REGULAR PAPER

Combating multimodal fake news on social media: methods, datasets, and future perspective

Sakshini Hangloo[1] · **Bhavna Arora**[1]

Abstract

The growth in the use of social media platforms such as Facebook and Twitter over the past decade has significantly facilitated and improved the way people communicate with each other. However, the information that is available and shared online is not always credible. These platforms provide a fertile ground for the rapid propagation of breaking news along with other misleading information. The enormous amounts of fake news present online have the potential to trigger serious problems at an individual level and in society at large. Detecting whether the given information is fake or not is a challenging problem and the traits of social media makes the task even more complicated as it eases the generation and spread of content to the masses leading to an enormous volume of content to analyze. The multimedia nature of fake news on online platforms has not been explored fully. This survey presents a comprehensive overview of the state-of-the-art techniques for combating fake news on online media with the prime focus on deep learning (DL) techniques keeping multimodality under consideration. Apart from this, various DL frameworks, pre-trained models, and transfer learning approaches are also underlined. As till date, there are only limited multimodal datasets that are available for this task, the paper highlights various data collection strategies that can be used along with a comparative analysis of available multimodal fake news datasets. The paper also highlights and discusses various open areas and challenges in this direction.

Keywords Fake news detection · Rumor detection · Transfer learning · Pretrained models · Text embedding · Deep learning · Multimodal

1 Introduction

The propagation of information on social media is a fast-paced process, with millions of individuals participating on these sites. However, unlike traditional news sources, the trustworthiness of content on social media sites is debatable. In the last decade, there is an upsurge in the use of social media and microblogging platforms such as Facebook and Twitter. Billions of users on daily basis use these platforms to convey their opinions through messages, pictures, and videos all over the world. Government agencies also utilize these platforms to disseminate critical information

✉ Sakshini Hangloo
  sakshini.hangloo@gmail.com

1 Department of Computer Science & Information
  Technology, Central University of Jammu, Bagla (Rahya
  Suchani), District-Samba, 181143 Jammu, J&K, India

using their official Twitter handles and verified Facebook accounts, since information circulated through these platforms can reach a large population in a short amount of time. Many deceptive practices, including as propaganda and rumor, might however, deceive consumers on a daily basis. Fake news and rumors are quite common in these COVID times, and they are widely circulated, causing havoc in this difficult time. People unknowingly spreading false information is a considerably more serious problem than systematic disinformation tactics. Previously, attempts to influence public opinion were gradual, but now, rumors are targeted at naive users on social media. Once people mistakenly transmit incorrect or fraudulent content, it spreads across trusted peer-to-peer networks in all directions and as a result, in the current situation, the requirement for Fake News Detection (FND) is unavoidable.

Despite ongoing research efforts in the field of FND, ranging from comprehending the problem to building a framework to model evaluation, there is still a need to construct a reliable and efficient model. Various approaches to

fake news and rumor detection have been formulated ranging from detection methods based on content [1]–[6], propagation data [7]–[11], user profile [12]–[14], event-specific data [15]–[17], external knowledge [16, 18, 19], temporal data [20]–[24], multimodal data [4, 16, 25]–[28] etc. Source detection [29]–[34], Bot detection [35]–[37], Stance detection [38]–[43], Credibility analysis [44]–[48] are other related areas.

Earlier solutions used ML based approaches [32, 49]–[53] but suffer from the problem of manual feature engineering. With the advancement of Deep Learning (DL) based approaches for computer vision and NLP (Natural Language Processing), recent years have seen a paradigm shift from ML (Machine Learning) to DL-based fake news detection solutions. The DL models are trained using Convolution Neural Network (CNN), Recurrent Neural Network (RNN), Recursive Neural Networks (RvNN), Multi-Layer Perceptron (MLP), Generative Adversarial Networks (GANs) and many more.

It is imperative to spot fake news at the earliest before it reaches the masses. The multimodal aspect of the news article makes the content look much more credible than its counterparts. Most of the existing work focuses on text-only content or the network structure and ignores the most important aspect of the news content i.e., the visual content and consistency between text-image. Currently, due to scarcity of multimodal training datasets, transfer learning and various pre-trained models, like VGGNet [54], ResNet [55], Inception [56], Word2Vec, GloVe, BERT [57], XLNet [58] etc. are utilized for a more efficient DL-based solution. A comparative analysis of available multimodal fake news datasets is provided in Sect. 4. Although various techniques and methods have been developed in the last decade to counter fake news there are still several open research issues and challenges as mentioned in Sect. 5. By evaluating several existing strategies and identifying potential models and approaches that can be used in this area, this paper aims at contributing to the ongoing research in the field of automatic multi-modal fake news detection. Our survey seeks to give an in-depth analysis of current state-of-the-art Multimodal Fake News Detection (MFND) frameworks, with a particular focus on DL-based models. Table 1 compares various

