116	3.95%
India	80	2.73%
Italy	77	2.62%

(TP- Total Publications)

Table 2.8 identifies most productive regions, the prominent regions with an input of 1,039 (35.40%) is the USA. The USA is followed by China 324 (11.04 %), England 290 (9.88%), Canada 157 (5.35%), Australia 144 (4.91%), Germany 138 (4.70%), Spain 117(3.95%), France 116 (3.95%), India 80(2.73%), Italy 77(2.62%).USA has received more attention from scholars and have more far-reaching significances than those published by other countries. The United States is the most prolific publisher of high-quality research in the world having high citation count. Thus, among the top ten productive countries, China and the USA are the leading countries in the research.

2.8.9 Prominent Journals

Table 2.9 illustrates the most prominent journals in rumour detection. 1057 journals published the paper in this field. The top 10 most prominent journals are described in Table 2.9.

Table 2.9 Productive journals

Source Titles	TP	TC	h_index	g_index	m_index	PY_START
Plos One	67	1684	21	40	1.75	2011
Profesional De La Informacion	37	444	11	20	2.2	2018
Information Processing & Management	29	351	9	18	1.285714	2016
International Journal of Communication	28	104	7	9	0.5	2009
Scientific Reports	28	769	12	27	1.2	2013
Social Media + Society	28	273	7	16	1.4	2018
Media And Communication	27	196	6	13	1	2017
Information Sciences	23	259	7	16	1	2016
Library Journal	22	0	0	0	0	1994
Complexity	67	50	4	5	0.666667	2017

(TP- Total Publications, TC- Total citations, h-index- Hirsch Index, PY start- Publication Starting Year).

From Table 2.9 it can be identified that the ‘Plos One’ with (h-index= 21) is the most productive journal. It has total publication 67 with 1684 total citation and its first publication start in the year 2011. Further it is discovered that the Profesional De La Informacion (TP=37) is the second most prolific journal having (h-index=11) and it started publishing in rumor detection since 2018. In terms of publications ‘Profesional De La Informacion’ further followed ‘Information Processing & Management’ (29), 'International Journal of Communication' (28), and ' Scientific Reports ' (28), in that order. From Table 6 it is revealed that Plos One and Complexity both journal having highest number of publications but Complexity having least number of total citation.

2.8.10 Prominent Authors

During the period under study, 7209 authors authored one thousand nine hundred seven publications individually or collaboratively. Lotka's law defines author productivity in the given field (Macroberts, 1982). The topmost productive authors with their affiliation, publication start year, total publications are tabulated in Table 2.10.

Table 2.10 Productive authors

Author	TP	TC	h_index	g_index	m_index	PY_start
Difonzo N	21	989	16	21	0.552	1994
Zhang Y	19	467	6	19	0.75	2015
Bordia P	18	1002	17	18	0.586	1994
Pennycook G	18	1384	12	18	2.4	2018
Rand Dg	18	1386	12	18	2.4	2018
Li J	13	79	5	8	0.833	2017
Wang Y	13	464	5	13	0.333	2008
Zhang Z	12	120	5	10	0.625	2015
Wang J	12	389	6	12	0.462	2010
Zhu L	12	114	4	10	0.571	2016

(PY_Start- Publication Starting Year, TP - Total Publications, TC-Total citations)

2.8.11 Prominent Institutes

The most prolific organizations in the area of rumor detection illustrated in Table 2.11. 1682 organizations contributed to the publications. The top ten universities along with total publications and percent with 2935 records shown in Table 2.11. The leading university with 63(2.147) publications is University of California.The University of California is distantly followed by the University Of London 52(1.772%), University of Texas System 41(1.397%), Nanyang Technological University 39 (1.329%), State University System of Florida 34 (1.158%), Massachusetts Institute Of Technology Mit and University of Cambridge 32(1.09%), Harvard University and Pennsylvania Commonwealth System of Higher Education Pcshe 31(1.056%), University Queensland 22(1.154) in such order.

