Journal of Theoretical and Applied Information Technology
31st March 2023. Vol.101. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A SYSTEMATIC LITERATURE SURVEY FOR DETECTING RUMORS BASED ON MACHINE LEARNING AND DEEP LEARNING

THANAA MOHAMED HASSAN[1] . YEHIA MOSTAFA HELMY[2] . DOAA S. ELZANFALY[3]

[1]Lecturer, Faculty of Commerce & Business Administration - Helwan University, Egypt

[2]Proffesor, Faculty of Commerce & Business Administration - Helwan University, Egypt

[3]Associate Professor, Faculty of Computers & Artificial Intelligence – Helwan University, Egypt

E-mail: [1]thanaahassan@hotmail.com, [2] ymhelmy@commerce.helwan.edu.eg, [3]doaa.saad@fci.helwan.edu.eg

ABSTRACT

Rumor detection is considered to be one of the key research areas in Social Network Analysis and it is a vital in preventing the propagation of misinformation in social networks. Several rumor detection techniques have been introduced in the recent years. These techniques presented the problem as a classification issue, for example binary ones (rumor or non-rumor). The majority of these techniques are based on machine learning (ML). The main obstacle in these techniques is mainly correlated to feature extraction from a selection of dataset. The manual extraction of features affects the efficiency of rumor detection for most of these works, due to the time and effort required. Another technique applied recently is the Deep networks which have been suggested as a means to streamline feature extraction and to offer a strong and superior capability for learning abstract representations. In general, spotting trending rumors requires the development of a robust yet flexible model that can capture long-range connections between postings and produce different representations for accurate early discovery. The aim of this paper is to present a number of carefully studied works in the area of rumor detection by focusing on machine learning and deep learning sectors. These studies are examined in the literature review section leading to answer the research questions that are addressed in this paper. This review is significant and helpful for researchers as it will enable researchers to compare their work with current works due to the accessibility of the complete description of the used evaluation matrices, dataset characteristics, and whether they applied machine learning or by applying deep learning model per each work. Furthermore, said review will also discuss the challenges that researchers in this area have faced and suggested a few potential avenues for further research.

Keywords: *Rumor Detection, Rumor Tracking, Deep Learning, Machine Learning, Social Media Analytics*

1 INTRODUCTION

Users of social media have the facilities to generate a huge amount of information, whether it's real or misleading, this amount reaches and have an impact on millions of other users. False information has recently been used in information warfare as a weapon. How to detect incorrect material on social media efficiently and effectively which become a difficult problem. As a consequence, it is crucial to identify and suppress harmful rumors before they seriously affect people's lives.

According to the definition b Pathak et al. in [1], a rumor is information with highly questionable veracity. However, rumors may be true, and others may be false, and still others may continue to be unconfirmed. Not all false information is considered a rumor. Misinformation is also referred to as mistakes that individuals honestly and sincerely make. Others could be malicious intentional rumors spread with the intent of tricking people who don't believe them. Said are classified as false news and are categorized depending on the intent of the originator.

Rumor detection is commonly considered as a four-step process. The first phase in the process begins by collecting info from various social media platforms. It is necessary to convert the acquired data into a consistent organized format in order to extract the necessary features. Preprocessing tasks include consolidation, cleaning, transformation, and reduction. The significant features are then extracted (such as content-based, pragmatic, and network-specific features), and the dataset is categorized as a rumor or not using a variety of machine learning or deep learning algorithms.

Based on Cao et al [2], the majority of automatic rumor detection systems see the rumor detection default as a binary classification task that can be categorized using one of the below paradigms:

- In specifically, the hand-crafted features-based machine learning (ML) paradigm: To explain how rumors are dispersed in high dimensional space, they employ handcrafted features. Feature engineering is one of the essential prerequisites for these techniques. Depending on the classifier that is employed to differentiate between hyper planes, these techniques extract features from both the textual and the visual material. [3] [4].
- Networking paradigm: this method is used to assess the network's credibility and legitimacy by combining various heterogeneous structural social networking features (such as the number of followers, content responses, timestamp, and so on) with optimal graph-based algorithms [4] [5].
- Deep learning (DL) paradigm: in order automatically aggregate and learn multi-modal properties, deep learning is applied. While both DL and ML are based on learning from data, ML uses hand-crafted features, whereas the DL paradigm eliminates the need for feature extraction activities by allowing the classifier to correctly learn and acquire the required feature during training. With the feature extraction phase gone, DL performance has significantly improved in the paradigm.

This study presents a variety of well-researched works in the machine learning and deep learning sectors, and uses this compilation to give a thorough strategy for recognizing machine and deep learning. Support Vector Machine (SVM), Logistic Regression (LR), Bernoulli Naive Bayes (BNB), SGD Classifier (SGD), K-Nearest Neighbor (K-NN), and Decision Tree were all used as machine learning approaches (J48). Ensemble machine techniques such as Random Forest (RF), AdaBoost (Ada), and Bagging were also used (Bag). Many deep learning methods, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), bias error correction trees (BERTs), and others, were described. Our systematic review then analyses these papers in depth to address the six research goals we specified to gain a comprehensive understanding of the current state of the art in rumor detection using machine learning and deep learning techniques. In addition, our systematic review details the problems and obstacles that have been encountered by researchers in this field and proposes exciting new avenues for study. Researchers in this field will find our study useful because it provides a comprehensive description of the performance matrices, dataset features, and the model utilized for each work, allowing for easier comparison with the current works. Our analysis will also help scientists locate annotated datasets that can serve as test beds for evaluating their own methods against industry standards.

The rest of the paper is organized as follows: A survey of the related studies to our literature survey are provided in Section 4. Section 5 presents a detailed literature review methodology as well as an analysis of the studies considered. The various ML and DL architectures utilized in rumor detection are examined in detail in Section 6. Section 7 discusses the challenges and unresolved issues raised by our survey. Finally, Section 8 will give a paper's conclusion.