Table 1 A relative comparison of proposed work with various related surveys

Ref.	Discussion	1	2	3	4	5	6	7
[59]	Proposes various visual and statistical features of a visual content	✔	✗	✗	✗	✗	✗	✗
[60]	Presents a comprehensive review of fake news detection techniques on social media from the data mining perspective	✔	✗	✗	✗	✔	✗	✔
[61]	Provides an overview of techniques of developing a rumor classification system consisting of detection, tracking, stance classification, and veracity classification modules	✔	✔	✗	✗	✔	✔	✔
[22]	Examined and compared the relative strength of the user, linguistic, network and temporal features of rumors over time	✔	✗	✗	✗	✗	✗	✗
[62]	provides an extensive study of automatic rumor detection on three paradigms: the hand-crafted feature-based approaches, the propagation structure-based approaches and the neural networks-based approaches	✔	✗	✗	✗	✔	✗	✔
[63]	Survey provides a review of techniques for manipulation and detection of face images including Deep-Fake methods. In particular, facial manipulation are reviewed based on following four types: attribute manipulation, face synthesis, identity swap (DeepFakes), and expression swap	✔	✗	✗	✗	✔	✗	✔
[64]	Gives an understanding of fake news creation, source identification, propagation patters, detection and containment strategies	✔	✔	✗	✗	✔	✗	✔
[65]	Presents a detailed review of state-of-the-art FND methods using DL, open issues along with future directions are also suggested	✔	✗	✔	✗	✔	✗	✔
[66]	Reviews the methods for detecting fake news from four verticals: the false information, writing style, propagation patterns, and the source credibility	✔	✗	✗	✗	✗	✗	✔
[67]	Presents an overview of the state-of-the-art fake news detection methods utilizing users, content, and context features	✔	✔	✗	✗	✔	✗	✔
[68]	Provides an overview of the different forms of fabricated content on social media ranging from text-only to multimedia content and discusses various detection techniques for the same	✔	✗	✗	✗	✗	✗	✔
[69]	proposed work explores the problem of rumors detection using textual content of social media on collected Twitter data	✔	✗	✗	✗	✗	✔	✔
[70]	Compares, reviews and provides insights into twenty-seven popular fake news detection datasets	✗	✗	✗	✗	✔	✗	✔
Present Study	The prime focus is on various deep learning approaches to fake news detection on social media keeping the multimodal data under consideration	✔	✔	✔	✔	✔	✔	✔

Notes: 1: Overview of ML/DL-based FND; 2: Open tools and initiatives; 3: DL frameworks & tools; 4: Review of MFND frameworks; 5: Datasets; 6: Data collection; 7: Open issues; Notations: ✔:Considered;✗: Not considered

existing fake news detection surveys with our survey, demonstrating that our survey not only uses more recent state-of-the-art MFND methods, but also includes widely used DL frameworks and tools, various data collection strategies from online platforms, a comparison of available datasets, and finally open issues and future scope in this direction.

1.1 Scope of the survey

This survey is motivated by the increase in the usage of social networking sites for the spread of fake news where the multimodal nature of post/tweet increases the difficulty level of the detection task. The motivations of this paper can be summed up as follows, (1) Analyzing the text-only content of an article is not sufficient to model a robust and efficient detection system. In this era of social media, it is highly imperative to consider the visual content apart from the textual context and social context to get a complete understanding of overall statistics. (2) Promising DL frameworks and transfer learning approaches are reviewed in this paper, which are advantageous for addressing the challenges and producing an improvement over the existing detection frameworks. (3) The studies performed for the detection of online fake news are diverse but these suffer from a lack of multimodal datasets. So, this study also gives an overview of some existing multimodal datasets and highlights various data collection strategies as well. This survey presents a comprehensive review of the state-of-the-art multimodal fake news detection on online media which was absent in the previous surveys.

An exhaustive comparative analysis of various research surveys is compiled in Table 1 to provide an insight into the dimensions that have not been covered previously. Different from the previous studies, in this work, the prime focus is on various deep learning approaches including the transfer learning and pre-trained models used for fake news detection on social media keeping the multimodal data under consideration. Apart from this, the paper also highlights the data collection methods and the datasets available for this task. Discussion on open areas and future scope is also provided at the end.