Table 2.11. Top ten most productive organizations in the area of rumor detection

Affiliations	TP	% TP
University of California System	63	2.147
University of London	52	1.772
University of Texas System	41	1.397
Nanyang Technological University	39	1.329
State University System of Florida	34	1.158
Massachusetts Institute of Technology Mit	32	1.09
University of Cambridge	32	1.09
Harvard University	31	1.056
Pennsylvania Commonwealth System of Higher Education Pcshe	31	1.056
University Queensland	22	1.154

(TP-Total Publication)

2.8.12 Research area

Rumors occur in various fields attracting researchers to study it including its growth, reasons and other issues. Table 2.12 depicts the top 10 fields researched for rumours led by 503 research articles published in computer science by different authors. The computer science is distantly followed by Communication 474(16.15%), Psychology 337(11.48%),History 317(10.80%), Business Economics 288(9.81%), Information Science Library Science 265(9.03%), Science Technology other Topics 236(8.04%), Mathematics 214(7.29%), Government Law 199(4.23%), Social Sciences Other Topics 124(4.23%).

Table 2.12. Top ten research areas in rumor detection

Research Areas	TP	% (TP)
Computer Science	503	17.14%
Communication	474	16.15%
Psychology	337	11.48%
History	317	10.80%
Business Economics	288	9.81%
Information Science Library Science	265	9.03%
Science Technology Other Topics	236	8.04%
Mathematics	214	7.29%
Government Law	199	6.78%
Social Sciences Other Topics	124	4.23%

(TP-Total Publication)

Rumors occur in various fields attracting researchers to study it including its growth, reasons and other issues. Table 2.12 depicts the top 10 fields researched for rumours led by 503 research articles published in computer science by different authors. The computer science is distantly followed by Communication 474(16.15%), Psychology 337(11.48%),History 317(10.80%), Business Economics 288(9.81%), Information Science Library Science 265(9.03%), Science Technology

other Topics 236(8.04%), Mathematics 214(7.29%), Government Law 199(4.23%), Social Sciences Other Topics 124(4.23%).

2.8.13 Keyword Analysis

Garfield described the concept of keyword mining (Garfield,1990). Many studies examined author keywords as a measure of topic development (Liu et al., 2014). Keywords are simplification and abbreviation of the paper (Wang, 2018). Table 2.13 presents the top twenty keywords used in the article. Extraordinary keywords attract the researchers in the particular field (Zhao, 2013).

Table 2.13 presents the top twenty words used in the author keyword. Table 2.14 presents the most used word in the paper title and most frequently used words in abstract are shown in table 2.15 . Extraordinary keywords attract the researchers in the particular field [67].

Table 2.13 . The most popular words In Author's keyword

Words	**Occurrences**	**Words**	**Occurrences**
fake news	507	social networks	54
social media	290	rumors	53
misinformation	196	communication	49
disinformation	142	post-truth	44
rumor	111	social	44
covid-19	87	rumor spreading	42
twitter	77	fact-checking	40
journalism	62	facebook	39
media	62	information	39
news	54	credibility	38

Table 2.14 . The most frequently used words in paper titles

Words	Occurrences	Words	Occurrences
news	748	political	132
fake	585	online	127
rumor	434	spreading	108
social	409	analysis	102
media	318	covid-	95
rumors	296	rumour	89
networks	189	based	84
model	159	rumours	82
misinformation	135	disinformation	78
detection	134	study	75

Table 2.15. The most frequently used words in Abstract

Words	Occurrences	Words	Occurrences
news	3647	misinformation	774
social	2454	article	768
fake	2206	network	758
media	2089	paper	736
rumor	1534	public	731
study	1315	spread	721
rumors	1223	networks	685
model	1178	analysis	678
political	941	data	654
online	793	based	632

Table 2.13 depicts that the words "misinformation", "false memories", “social media” frequently appear in articles on rumours during 1989-2019. Similarly Table 2.14 shows that the other high occurrence keywords are “information”, “media”, “communication”, and “social media”. These keywords are part of data retrieval shown in Table 2.5.

2.8.14 Net-Map Analysis

The theory of keyword abstraction was used to understand the development of the field. The current study illustrates that numerous researchers (Li et al., 2009) extract keywords to inspect the area's progress. The subsequent records were created via VOSViewer software to calculate the occurrence of the keyword. Co-occurrence means the combination of two keywords in an article (Bhardwaj et al., 2020). Different colours were used to describe the keywords terms in the network. The items collected in a similar cluster defined the same topic. The common keywords among clusters defined the largest node of the cluster (Watts and Strong, 1998). The alteration of colours among the cluster reveals the gain of the study. Figure 2.5 displays the co-occurrence of author-supplied keywords through a mean network diagram designed via VOSviewer.

The minimum existence of the words in the diagram is five. There are 986 items, five clusters, 43,688 links among all items, and the net map's total strength is 63,836. Figure 2.5 assembled the keyword items among five major nodes - War, Rumor, Model, Experiment ,Covid. Table 2.16 illustrates the total number of cluster and their associated key terms used in the network diagram.