2 RELATED SURVEY STUDIES

Numerous research studies have looked closely at the characteristics of rumors. Some researchers examined rumor detection in social media from various aspects. The review paper by Zubiaga et al. [6] gives a summary of the research into social media rumors and various detection techniques while paying less attention to the DL algorithms. The

authors of [2] classified previous research findings to three major paradigms: (i) machine-learning-based approaches, where feature extraction is the primary and most important step in developing reliable classification algorithms; (ii) propagation-based approaches, in which mining relationships among entities observes rumors; and (iii) neural network-based approaches.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1: Existing Surveys on Rumor Detection

Research Paper	ML	DL	Techniques
[6]	✓	×	Covered different aspects of rumor classification systems covering ML methods and neglecting DL methods
[2]	×	P	Covered only two categories: RNN-based models and CNN-based models and gave more attention on the rumor definition and characteristics
[7]	×	×	Discussed deeply characteristics of rumors and features of false information; -Poorly covered ML and DL methods
[8]	✓	P	Briefly discussed DL methods giving more consideration on the feature-based methods
[9]	×	✓	The work focuses on detecting rumor using only DL methods
[10]	×	P	Several DL methodologies have been used in order to review the issues that are now present, and provide the necessary fixes, and validation of MID in online social networking. Poorly covered the DL methods.
[11]	×	✓	They list all NLP techniques in detail and discuss their advantages and disadvantages. They conducted a thorough examination of recently published DL-based studies.

DL: Deep Learning; P: partially covered; X: not covered at all; ✓: Covered

In their 2018 study, Kumar and Shah [7] highlighted the ways in which actors distribute false information as well as the methods by which algorithms are created to identify it. Zhou and Zafarani [8] identified some prospective features of fake news while also using DL approaches to identify it. Al-Sarem gave a thorough analysis of the literature on rumor detection on microblogging platforms like Twitter and Weibo using only DL methods [9]. Islam, Md. Rafiqul, et al. [10] examined thoroughly automated misinformation detection on (i) incorrect information, (ii) rumors, (iii) spam, (iv) fake news, and (v) disinformation using DL techniques in their survey article. Mridha [11] has discussed significant evaluation metrics in fake news identification by presenting a survey of DL techniques.

The most important studies on rumor identification using both ML and DL algorithms will be reviewed in detail in the current paper. Said survey will also put highlight on open datasets that have already been researched by other studies.

In summary, the majority of studies that have been published focus on rumor detection in general rather than focusing on the potential applications of ML and DL for rumor detection. Existing studies on rumor identification on social media are presented in Table 1.

In this paper, the main focus is on detecting rumor using both ML and DL methods. Following the SLR methodology.

3 REVIEW METHODOLOGY

In this survey, we utilized and adhered to the systematic review methodology provided in [12]. There will be four major subsections in this section: the first sub-part will deal with the review protocol plan, the second with the research question, the third with the information source, and the fourth with the selection criteria.

3.1 Systematic Review Protocol Planning

In the first step, we set out the methods to be used in the review and provides an explicit plan for the work. We began by choosing research studies from several libraries and databases. Following this, number of the selected and chosen research has been reduced in accordance with the criteria of inclusion and exclusion. Then, decision about the review questions should be formulated to conduct and carry out the suggested proposed study.

3.2 Review Questions

The proposed review questions one of its goals is to deeply analyze the advantages that ML and DL applications offer for rumor detection systems. Additionally, it lists the public datasets that are currently used to perform rumor detection and all the possible evaluation matrix techniques to analyze the rumor detection performance.

RQ1: How and what types of publications have been distributed over the last five years?

RQ2: Which datasets are most frequently used for rumor analysis?

Journal of Theoretical and Applied Information Technology
31st March 2023. Vol.101. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

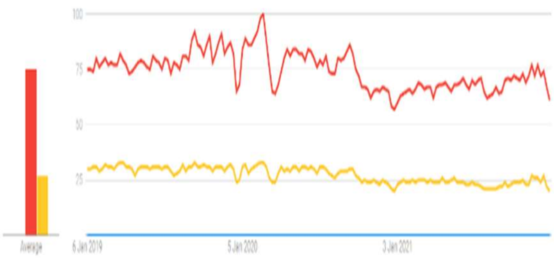

Figure 1: Machine Learning Versus Deep Learning Extracted from Google Trends

RQ3: What are the ML techniques used most for detecting rumors? Which of these techniques reached the highest performance?

RQ4: What are the DL techniques used for detecting rumors? Which of these techniques reached the highest performance?

RQ5: Which evaluation matrix techniques are mostly used for analyzing rumor detection performance?

RQ6: What are the main challenges and potential directions for the field of rumor detection??

3.3 Source of Information

A variety of libraries have been chosen to undertake the systematic literature review, including:

- IEEE Explore ()
- ACM Digital Library ()
- Science Direct ()
- Google Scholar ()

A number of keywords were used in order to search for all the appropriate papers “Rumor detection" or “Fake information” WITH either:

— “Machine Learning” or “[the name of any ML technique]”

— "Deep Learning”, "Deep Neural Networks" or [the name of any DL approach].

Following the search process, the papers that are generated include both rumor detection and DL or rumor detection and ML. We also included papers through Google Scholar only when the other identified digital libraries were unable to locate them.

Additionally, we have restricted our inclusion to only those publications done on Twitter and other microblogging sites like Sina Weibo. Additionally, the most comprehensive version of a study is provided if it appears in multiple journals or conference proceedings.

Concerning the exclusion criteria, we excluded all research researches that had nothing to do with the identified questions, as well as any duplicate works or older pieces published before 2018.

3.4 Selection Criteria

There are numerous surveys in the said field of rumor detection, and a set of strict criteria has been established in order to select the most pertinent and relevant research for our survey. For these standards, there are both inclusion and exclusion requirements. We included the years 2018-2021 in the inclusion criteria. Figure 1 depicts a global comparison of the current trends in the area of social networking analysis for the ML and DL domains. In comparison to the use of DL (shown in yellow color on Figure 1, this can be observed that using classical ML methods is more widespread, the red line in Figure 1.

3.5 Analysis of Related Studies

Concerning this section, we will find, review, in addition to evaluating the results of the relevant papers, as well as providing answers to previously specified questions. The solution to RQ1 may be found in Table 2, which categorizes chosen publications on both ML and DL based on the publishing type, the year of publication, and the technique used.

Number of citations is an important indicator of the quality of published research work. According to **Error! Reference source not found.**, approximately 38% of published research studies have more than 15 citations and are having large, sizable audience. The most frequently study was Guo, H.'s (2018) paper which had 131 citations, and Ma, J.'s (2019) study came in second with 115.

3.6 Rumor Detection in ML and DL

3.6.1 Dataset

PHEME, Kaggle, Newly Emerged Rumors, Liberia - Ebola 2015, and Credibility Corpus are among the publicly available datasets. In the following sections, we will pay special attention to the datasets being used in said research in order to track rumors using ML and Dl approaches.