1.2 Contribution

The key contributions of this paper are as follows:

- The paper gives a brief introduction to fake news its related terms and provides a clear taxonomy that focuses on different methods for Fake News Detection (FND).
- The paper highlights various DL models and frameworks that are used in the literature and the benefits of using the pre-trained models and the approach of transfer learning

are highlighted. Critical analysis of different learning techniques and DL frameworks has also been presented.
- The paper discusses and reviews the various state-of-the-art Multimodal Fake News Detection (MFND) frameworks that are presented in the literature.
- The paper discusses various data collection techniques using APIs and Web crawlers in addition to a comparative analysis of various benchmark multimodal datasets.
- Finally, open issues and future recommendations are provided to combat the issue of fake news.

1.3 Methods and materials

This study is conducted using a suitable methodology to provide a complete analysis of one of the essential pillars in fake news detection, i.e., the multimodal dimension of a given article. To conduct this systematic review, various relevant articles, studies, and publications were examined. Before gathering the essential information for the conducted survey, quality checks are performed on the identified data with a focus on the most cited paper. In this work, the prime focus is on state-of-the-art research on multimodal fake news detection for assessing the authenticity of a news piece using deep learning algorithms. To obtain relevant literature, high-quality, highly cited, and reliable peer-reviewed publications, as well as conferences proceedings, are preferred. Other sources that are referred to for this study include books, technical blogs, and tutorial papers. For the search criteria, keywords like fake news detection, rumor detection, multimodal feature extraction, deep learning, and pretraining have been used. We have analyzed and acknowledged several works related to the reviewed theme of the proposed survey.

1.4 Organization

Figure 1 describes the organization of the presented survey. Section 1 presents the introduction as well as the overall scope of the paper. In Sect. 2 an introduction to fake news along with a taxonomy of fake news detection techniques has been presented. This section also disuses existing techniques and solutions to curb and combat fake news in this era of social media. Section 3 gives an overview of various DL models and transfer learning approaches that are widely in NLP, computer vision, and related fields. This section further presents a review of the DL-based state-of-the-art frameworks for Multimodal Fake news detection (MFND). Various data collection techniques and the details about the related datasets are discussed in Sect. 4. Section 5 and Sect. 6 deal with the challenges, open issues, and future direction, and the conclusion based on the survey.
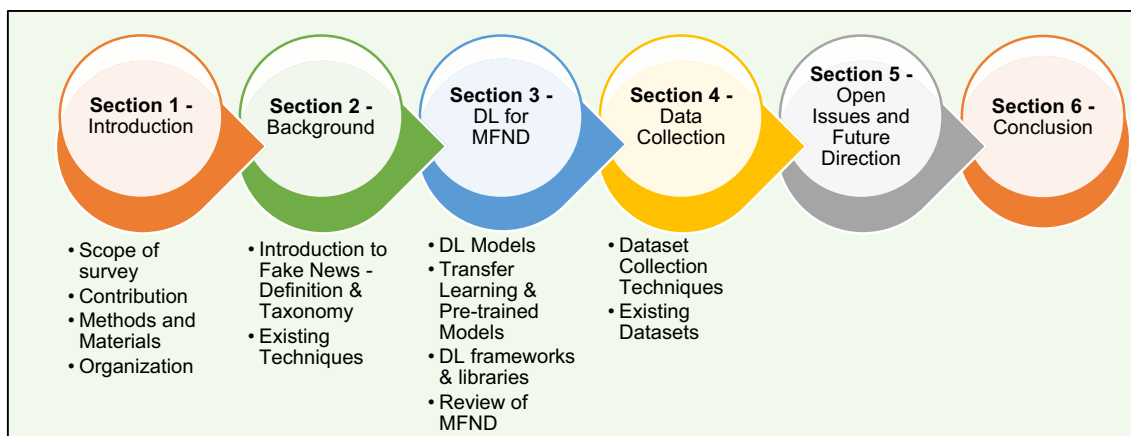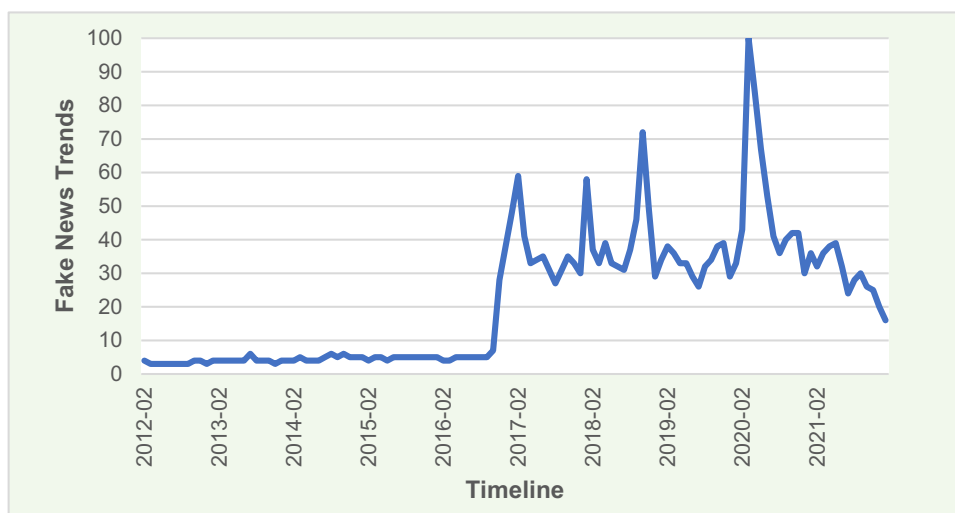
Fig. 1 Roadmap of the proposed survey