The cluster with the key term war is the biggest node that specifies the subject in the preliminary stage of investigation related to rumor detection among social media. Total links produced by this cluster are 308, with a strength of 6172 and key terms used for 372 times - the associated keyword terms being market, price, aggression, victim, share, gossip, child etc. The initial phase of research led to social media researchers from diverse fields worked in various areas. Social media was primarily used in propaganda to spread information or ideas to influence feelings and actions. Figure 2.5 also displays perceptions learned in the early stage, which investigators further used in social media. It is shown by various keywords like transmission, firm, story, rule, uncertainty etc. in the cluster having the largest node rumor. The total links produced by this cluster is 348 with a strength of 36274 and 2429 occurrence of key terms. Here, the presence of Nigeria indicates the concern of Nigerian authors in rumor detection. As the social media network became popular many new social media sites like Facebook, Twitter, YouTube, WhatsApp were developed for numerous purposes. People use these platforms for various purposes like education, medical, news, information diffusion and many more. Hence, social media was mainly used for political purposes during this period, as shown by the keywords terms politics, election etc. During the U.S. constitutional voting, applicants and followers enthusiastically used Facebook and Twitter for promotions and expressing their sentiments (Jin et al., 2017).. Much false information disseminated during the presidential election. Political parties used social networks for their promotion, as the masses relied less on traditional media and more on new communication channels. Various algorithms developed to check the accuracy of different social media.

Table 2.16 . Net- map analysis

Cluster No.	Top item name	Links	Strengths	Main topic	Occurrences
1	War	308	6172	Market, price, aggression, victim, share, gossip,child	372
2	Rumor	348	36274	Transmission, firm,story,rule, uncertainity,investor	2429
3	Model	344	5408	Network, algorithm, node,performance, message	277
4	Experiment	302	5901	Credibility, trust, accuracy, participant, reasoning	302
5	Covid	301	8768	Belief, disinformation, practice, country, fear	384