Table 2: Selected Papers Classification (Citation via Google Scholar)

Publication		Type of Publication		Technique Applied	
Source	Citation Number	Journal	Conference	ML	DL
[13]	2	✓		✓	
[14]	1	✓		✓	✓
[15]	128		✓	✓	
[16]	22	✓		✓	
[17]	7	✓			✓
[18]	10	✓		✓	✓
[19]	1	✓		✓	
[20]	2	✓		✓	
[21]	2		✓	✓	
[22]	2		✓	✓	
[23]	65	✓		✓	
[24]	19	✓			✓
[25]	4		✓		✓
[26]	1		✓		✓
[27]	16	✓		✓	
[28]	5		✓	✓	
[29]	21		✓		✓
[30]	115		✓		✓
[31]	4		✓		✓
[32]	7		✓	✓	
[33]	131		✓		✓

- *PHEME dataset*

Zubiaga et al. [34] created the PHEME public dataset. The dataset is derived from Twitter through crawling tweets with the Twitter streaming APIs. In addition to, including both rumors and non-rumors tweets sent out throughout breaking news.

The authors of [24], [30], and [31] employed the PHEME dataset that is considered a benchmark for stance identification, in order to test several approaches on the rumor detection task. In [24], they ran many experiments including the publicly accessible dataset PHEME to figure out the optimal hyper parameter values for the proposed CNN model. [30] used two datasets, PHEME with 1,123 rumors and 1,123 non-rumors, which was gathered based on 5 breaking news, so it claims overlapped more than TWITTER, which was collected based on 498 non-rumors and 494 rumors published on snopes.com. The [31] dataset contains 5,802 discussion samples, 1,972 of which are rumor samples and 3,830 of which are non-rumor samples.

- *Twitter and Weibo dataset*

Rather than using publicly available datasets, a group of academics prefers to create and collect their own. Researchers can obtain sample data mainly on social networking sites as Twitter and Sina Weibo via APIs.

Without any prior preparation: [13] and [17] have used the available ArCOV-19 dataset that includes Arabic tweets concerning the COVID-19 pandemic and covers the same period from 27 January to 30 May 2020. The overall number of tweets collected was lowered to 3157 tweets, including 1480 rumors (46.87%) and 1677 non-rumors (53.12%) while [18] collected an amount of significant Arabic tweets using the Twitter streaming application interface and Tweepy Python library for four months starting from January 1, 2020, to April 30, 2020, containing more than 4,514,136 million tweets also [21] have been streaming tweets in real-time using Tweepy.

With 130 total samples, they have created their own dataset for [16]. The basic questions that can be found on Twitter, Facebook, and numerous news stories served as the basis for the samples.

Table 3: Published Research Work

Citation Number	>=15	>=5 and <15	<5
Papers Count	8	4	9

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

For training purposes, the administrator of [19] provides data from Twitter; they have chosen 400 ham messages and 100 spam tweets. A total of 200 messages have been chosen for testing. The chosen datasets are stored as text files. Speaking about [20], a dataset called the Health Connected Rumors Dataset (HRRD), which compiles tweets concerning rumors about cancer treatments/disease, was developed. Using Twitter's streaming API, [23] compiled a dataset of tweets that covered the three-month period from 15 January to 15 April 2020. A total of 121,950 credible tweets and 287,534 non-credible tweets were identified and categorized from the total number of tweets. Work in [33] used a twitter dataset of rumor events (498) and non- rumor event (493). [29] made use of Ma et al. public's Twitter dataset, which included 992 distinct identified groups. [26] and [33], carried out studies using Weibo data. The dataset contains 2313 rumors and 2351 non-rumor samples, each of which includes a different type of data. In [27], they proposed an excellent research tool regarding rumor spreading and rumor refutation spread analyses. The amount of data collected was 38365 microblogs from 3999 Weibo users (3793 valid). [28] was publicly obtained from Twitter data consisting of rumor and non-rumor labeled tweets. 18571 data were employed in the experiments.

- Kaggle

The dataset is available in CSV format and includes three files, each of which contains a list of websites that have been independently cited on Snopes.com, Emergent.info, and Politifact.com. Work in [32] has obtained business reviews from the publicly accessible Kaggle dataset.

- News articles dataset

Speaking of [14], it has used two datasets, one of which was manually compiled from a large scale of news articles from a trustworthy online newspaper website. They then extract these data and store it in a CSV file as columns.

- Others

The other dataset of [14] was collected from several social media platforms like Facebook, WhatsApp, boom.live, etc. and manually extract in the same way as they previously done it with the fact data.

In [22], the datasets used in the experiment mainly have three sources. Data collected from micro-blogs through crawl tools, or through the internet or through microblogs platform. [15] collected both real news and fake claims that surfaced on social media on COVID-19 topic.

Fake claims are collected from various fact-checking websites like Politifact1, NewsChecker2, Boomlive3, etc., and from tools like Google fact-check-explorer4 and IFCN chatbot5. Real news is collected from Twitter using verified twitter handles.[25] Handcrafted COVID-19 dataset obtained from Chinese platforms with 3737 rumor related data.

- Summarizing dataset research studies

To answer the question RQ2, the section above provides detailed information about the existing datasets that are intensively used for conducting rumors detection task. Figure 2: Considered Studies For The Distribution Of Dataset Types shows that, (63%) some researchers prefer available public datasets or creating their own datasets by crawling Twitter or Weibo microblogs. In 17% of cases use rumor tracking websites, 12% of researchers use public PHEME dataset and 4% research propose to use news articles' datasets and Kaggle dataset.

Concerning this section's goal is to respond to research questions RQ3 and RQ4. We'll begin by explaining why ML is used in rumor detection. On the other hand, we identify the key distinctions between DL and traditional machine learning. In addition, we will discuss the distribution of ML and DL methods for rumor detection.

3.6.2 Machine Learning and Deep Learning Technique

- Purpose of Machine Learning

Machine learning-based techniques have emerged as one of the viable approaches for rumor detection on social media. A state-of-the-art depicting the use of supervised machine learning (ML) algorithms in the field of rumor detection on social media is presented. The most popular machine learning techniques used are SVM, Naïve Bayes, and Decision Trees. Techniques like KNN, Clustering, Naïve Bayes, Random Forests, Logistic Naïve Bayes, Natural Language Processing, social spam framework of analytics and detection have also been used.