2 Background

Social media in the past decade has become the most una-voidable part of our society. With the shift of trend from Web 1.0 to Web 2.0 people have started not only to consume but also to create and spread information online. But, is this information credible? Can these be trusted? Not always. Here is a popular quote by Mark Twain "A lie can travel halfway around the world while the truth is putting on its shoes". This became quite evident in the COVID times when the internet was flooded with all kinds of information related to Government advisories, home remedies, etc.

The graph in Fig. 2 below shows a clear picture of how in the past decade the cases of fake news have increased exponentially. One of the major reasons is the rise in the use of social media and the unchecked circulation of messages on the platforms.

2.1 Introduction to fake news—definition and taxonomy

Fake News is defined as "false stories that are created and spread on the Internet to influence public opinion and appear to be true" The issue of spread of fake news is not new and has been around for centuries but with the use of social media the whole dynamics of the proliferation of information has changed and is quite different from the slow-paced traditional media. These sites provide a platform for intentional propaganda and trolling. Propaganda, fake news, satire, hoax, misinformation, rumors, disinformation, etc., are some of the terms that are used interchangeably. Some of them are discussed below (Fig. 3).

i. *Propaganda:* It is a form of news articles and stories created and disseminated by political parties to shape public opinion.
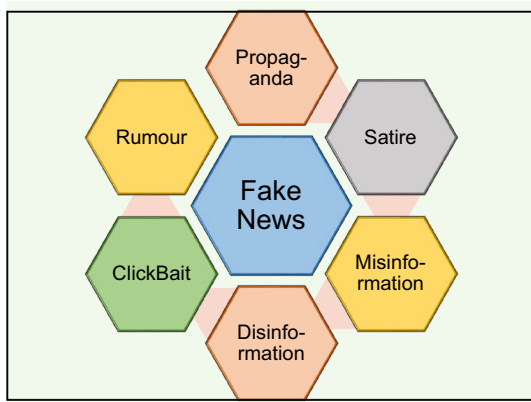
Fig. 2 Fake news trends (2012–2021) [71]

Fig. 3 Key terms related to Fake News

ii. *Misinformation:* It is purposely crafted erroneous information that is broadcast intentionally or accidentally, without regard for the true intent.

iii. *Disinformation:* It refers to a piece of misleading or partial information that is circulated to distort facts and deceive the intended audience.

iv. *Rumors and hoaxes:* These terms are often used interchangeably to refer to the purposeful fabrication of evidence that is intended to appear legitimate. They publish unsubstantiated and false allegations as true claims that are validated by established news outlets.

v. *Parody and Satire:* Humor is frequently used in parody and satire to provide news updates, and they frequently imitate mainstream news sources.

vi. *Clickbait:* Clickbait headlines are frequently used to attract readers' attention and encourage them to click, redirecting the reader to a different site. More advertisement clicks equal more money.

With the increased usage of propaganda, hoaxes, and satire alongside real news and legitimate information, even regular users find it difficult to discriminate between true and fake news. However, there are a number of online tools and IFCN-certified fact-checkers throughout the world such as BSDetector, AltNews, APF Fact Check, Hoaxy, Snopes, and PolitiFact that evaluate, rate, and debunk false news on online platforms[72].

Table 2 provides some of the Fact-checking sites and online tools that are used for debunking false news online. This table also gives an overview of the methodology and set of actions that are taken to detect and combat fake news on online platforms.

2.2 Existing detection techniques

A huge amount of content today is human-generated and most of these get published and people spread that information without even bothering about the credibility of these contents. Many technical giants are now committed to fight against the spread of fake information. Facebook in certain countries has started to work with third-party fact-checker organizations to help identify, review and rate the accuracy of information [73]. These fact-checkers are certified through the non-partisan International Fact-Checking Network (IFCN).