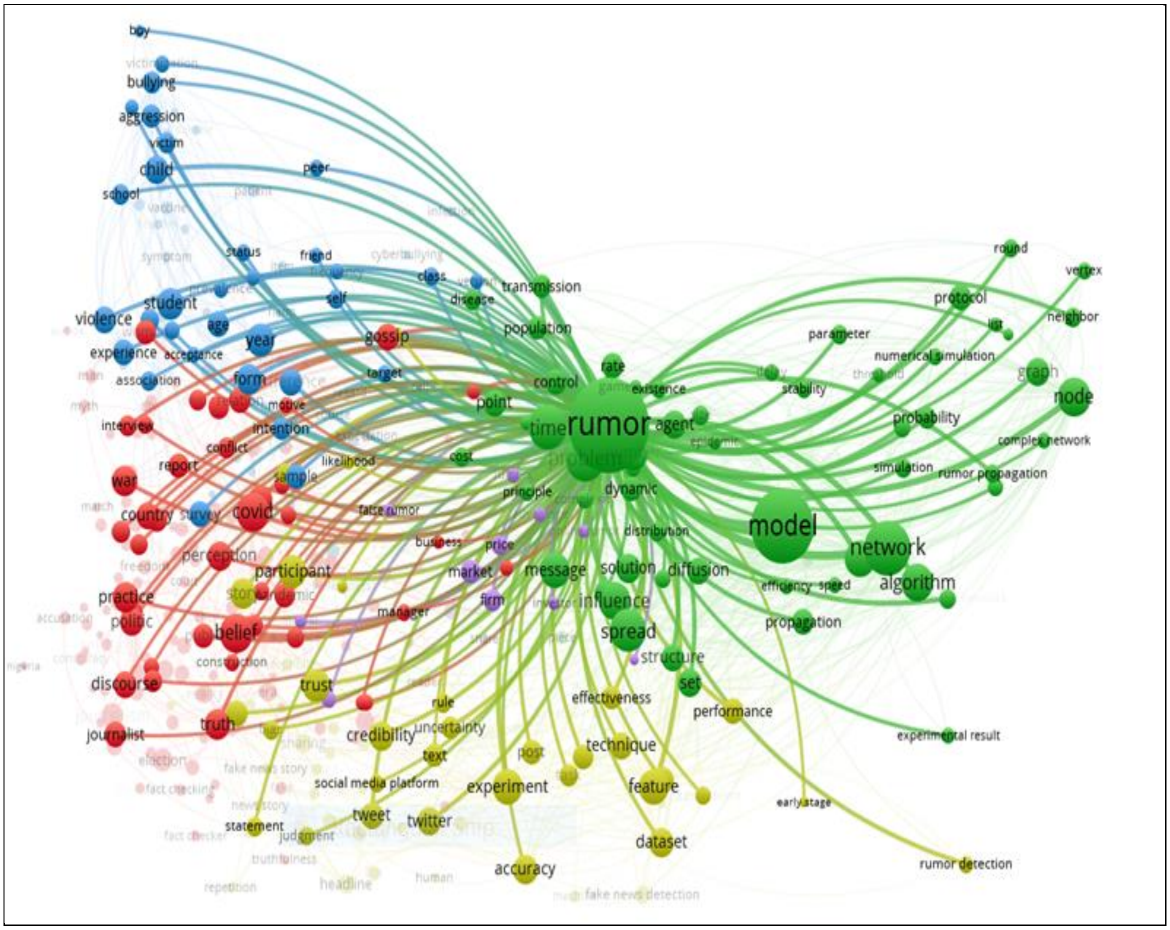

Fig. 2.5 Net-Map diagram of keyword analysis

Consequently, numerous rumours disseminated through social media sites. People misused these sites to spread panic in society through fake accounts created by individuals with vested interest to spread rumors and fake news.

With the development of social sites, researchers also developed numerous rumours detection model and applications. The cluster of the keyword term model in Figure 3 indicates this. Many keywords associated with this cluster are Network, algorithm, node,performance, message, etc. Total links produced by this cluster are 344, with a strength of 5408 and 277 occurrence of key terms.

algorithm, node,performance, message, etc. Total links produced by this cluster are 344, with a strength of 5408 and 277 occurrence of key terms.

The investigation dynamics of rumour dissemination is typically centred on rumour spreading model SIS (Li and Ma, 2017), SIR (Zhao et al., 2013), SEIR (Liu,2017) and SHIR model (Liu, 2016). The development in network science also improved certain significant features of complex networks (Barabási and Albert,1999).This inspired researchers to intensify research in a complex network by combining with rumour dissemination features. A model for rumour broadcast for small-world systems was defined (Zanette, 2002).

Researchers considered the topological features of the model and deliberated the properties of the association on transmission law (Vazque, 2006).The rumour dissemination among scale-free systems was also defined (Nekovee et al., 2007). Hence research in rumour detection was primarily accomplished through the growth of epidemic theory and networks, and then multidisciplinary research was carried out. The machine and deep learning techniques were applied to the various rumour detection models to check their efficiency.

Figure 2.5 also shows that as the number of model increases more experiments implemented by the researchers to mitigate the problem of rumour detection. The cluster with key term experiment revels this. Many key terms associated with this cluster are credibility, trust, accuracy, participant, reasoning etc. Many features, dataset and techniques are explored by the researchers to resolve the rumor detection problem.

Figure 2.5 also revealed that as social networking platforms grew, their usage increases in the medical field and shown by the cluster named with the keyword Covid.

Numerous keywords associated with this cluster are Belief, disinformation, practice, country, fear etc. 301 links were produced by this cluster, with a strength of 8768 and 384 occurrences of key terms.

Social networking sites were used for the medical field in a typical way (Obrien, 2016). Doctors used these sites to consult with their patient and colleagues on different healthcare issues and questions. The cancer patients used social networking sites to interact with each other and oncology professionals, form communities to obtain valuable information about their health (Attai et al., 2016). A study showed sentiment analysis of Twitter messages posted during the coronavirus pandemic (Rajput et al., 2020).

The year-wise development in social media increases . The journey of social media sites shows how social media users increased every month. Research has estimated 351.4 million active users using social media for various purposes. The price of social media sites increased daily. On average 8768 most companies spend $ 4,000 to $ 7,000 a month to manage social media. The above keyword analysis illustrates much improvement in the field of social media rumour detection. The researchers adopt different definitions to illustrate the terms related to social media rumour detection.

2.9 Chapter Summary

This chapter demonstrates a systematic literature study on rumour detection for identification of research gaps as well as future scope. Various affects and types of rumours are defined in this chapter. Numerous machine and deep learning methods are defined to know the researchers contribution towards rumour detection. Further the chapter shows the bibliometric study to know prominent authors, countries, journals, institutions, most cited articles, most used keywords for rumour detection by collecting the data from the Web of Science database.