Journal of Theoretical and Applied Information Technology
31st March 2023. Vol.101. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

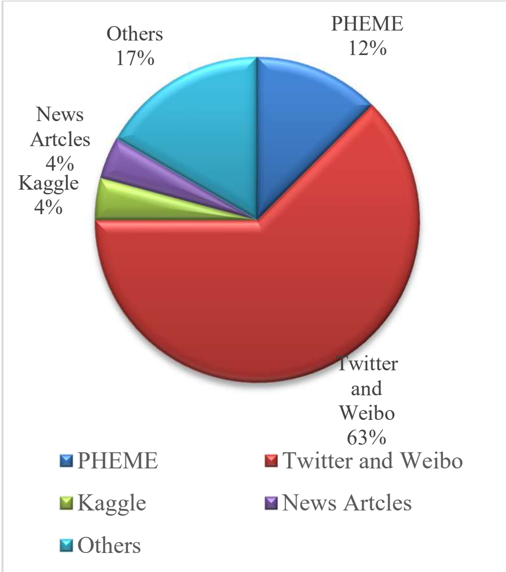

Figure 2: Considered Studies For The Distribution Of Dataset Types

- Purpose of Deep Learning

DL provides a computational performance with many processing layers that will allow the learning of data representations with multiple processing layers excluding any feature engineering [36]. On the other side, ML classifiers depend on feature engineering, which is commonly labor- and time-intensive. Research teams are increasingly interested in using DL for rumor detection to reduce the demand for specialized features. Furthermore, DL can find more relevant hidden features than conventional machine learning methods [37].

- Machine learning-based Rumor Detection

To answer RQ3, we will first go over the architecture, tools, and performance metrics of the ML techniques used in the research studies. The ML methods for rumors detection reported in the analyzed research studies are compiled in

Table 4 including the tools and frameworks employed as well as the evaluation results that were applied after using the ML approach.

- Out of the 12 research papers, 5 of them proposed models for rumor detection. [21] created an automated system that uses a number of medical keywords that will be updated automatically by using the Wikipedia API to feed real-time Twitter Data relevant to health field [22] applies a rumor detection model relying on Naive Bayes and NLP to the data volume of detecting micro-blog rumors. Using supervised machine learning classifiers, [32] created an automated system for detecting business rumors in online business reviews. [23] has suggested a methodology based on ensemble learning for evaluating the validity of a significant number of tweets. [16] has developed a web Search Engine Misinformation Notifier Extension (SEMiNExt) that combined machine learning and natural language processing (NLP) into this newly proposed extension.
- Out of the 12 studies, 5 suggested a comparison of various ML classifiers. In order to classify spam, [19] compared logistic regression and the Naive Bayes algorithm. They have also proposed a method for leveraging the Naive Bayes algorithm to block a specific group of nodes in order to recognize spammers on Twitter and instantly put an end to those rumors. [28] have applied a comparison for choosing the best technique to enhance the rumor detection problems, those techniques are the OneR (One Rule), Naive Bayes, ZeroR, JRip, Random Forest, Sequential Minimal Optimization, and Hoeffding Tree methods. A thorough evaluation is also given. Extreme gradient boosting (XGBoost), random forest (RF), and logistic regression (LR) were the four machine learning techniques that were investigated and contrasted by [27]. [14] tested various machine learning approaches with one deep learning mechanism (they combined the multilayer neural approach with three unique layers in the training environment) (Logistic Regression, K-Nearest Neighbor, Random Forest, and Support Vector Machine). [15] examined the performance of Decision Tree, Logistic Regression, Gradient Boost, and Support Vector Machine applying manual annotated dataset of 10,700 social media postings and research of both true and misleading news regarding COVID-19.
- Two research used data in Arabic. [20] developed their own dataset employing Arabic-language social media. [13] developed an in-depth proposal that covers two phases: detection and tracking using the Arcov-19 dataset.

Table 4: Summary of Machine Learning Related Studies

Paper Ref.	Dataset	Technique Applied	Evaluation Matrix	Findings
[20]	Health-Related Rumors Dataset (HRRD), an Arabic dataset that contains a tweets collection discussing rumors regarding cancer disease and treatment.	Machine learning techniques including Support Vector Machine (SVM), Logistic Regression (LR), Bernoulli Naive Bayes (BNB), SGD Classifier, K-Nearest Neighbor (K-NN), and Decision Tree (J48). In addition, an ensemble machine approaches were applied: Random Forest (RF), AdaBoost (Ada), and Bagging (Bag).	Accuracy, Precision, Recall, and F-1 Score	The Random Forest produced the best accuracy (83.50%).
[21]	Tweepy, an open-source Python module that allows programmers to interact with the Twitter API, was used as the dataset.	Using machine learning, they created an automated method to detect medical rumors on Twitter.	Precision, Recall, F1 score and Support	Atchived 90% accurate analysis of the user-based, content-based, and network-based aspects.
[22]	Three sources of datasets are used in the experiment. (1) Data gathered from microblogs using crawlers. (2) Open microblog data collections gathered from the web and microblogs (3) Microblog rumors gathered from microblogging platforms.	A rumor detection method was proposed by the study which combines a machine learning algorithm relying on Naive Bayes and NLP particles.	F1-score, ROC curve with two numerical axes: the true and false positive rates	This detection approach cuts processing time by almost 10 hours when compared to manual rumor detection. Additionally, it now correctly detects 15% more micro-blog rumors.
[19]	The datasets chosen are text files obtained from Twitter spammers. They chose 400 ham messages and 100 spam messages for training purposes. For testing, a total of 200 messages were chosen.	In this research, they examined the ideal strategy for social network rumor blocking. The ability of the analyses to recognize spam communications is tested using Naive Bayes as a supervised machine learning classification technique.	Recall, accuracy, precision, and F-measure	When comparing Naive Bayes with logistic regression, it was found that the Naive Bayes approach outperform with an accuracy of 97.5% has an accuracy of 88.5%.
[32]	Business reviews was obtained from Kaggle.	Applying supervised machine learning classifiers as: Logistic Regression, Support Vector Classifier (SVC), Naive Bayesian (NB), and K-Nearest Neighbors (KNN), the contribution of this research lied in the developing of an automated system for identifying business rumors from online business evaluations and reviews.	F measure, precision, recall, and accuracy are employed.	The Naive Bayesian classifier made a much better performance than any other machine learning classifiers that were examined performing an accuracy of 72.43%.
[28]	The data set utilized in the tests was acquired by Kwon et al. from publicly available Twitter data [35]. The tweets in this data collection are classified as rumors or	OneR (One Rule), Naive Bayes, ZeroR, JRip, Random Forest, Sequential Minimal Optimization, and Hoeffding Tree algorithms were applied for rumor detection	There applied a comparison between all those algorithms in terms of precision, recall, F-	In general, the accuracy of the classification algorithms shows that Random Forest and SMO was found to be higher than that of other methods while