Figure 4 shows some of the online claims that are debunked by fact-checking organizations. Twitter on the other hand took a step forward in May 2020 to curb the misinformation around COVID-19 by introducing new labels and warning messages [74] to provide the users with additional context and information about the Tweets. This has made it easier for the users to find facts and make informed decisions about the tweets. In Jan 2021, Twitter introduced BirdWatch [75], a pilot in the US which is a community-based approach that allows people to identify Tweets that they believe are deceptive and annotate these. The pilot participants can also rate the preciseness of the notes added by other contributors.

The research community is also working tirelessly and many papers have been published to combat rumors and fake news on social media platforms. The earlier approaches [51, 76]–[81] used various ML techniques like SVM, RF, NB, etc. but with the ever-increasing amount of data on social media platforms a shift to DL approaches [11, 23, 82]–[85] can be seen with includes the use of CNN, RNN, LSTM, GAN based approaches.

Figure 5 gives a detailed taxonomy of existing fake news detection methods and techniques. Table 3 below provides a detailed classification of prominent state-of-the art ML/DL FND techniques based on the proposed taxonomy.

In the case of online social media, the rumors proliferate in a short period and hence early detection becomes very important. By exploiting the dissemination structure on social media, Liu et al. [80] offers a model for early identification of misleading news. Each news story's propagation path is treated as a multivariate time series. It employs a hybrid CNN-RNN that gathers global and local fluctuations in user attributes along the propagation path. In just 5 min after it starts spreading, the model detects fake news on Twitter and Sina Weibo with an accuracy of 85 percent and 92 percent respectively. The work proposed by Varol et al. [86] works on the early detection of promoted campaigns on online platforms. It proposed a supervised computational framework that leverages temporal patterns of the message associated with trending hashtags on Twitter to catch how the posts evolve over time and successfully classifies it as either 'promoted' or 'organic'. In addition to this, it also used network structure, sentiment, content features, and user metadata and achieves 75% AUC score for early detection, increasing to above 95% after trending.

Table 2 Fact-checking sites and online tools that are used for debunking false news online

Name	Tool/Extension	Methodology/Action
AltNews	Fact-checking website	Continuously monitors social media and mainstream media platforms for identifying incorrect information related mainly to Indian politics and entertainment, and evaluates the veracity of a claim by Manual Fact-checking
APF Fact Check	Fact-checking website	It uses many simple tools to verify online information. Fact-checking is carried out by editors and a worldwide network of journalists
BS Detector	It is an extension of Google Chrome, Mozilla	Identifies and marks fake and satirical news sites, as well as other suspected news sources. It puts a warning label to the top of potentially dangerous websites, as well as identifies fake links on Facebook and Twitter
Emergent	Fact-checking website	Emergent is a real-time rumor tracker that assesses news credibility and gives a True, False, or Unverified label
Fact-Checker	Fact-checking website	A project of The Washington Post, grades news articles from zero to four "Pinocchios" based on the factual accuracy of their content
FakerFact	A Chrome and Firefox extension	Distinguishes a fake news article from the real one and categorizes it as opinion, satire, agenda-driven, journalism, and sensationalism
InVid Verification Plugin	Use with Chrome, Firefox	A plugin to debunk fake images and videos. The tool uses reverse-image searching to debunk fake videos and also provide the users with metadata to take informed decision
PolitiFact	Fact-checking website	Tests the statements made on the Internet by political analysts and politicians and rate them. Its journalists evaluate original statements and each statement receives a "Truth-O-Meter" rating as "True", "Mostly True", "Half True", "Mostly False", "False", and "Pants on Fire"
Snopes	Fact-checking website	Conducts in-depth fact-checking research on hot issues, which are frequently picked depending on reader interest. "True", "Mostly True", "Mostly False", "False", "Unproven", "Miscaptioned", "Misattributed" are some of the annotations used to classify the content
Reverse Image Search (TinEye)	Browser extension	Can be used to see if the image has been taken from somewhere online. The tool comes with a Compare feature, which can be helpful to see how your image differs from the original one
BOOM	Fact-checking website	Manually checks the posts, debunks fake news, and prevents further spread
SurfSafe	Browser Plugin	Alerts users about misinformation by scanning images and videos on the web pages they're looking at. Performs reverse-image search by looking for the same content that appears on trusted source sites and flagging well-known doctored images
YouTube Data Viewer	A web-based video verification tool	Simple tool for extracting hidden data and metadata from YouTube videos which is particularly valuable for locating original content