	not. The experiments utilized 18571 data.		measure, accuracy, ROC area, and PRC area metrics.	ZeroR have the lowest accuracy.
[27]	They crawled Sina Weibo's microblogs whereby they collected 38365 microblogs from 3999 Weibo users of which 3793 were legitimate.	Talking on this study provides a general framework of machine learning method were applied. For the sentiment and short text similarity analyses, natural language processing (NLP) was used as well as four machine learning techniques (logistic regression (LR), support vector machines (SVM), random forest (RF), and extreme gradient boosting (XGBoost)) were analyzed and also compared.	Precision, Recall as well as F1 Score were applied for evaluation.	Sentiment and similarity text analysis has proven to be important, and the highest optimal performance was found with the XGBoost.
[14]	They used two types of datasets related to the COVID-19 pandemic: facts and rumors. • A large number of news stories from reputable online newspaper sites were gathered manually and news blogs were collected for the fact data. • Speaking of rumor information will be gathered from various social media platforms including Facebook, WhatsApp, boom.live, and others.	A one deep learning mechanism as well as numerous machine learning techniques (Logistic Regression, K-Nearest Neighbor, Random Forest, and Support Vector Machine) were applied.	F1- score Matrix.	The news classification schema based on deep learning achieved an efficiency of almost 90% succeeded in providing the highest f1-score.
[13]	The dataset used is publicly accessible in the "tweet verification" subdirectory.	This study's method is divided into two stages: detection and tracking. They used two models in the detection phase: the stacking ensemble approach, which analyses tweet texts, and Support Vector Machine that is a genetically based algorithm, that will work on both user and tweet features. Various similarity-based techniques discovered during the tracking phase in order to collect the top 1% of tweets that were similar to a target tweet or post in order to locate the rumors source.	F1-score, recall, precision, and accuracy were applied.	The experiments results were very interesting in the tracking phase, in terms of ROUGE L precision, recall, and F1-Score for similarity approaches, as well as providing promising results in terms of accuracy, precision, recall, and F1-Score for rumor detection with an accuracy of 92.63%.
[15]	They gathered both COVID-19 true and false information from social media. They also gathered fake claims from fact-checking websites such as Politifact1, NewsChecker2, Boomlive3, and others, in addition to that tools such	They used the term frequency-inverse document frequency (TF-IDF), and they benchmarked the annotated dataset using the four machine learning baselines which are Decision Tree, Logistic Regression, Gradient Boost, and Support Vector Machine.	Accuracy, recall, precision, and F1-score are considered for evaluation.	SVM have achieved the best results on the test set, providing F1- score with 93.32%.

	as Google fact-check-explorer4 and the IFCN chatbot5. Twitter verified handles which are typically used to gather real news.			
[23]	A dataset of tweets that was gathered over a three-month period, from January 15- April 15, 2020, using the Twitter's streaming API. Their search was conducted using the related hashtags and keywords to COVID-19.	This paper attempt to provide an ensemble-learning-based approach in order to evaluate the large number of tweets validity. They ensemble learning was used in order to combine six machine learning algorithms and run numerous experiments on the gathered and labelled dataset. This model was also improved after combining the ensemble-stacking model with the chosen machine-learning models.	A standard deviation and average accuracy of each model was obtained and used as the 1-sigma error.	Using the proposed framework a high degree of accuracy in distinguishing between credible and non-credible tweets was demonstrated and obtained clearly for the COVID-19 information.
[16]	A data set was made. They treat each sentence in a search query as a separate sample, speaking on their data set which includes a 130 samples on total. Basic questions from popular social media platforms such as Twitter and Facebook, as well as numerous news articles, were obtained in order to serve the samples.	This study had been presenting a web Search Engine Misinformation Notifier Extension (SEMiNExt), that was created by combining both the machine learning and natural language processing.	Recall, accuracy, precision, and F1 score were used	After this study it had been proven that the SEMiNExt under an artificial neural network (ANN) performs the best in terms of accuracy, F1-score, precision, and recall values of 93%, 92%, and 92%, respectively.

- Deep learning-based rumor detection

To respond to RQ4, we first review the architecture, libraries, tools, and performance metrics of the DL techniques used in the research studies.

Table 5 summarizes the DL techniques proposed in the rumors detection field that were discovered in the studies considered. It also identifies and employs the frameworks, tools, and libraries used, as well as the performance achieved after employing the DL method.

- Five studies proposed models for rumor detection. For rumor identification in [29], they suggested combining CNN and an attention-residual network to capture long-range dependency. Convolution DL CNN architecture is the foundation of the model that was proposed in [31]. Then, they compared their results to those of the DL- and machine-learning-based rumor detection methods that are currently in use. A novel was proposed to deep hybrid learning model in [17] for detecting COVID-19-related rumors on social media, relying on concatenated parallel convolutional neural networks and a long short-term memory (LSTM–PCNN). The pre-trained BERT model has been combined with the TextCNN and TextRNN models in [25]. For rumor detection, [33] proposed a hierarchical LSTM network with social attention.
- In [18] the paper is focusing on the feature representation they compared the performance of word2vec and FASTTEXT on both ML and DL techniques to identify Arabic misinformation related to COVID automatically.

Table 5: Summary of Deep Learning Related Studies

Paper Reference	Dataset	Dataset Size	Technique Applied	Feature Representation	Findings	Limitation
[29]	Two Twitter datasets one proposed by Ma et al and the other is a Twitter Monitor defines a keyword-based query method.	From the first dataset 992 events and from the other 1,409	Attention-residual network combined with CNN	Content feature to generate word vector matric	Attention-residual network mixed with CNN beats other state-of-the-art models concerning both rumors identification in addition to early rumor detection when compared to prior research that exclusively relied on RNN.	Need to compare their findings with other attention models to measure their effectiveness.
[24]	The publicly available dataset PHEME	Not stated	Deep learning model based on (CNN)	Word Vector	The proposed model is built on CNN architecture demonstrating one of the best performances in comparison to the existing rumor detection approaches using and depending on both DL and machine learning techniques.	To assess the performance of their model, it is necessary to investigate how robust this discovery is on different NLP tasks.
[17]	ArCOV-19 dataset (Arabic COVID-19)	3,157 tweets total, including 1677 non-rumors and 1480 rumors.	Three parallel and concatenated convention al neural networks used together with the long short-term memory, (LSTM) (PCNN).	word2vec, GloVe and Fast Text model	Speaking on this study a novel hybrid deep learning model for COVID-19 detection was created (LSTM–PCNN). As a result of this study the model produced one of the highest results, with an accuracy of 86.37%.	More investigation using Arabic Tweets should be considered with other DL methods
[25]	COVID-19 dataset created by hand and collected	They supplied 3737 rumor-related	TextCNN and TextRNN are used	BERT Model was performed	Applying BERT model had outperformed the other models. The BERT, TextCNN,	It is impossible to rerun this model as the dataset used

	from Chinese platforms	data points.			and TextRNN models achieved an accuracy higher than 90%.	is not accessible.
[18]	Arabic COVID-19 obtained via Twitter	8786 Tweets were used.	Applied both ML and DL techniques	They used Word2vec, FASTTEXT, TF-IDF, and n-gram.	Fast Text performed better for DL approaches, but TF-IDF word level performed better for ML. The XGB classifier was the most effective at detecting Arabic misinformation.	Preprocessing limitation
[26]	A set of publicly accessible data sets	2313 rumors and 2351 non-rumors samples were used	They applied the attention-based DL method,	Not Mentioned in this study	User comments is used to provide a crystal-clear sign for spotting the rumor trend.	Need to shorten training time and improve model performance
[30]	PHEME and Twitter	Twitter (498 non rumors and 494 rumors) PHEME (1,123 rumors and 1,123 non rumors)	Generative adversarial network	Use a stochastic gradient descent algorithm with small batches, 200 epochs, a hidden/concealed vector of 100, and a word size of 5000.	Comparing the proposed method to modern methods, the results are significantly superior. On the PHEME dataset, accuracy was 78.1%, and on the TWITTER dataset, it was 86.3%. Low-frequency strong non-trivial patterns has been detected by the model.	Thy must improve and enhance the accuracy of the model.
[31]	PHEME	There are 1,972 rumor samples and 3,830 non-rumor samples.	Three recurrent neural networks (RNN) were used as the following: Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Bi-	In this study they didn't use the contents they used the tweets' propagation patterns instead to identify fraudulent	They eliminated the dimensionality of the input feature set that helps in improving and enhancing the training time. Due to the nature of the time-series data generated which had reduced the computational complexity of classification. Increasing the micro-averaged F1 score by 4.6% in comparison to other baselines.	Evaluation matrix applied needs to be more accurate.

Journal of Theoretical and Applied Information Technology
31st March 2023. Vol.101. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

			directional Recurrent Neural Network (Bi-RNN) at the same time they applied also one convolutional neural network (CNN).	information		
[33]	Twitter: All rumor events are indexed using keywords gleaned from Snopes' fake news database. Weibo Dataset: Non-rumor events are collected during crawling posts in general threads, on the other side rumors are confirmed by the Community Management Center of Sina.	On Twitter, there are rumor events (498) and non-rumor events (493). While on Weibo there are rumor event (2,313) and non-rumor event (2,351)	For rumor detection, they have proposed a unique hierarchical network with social attention (HSA-BLSTM).	In this study paper, the social features are user profile, propagation and post texts	Extensive experiments and tests using both datasets (Weibo and Twitter) have shown that the proposed model (HSA-BLSTM) gave better results than any other art models significantly showing an accuracy of 94.3% and 84.4% respectively, for HSABLSTM.	Their model's accuracy needs to be investigated much more, and it needs to be compared to other deep learning models.

3.6.3 Rumor Detection Evaluation Matrix

To answer RQ5, we should start by reviewing the different techniques of the evaluation matrix that are mostly used for analyzing rumor detection performance and then stating the ML and DL evaluation techniques proposed in the rumor's detection field found in the research studies.

- Evaluation Matrices

Concerning classification problem, we should be prepared with various assessment metrics to analyze the classification algorithm. Including Confusion Matrix, Precision, Recall, Accuracy, F measure, and Area under ROC curve (AUC).

- *Confusion Matrix*

Confusion matrix for the classification problem with two classes is shown in Figure 3. According to the graph, there could be four different forecasts for the results. Whether true positive or true negative outcomes both are correct classifications, however false positive in addition to false negative outcomes result in two types of errors.

A false positive example is a negative class that is inaccurately classified as positive, and a false negative example a positive class that is inaccurately classified as negative. Kohavi and Provost [38] 3define the context of our research entrance to the confusion matrix as below:

(a) Representing the number of correct negative forecasts,
(b) the number of correct positive forecasts,
(c) the number of correct negative forecasts, and
(d) the number of correct positive forecasts.

		Predicted class	
		Negatives	Positives
Actual class	Negatives	*a*	*b*
	Positives	*c*	*d*

Figure 3: Confusion Matrix

- *Precision*

It can be viewed and described as either the fraction of correctly predicted positive classes that occurred, or it can be viewed through number of accurate outputs provided by the model. Said can be observed by applying the formula below:

Precision= TP/TP+FP

It is the percentage of total positive classes that our model correctly predicted.

- *Recall*

The recall rate is calculated as follows and should be as high as possible.

Recall= TP/TP+FN

- *Accuracy*

It is considered one of the most important concerns when assessing the accuracy of a classification task. This shows how the model frequently predicts the outcome correctly. It can be calculated as the ratio of the classifier's total number of predictions to the number of correctly accurate predictions it made. The formula is as follows:

Accuracy = TP + TN / TP + FP + FN + TN.

- *F-measure*

It is challenging comparing two models with low precision but high recall, or vice versa. Following, the F-score can be used regarding this. This score enables us at the same time to evaluate recall and precision. The F-score is at its highest when recall and precision are both equal. It can be calculated using the formula below:

F-measure= 2*Recall*Precision / Recall + Precision

- *Receiver Operating Characteristic Curve (ROC curve)*

The ROC is a graph that displays and illustrates the classifier's performance across all possible thresholds. A graph depicts the true positive rate (on the Y-axis) and the false positive rate (on the x-axis).

- ML and DL Evaluation Matrices in the related Research Studies

Speaking of machine learning papers, [13], [15], [19], [20], [21], [27], and [32] have employed Precision, Recall, F-1 score, and Accuracy in their publications. [14] uses the F1 scoring measures. In [16], they use a confusion matrix table to visually demonstrate the algorithm's performance and offer insight. When determining if a system model is valid, [22] takes into account both index values as well as the F 1-Score, which is the weighted average of recall and accuracy. Additionally, the ROC Curve, whose two number axes are the true positive rate in addition to false positive rate, has been applied. In [23], they calculated the 1-sigma error by taking the average accuracy of each model and the standard deviation. They compared the algorithms in [28] using precision, recall, F-measure, accuracy, ROC area, and precision-recall curve (PRC) area measurements.