Various approaches for early rumor detection have been explored. In case of the content-based approach [18, 50, 59, 90, 103], the content (text + images) within the article is considered, in contrast to the social context-based approaches [7, 13, 118]–[120] where the propagation structures, data from the user profile is considered. The content-based approaches have performed better as compared to the context-based approaches because the propagation structure and user data become available only after the news has traveled masses. Monti et al. [109] proposed a geometric deep learning approach (a non-Euclidean deep learning approach) for fake news detection on Twitter that uses a GRU-based propagation Graph Neural Network to utilize the network structure. In addition to the spreading patterns, it also uses features from the user profile, social network structure, and content. Dou et al. [121] proposes a framework, UPFD, which simultaneously captures various signals from user bias along with the news content to analyze the likelihood of user to forward a post based on his/her existing beliefs. Wu et al. [11] proposes a novel method to construct the network
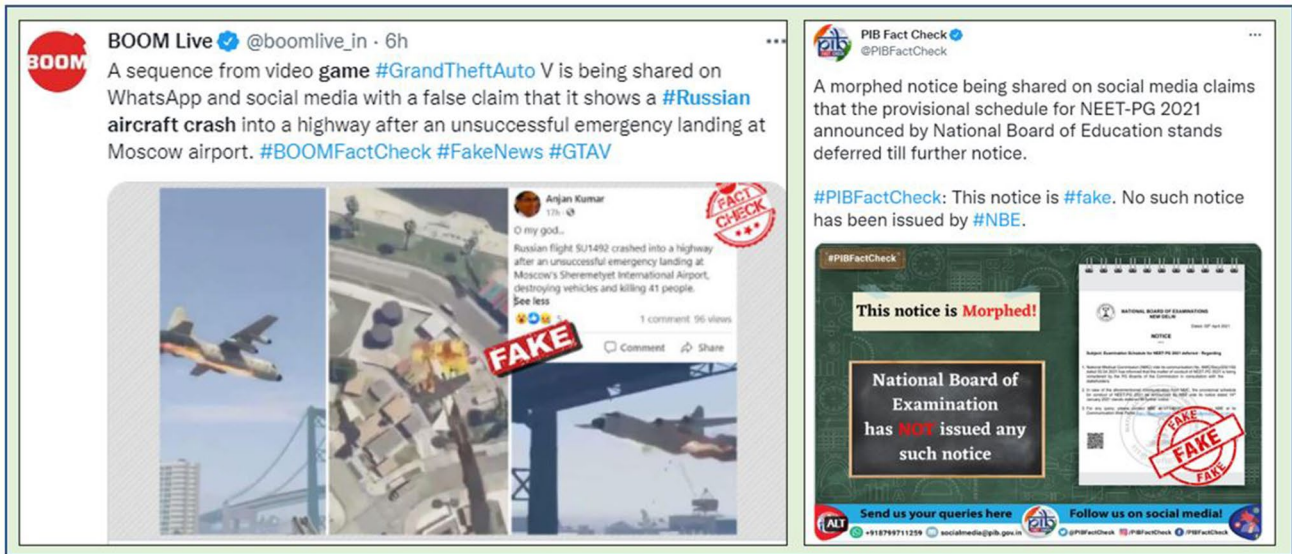
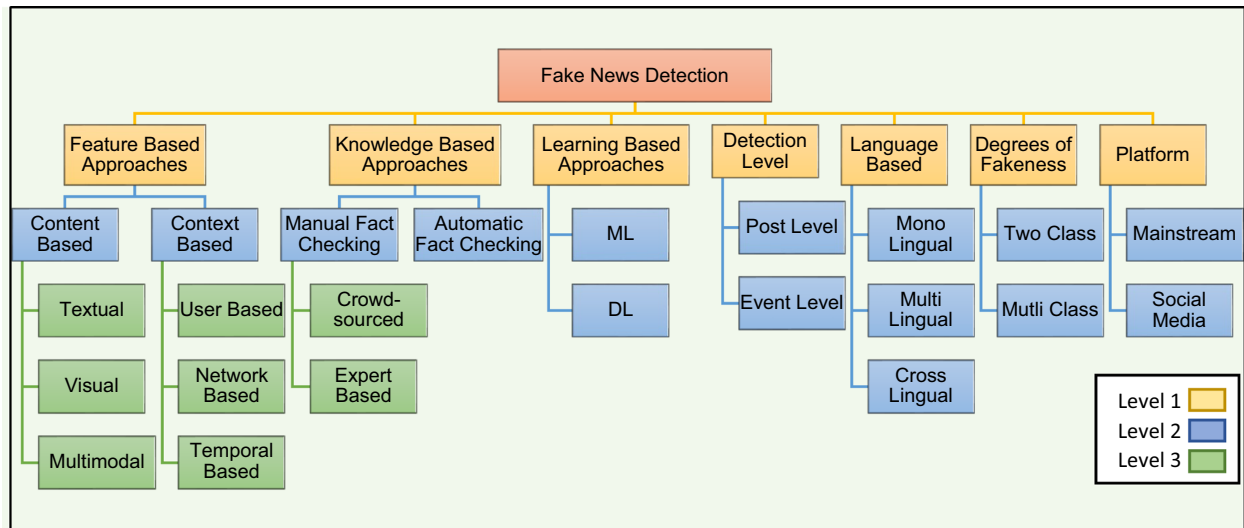**Fig. 4** False claims debunked by fact-checking organizations

Fig. 5 Taxonomy of Fake News Detection models

graph by taking the "who-replies-to-whom" structure on Twitter.