Regarding deep learning papers, evaluation matrix used to create all of these measures in [17], [24], [25], [26], [30], and [33] was accuracy, precision, recall, and F-score. [31] considered evaluation metric known as the F1 score. [18] included evaluation measures such as the Area under the ROC Curve (AUC), precision, recall, and F1. In [29], they used Accuracy, Precision, and F1 Score to evaluate the performance of their model in various aspects.

4 CHALLENGES AND OPEN ISSUES

This section is responding to RQ6 question, which is concerned with the major future directions in addition to challenges proposed in the field of rumor detection, by highlighting unresolved problems.

4.1 Rumor Detection Problems and Obstacles

When researching and examining the material of rumor detection on social media networks, academic researchers face numerous challenges.

We'll go over a few of the challenges that academics who use social media data face.

a. Applying means of hashtags and keywords to find tweets about a specific topic does not guarantee that all relevant keywords-based queries and hashtags, therefore, the dataset will be biased.

b. Quick and accurate detection aids in limiting the spread of rumors and mitigating their negative effects on society. As a result, the data sets related to source detection must be collected in real-time.

c. Source identification and tracking across linked social networks. A number of social networking websites, including Facebook, Twitter, and others, provide accounts for users. This various social networks that people use can occasionally be used to spread rumors.

d. Data retrieval in case of using different keywords or hashtags to access different data which requires additional data preparation and transformation, including format conversion and filtering out irrelevant data. Furthermore, to obtain data written in a specific language we use is considered one of the huge problems.

4.2 FUTURE PATH AND DIRECTIONS

Because ML and DL are promising technologies for rumor detection, researchers should take note and exercise caution when choosing the best architecture. The following suggestions provide information regarding the response to research question RQ6:

- The development of a linkable or interconnected social network and the discovery of the genuine sources within a linkable social network might be seen as a wide and challenging area that requires further study.
- The majority of research studies rely solely on textual elements. It is preferable to use DL when working with visual formats as videos and images.
- The dataset is gathered using keywords that were previously identified as being associated with rumors. Most of the time, data retrieved is solely written in the language of the queries that were actually run. Because of this, it will be crucial to keep an eye out for rumors written in other languages in the future.
- The researchers are urged to be as specific as possible about the tools they used and the evaluation performance matrix obtained. The scientific community will be able to reproduce the findings and enable additional developments in the process.

5 CONCLUSION

The spread of rumors has been proven to be incredibly unsafe for society. Therefore, we should find a way to diffuse these rumors. Machine learning (ML) and deep learning (DL) have gained great success in the field of rumor detection. Several ML and DL-based methods have been proposed for rumor detection on social networks. In this paper, six review questions have been proposed: the types of publications that were distributed during the last five years; the datasets that are most frequently used; the techniques of ML and DL and their performance; the evaluation matrix techniques mostly used for rumor analysis; and finally, the challenges and potential directions for the rumor detection field. To answer those review questions, we conducted methods using 21 published papers in the field of ML and DL. Specifically, we provided a comparison for both DL and ML based on the methods, performance evaluation, tools, and architecture. We also provided the dataset's principle based on its type, source, and content. Finally, we highlighted the challenges and future directions for research on rumor detection

REFRENCES

[1] Pathak, Ajeet Ram, Aditee Mahajan, Keshav Singh, Aishwarya Patil, and Anusha Nair. "Analysis of techniques for rumor detection in social media." Procedia Computer Science 167 (2020): 2286-2296.

[2] Cao, Juan, Junbo Guo, Xirong Li, Zhiwei Jin, Han Guo, and Jintao Li. "Automatic rumor detection on microblogs: A survey." arXiv preprint arXiv:1807.03505 (2018).

[3] Castillo, Carlos, Marcelo Mendoza, and Barbara Poblete. "Information credibility on twitter." In Proceedings of the 20th international conference on World wide web, 2011, pp. 675-684.

[4] Jin, Zhiwei, Juan Cao, Yazi Zhang, and Yongdong Zhang. "MCG-ICT at MediaEval 2015: Verifying Multimedia Use with a Two-Level Classification Model." In MediaEval. 2015.

[5] Jin, Zhiwei, Juan Cao, Yongdong Zhang, and Jiebo Luo. "News verification by exploiting conflicting social viewpoints in microblogs." In Proceedings of the AAAI conference on artificial intelligence, vol. 30, no. 1. 2016.

[6] Zubiaga, Arkaitz, Ahmet Aker, Kalina Bontcheva, Maria Liakata, and Rob Procter. "Detection and resolution of rumours in social media: A survey." ACM Computing Surveys (CSUR) 51, no. 2 (2018): 1-36.

[7] Kumar, Srijan, and Neil Shah. "False information on web and social media: A survey."arXiv preprint arXiv:1804.08559 (2018).