Most of the existing works are limited to using the datasets that cover domains such as politics, and entertainment. However, in a real-time scenario, a news stream typically covers various domains. Silva et al. [93] proposed a novel fake news detection framework to determine fake news from different domains by exploiting domain-specific and cross-domain knowledge in news records. To maximize domain-coverage, this research merges three datasets (PolitiFact, GossipCop and CoAID).

The content available in a post/ tweet in a microblogging site is very limited and hence detection only based on the content available in that particular post i.e. Post-level becomes difficult in such scenarios but leveraging a complete event is beneficial. [15, 16] are some state-of-the-art Event-level detection models. An event not only includes the particular post but also post-repost structure, comments, likes-dislikes, etc. and this auxiliary information makes the detection process efficient. Guo et al. [101] uses a framework of Hierarchical Networks with Social Attention (HSA-BLSTM) that aims to predict the credibility of a group of

Table 3 Classification of prominent state-of-the art ML/DL FND techniques based on the proposed taxonomy

	Level 1	Level 2	Level 3	Related work
Fake News Detection Methods	Feature Based	Content	Single-Modal	[17, 50, 59, 87]–[90]
			Multi-Modal	[4, 26, 28, 91]–[100]
		Context	Network	[10, 11, 51, 85, 101]
			User	[13, 14, 53, 102]
			Temporal	[21]–[24, 86, 103]
	Knowledge Based	Automatic	–	[16, 19, 99]
		Manual	Expert Based	[18, 40, 104, 105]
			Crowdsourced	[103, 106, 107]
	Learning Based	ML	–	[51, 76, 88, 108]
		DL	–	[11, 26, 85, 92, 96, 101, 109]–[111]
	Detection Based	Post-level	–	[18, 25, 106]
		Event Level	–	[17, 112, 113]
	Language Based	Mono-Lingual	–	[22, 25, 27, 92]
		Multi-Lingual	–	[87]
	Degree of Fakeness	Two-Class	–	[16, 25, 26, 95, 114]
		Multi-Class	–	[106, 107]
	Platform	Main-Stream	–	[18, 106, 115]
		Social Media	–	[60, 82, 86, 116, 117]

posts (reposts and comments) that constitute an event. The model uses user-based features and propagation patterns in addition to the post-based features.

Another popular approach to FND is Evidence-based or Knowledge-based Fact-Checking where the article at hand is verified with external sources. While manual fact checking is done either manually by expert journalists, editors or is sometimes crowdsourced on the other hand Automatic fact checking [18, 115] is done by employing various ML/DL techniques. A novel end-to-end graph neural model, CompareNet [19], compares the news to the knowledge base (KB) through entities for automatic fact checking.

3 Deep learning for multimodal fake news detection

With the rapid development of social media platforms, news content has transformed from traditional text-only pieces to multimedia stories with images and videos that provide more information. Multimodal news items engage more readers than standard text-only news articles because the photos and videos related to these articles make them more credible. The majority of online users are impacted by such material, unwittingly spreading false information, and becoming a part of this entire vicious network. With the growing quantity of articles on the Internet that include visual information and the widespread usage of social media networks,

the multimodal aspect is becoming increasingly important in better comprehending the overall structure of the content.

Due to exceptionally promising outcomes in several study areas, including Computer Vision and Natural Language Processing, Deep Learning has become one of the most frequently researched domains by the research community in recent years. Feature extraction, which is a laborious and time-consuming process in traditional machine learning algorithms, is done automatically by deep learning frameworks. Furthermore, these frameworks can learn hidden representations from complex inputs, both in terms of context and content, giving them an advantage in false news detection tasks when identifying relevant features for analysis is difficult.