[8] Zhou, Xinyi, and Reza Zafarani. "A survey of fake news: Fundamental theories, detection methods, and opportunities." ACM Computing Surveys (CSUR) 53, no. 5 (2020): 1-40.
[9] Alkhodair, Sarah A., Steven HH Ding, Benjamin CM Fung, and Junqiang Liu. "Detecting breaking news rumors of emerging topics in social media." Information Processing & Management 57, no. 2 (2020): 102018.
[10] Shah, Kapil. "A Review: Rumors Detection on Twitter Using Machine Learning Techniques." International Journal of Scientific Research & Engineering Trends (2020).
[11] Mridha, Muhammad F., Ashfia Jannat Keya, Md Abdul Hamid, Muhammad Mostafa Monowar, and Md Saifur Rahman. "A comprehensive review on fake news detection with deep learning." IEEE Access 9 (2021): 156151-156170.
[12] Keele, Staffs. "Guidelines for performing systematic literature reviews in software engineering." EBSE Technical Report ver. 2.3 (2007).
[13] Qasem, Sultan Noman, Mohammed Al-Sarem, and Faisal Saeed. "An ensemble learning based approach for detecting and tracking COVID19 rumors." Computers, Materials and Continua (2021): 1721-1747.
[14] Paul, Tulika, Samarjit Roy, Satanu Maity, Abhishek Bhattacharya, Soumi Dutta, and Sujata Ghatak. "COVID-R: a deep feature learning-based COVID-19 rumors detection framework." In Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2020, Volume 2, Springer Singapore, 2021: pp. 907-917.
[15] Patwa, Parth, Shivam Sharma, Srinivas Pykl, Vineeth Guptha, Gitanjali Kumari, Md Shad Akhtar, Asif Ekbal, Amitava Das, and Tanmoy Chakraborty. "Fighting an infodemic: Covid-19 fake news dataset." In Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021, Collocated with AAAI 2021, Virtual Event, February 8, 2021, Revised Selected Papers 1, pp. 21-29. Springer International Publishing, 2021.
[16] Shams, Abdullah Bin, Ehsanul Hoque Apu, Ashiqur Rahman, Md Mohsin Sarker Raihan, Nazeeba Siddika, Rahat Bin Preo, Molla Rashied Hussein, Shabnam Mostari, and Russell Kabir. "Web search engine misinformation notifier extension (SEMiNExt): A machine learning based approach during COVID-19 Pandemic." In *Healthcare*, 2021, vol. 9, no. 2, p. 156. MDPI.
[17] Al-Sarem, Mohammed, Abdullah Alsaeedi, Faisal Saeed, Wadii Boulila, and Omair AmeerBakhsh. "A novel hybrid deep learning model for detecting COVID-19-related rumors on social media based on LSTM and concatenated parallel CNNs." Applied Sciences 11, no. 17 (2021): 7940.
[18] Alqurashi, Sarah, Btool Hamoui, Abdulaziz Alashaikh, Ahmad Alhindi, and Eisa Alanazi. "Eating garlic prevents COVID-19 infection: Detecting misinformation on the Arabic content of Twitter." arXiv preprint arXiv:2101.05626 (2021).
[19] Vanmathi, C., and R. Mangayarkarasi. "An Analysis of Machine Learning Approach for Detecting Automated Spammer in Twitter." In Advances in Distributed Computing and Machine Learning: Proceedings of ICADCML 2020, pp. 443-451. Springer Singapore, 2021.
[20] Saeed, Faisal, Wael MS Yafooz, Mohammed Al-Sarem, and Essa Abdullah Hezzam. "Detecting health-related rumors on Twitter using machine learning methods." International Journal of Advanced Computer Science and Applications 11, no. 8 (2020).
[21] Dito, Fatima Mohammed, Haleema Adnan Alqadhi, and Abdulla Alasaadi. "Detecting medical rumors on twitter using machine learning." In 2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT), 2020, pp. 1-7. IEEE.
[22] Wenfeng, Gan, Zhang Hong, and Cheng Ruoyi. "A Study on Online Detection of micro-blog Rumors Based on Naive Bayes Algorithm." In 2020 Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), 2020, pp. 22-25. IEEE.
[23] Al-Rakhami, Mabrook S., and Atif M. Al-Amri. "Lies kill, facts save: Detecting COVID-19 misinformation in Twitter." *Ieee Access* 8 (2020): 155961-155970.
[24] Alsaeedi, Abdullah, and Mohammed Al-Sarem. "Detecting rumors on social media based on a CNN deep learning technique." *Arabian Journal for Science and Engineering* 45 (2020): 10813-10844.
[25] Chen, Shuaipu. "Research on Fine-Grained Classification of Rumors in Public Crisis——Take the COVID-19 incident as an example."

In *E3S Web of conferences*, vol. 179, p. 02027. E3S Web of Conferences, 2020.

[26] Wang, Lina, Wenqi Wang, Tong Chen, Jianpeng Ke, and Benxiao Tang. "Deep Attention Model with Multiple Features for Rumor Identification." In *Frontiers in Cyber Security: Third International Conference, FCS 2020, Tianjin, China, November 15–17, 2020, Proceedings 3*, pp. 65-82. Springer Singapore, 2020.

[27] Wang, Shihang, Zongmin Li, Yuhong Wang, and Qi Zhang. "Machine learning methods to predict social media disaster rumor refuters." *International journal of environmental research and public health* 16, no. 8 (2019): 1452.

[28] Bingol, Harun, and Bilal Alatas. "Rumor Detection in Social Media using machine learning methods." In *2019 1st International Informatics and Software Engineering Conference (UBMYK)*, 2019, pp. 1-4. IEEE.

[29] Chen, Yixuan, Jie Sui, Liang Hu, and Wei Gong. "Attention-residual network with CNN for rumor detection." In *Proceedings of the 28th ACM international conference on information and knowledge management*, 2019, pp. 1121-1130.

[30] Ma, Jing, Wei Gao, and Kam-Fai Wong. "Detect rumors on twitter by promoting information campaigns with generative adversarial learning." In *The world wide web conference*, 2019, pp. 3049-3055.

[31] Kotteti, Chandra Mouli Madhav, Xishuang Dong, and Lijun Qian. "Rumor detection on time-series of tweets via deep learning." In *MILCOM 2019-2019 IEEE Military Communications Conference (MILCOM)*, 2019, pp. 1-7. IEEE.

[32] Habib, A., Akbar, S., Asghar, M. Z., Khattak, A. M., Ali, R., & Batool, U. “Rumor detection in business reviews using supervised machine learning.” In 2018 5th International Conference on Behavioral, Economic, and Socio-Cultural Computing (BESC), November 2018, (pp. 233-237). IEEE.

[33] Guo, Han, Juan Cao, Yazi Zhang, Junbo Guo, and Jintao Li. "Rumor detection with hierarchical social attention network." In *Proceedings of the 27th ACM international conference on information and knowledge management*. 2018, pp. 943-951.

[34] Zubiaga, Arkaitz, Maria Liakata, and Rob Procter. "Learning reporting dynamics during breaking news for rumour detection in social media." *arXiv preprint arXiv:1610.07363* (2016).

[35] Kwon, Sejeong, Meeyoung Cha, and Kyomin Jung. "Rumor detection over varying time windows." *PloS one* 12, no. 1 (2017): e0168344.

[36] Bengio, Yoshua, and Yann LeCun. "Scaling learning algorithms towards AI." *Large-scale kernel machines* 34, no. 5 (2007): 1-41.

[37] Nguyen, Tu Ngoc, Cheng Li, and Claudia Niederée. "On early-stage debunking rumors on twitter: Leveraging the wisdom of weak learners." In *Social Informatics: 9th International Conference, SocInfo 2017, Oxford, UK, September 13-15, 2017, Proceedings, Part II 9*, 2017, pp. 141-158. Springer International Publishing.

[38] Kohavi, Ron. "Glossary of terms." *Special issue on applications of machine learning and the knowledge discovery process* 30, no. 271 (1998): 127-132.