3.1 Deep learning models

The Deep Learning techniques can be broadly classified as Discriminative and Generative models among these are Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) are the most widely implemented paradigms. As shown in Fig. 6, the DL models can be broadly classified as Discriminative and Generative models. The models are described under:

A. *Discriminative Models:* These models are supervised learning-based models and are used to solve classification and regression problems. In recent years, several discriminative models (mostly CNN, RNN) have pro-

Fig. 6 Classification of Deep Learning Models
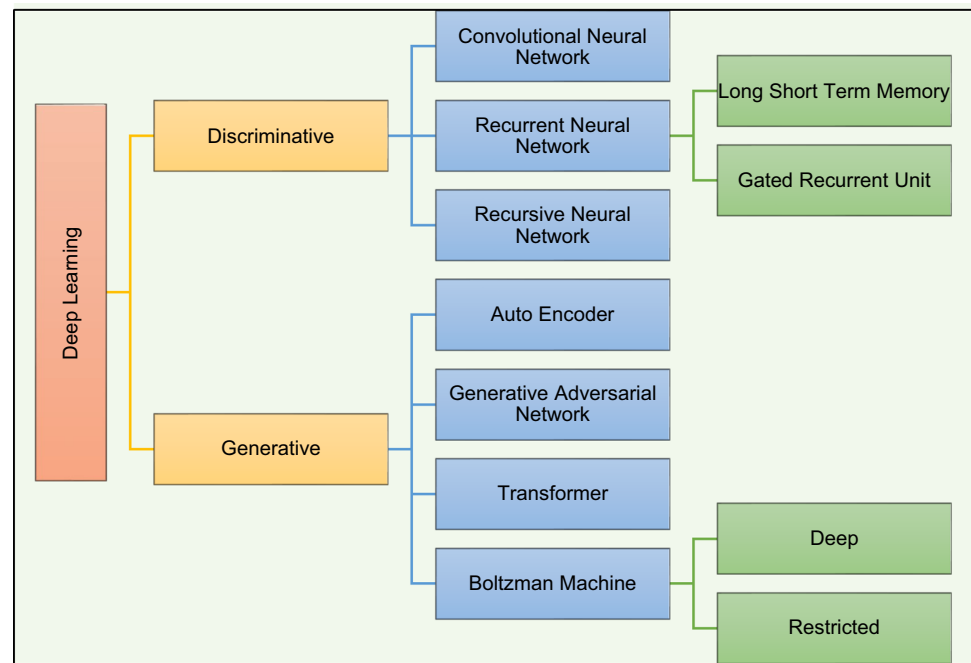
vided promising results in detecting fake news on social media platforms.

B. *Convolutional Neural Network (CNN)***:** CNN is a type of neural network that has an input layer, an output layer, and a sequence of hidden layers, in addition to this they include pooling and convolution operations for applying a variety of transformations to the given input. For the Computer Vision tasks, CNNs have been extensively explored and are regarded as state-of-the-art. CNN is also becoming increasingly popular in NLP tasks.

C. *Recurrent Neural Network (RNN):* RNNs are powerful structures that allow modeling of sequential data using loops within the neural network. Deep neural networks (RNN) have shown promising performance for learning representations in recent research. RNN is capable of capturing long-term dependency but fails to hold it as the sequence becomes longer. LSTM and GRU, the two variants of RNN, are designed to have more persistent memory and hence make capturing long-term dependencies easier. Additionally, the two networks also solve the issue of vanishing gradient problem that was encountered in traditional RNNs. LSTM includes memory cells for holding long-term dependencies in the text and includes input, output, and forget gates for memory orchestration. To further capture the contextual information, bidirectional LSTM (Bi-LSTM) and bidirectional GRU (Bi-GRU) are used to model word sequences from both directions.

D. *Recursive Neural Network (RvNN):* A recursive neural network is a deep neural network that applies same set of weights recursively over a structured input, to pro-

duce a structured prediction over variable-size input structures, by traversing a given structure in topological order. Ma et al. [122] proposes two recursive neural models based on a bottom-up and a top-down tree-structured neural networks for representing propagation structure of tweet. A tree-structured LSTM with an attention mechanism is proposed in [123] learns the correspondence between image regions and descriptive words.

Fig. 7 provides a descriptive overview of the most widely used descriptive models, namely CNN, RNN and LSTM. In addition to these, several recent works have exploited a combination of RNNs and CNNs in their models for increased efficiency. Nasir et al. [124] proposed a hybrid CNN+RNN model that can generalize across datasets and tasks. This model use CNN and LSTM to extract local features and learn long-term dependencies respectively.

B. *Generative Models:* These models are used in absence of labeled data and belong to the category of unsupervised learning. Among various generative models that have been used widely to solve a vast domain of problems, Generative Adversarial Network (GAN), Auto Encoder (AE), Transformer-based network and Boltzman Machine (BM) are mostly used and have also shown promising results in the field of FND.

i. *Auto Encoder (AE):* An autoencoder is a feed-forward neural network that regenerates the input and

Fig. 7 Discriminative Models (a) RNN (b) LSTM (c) CNN. [126]
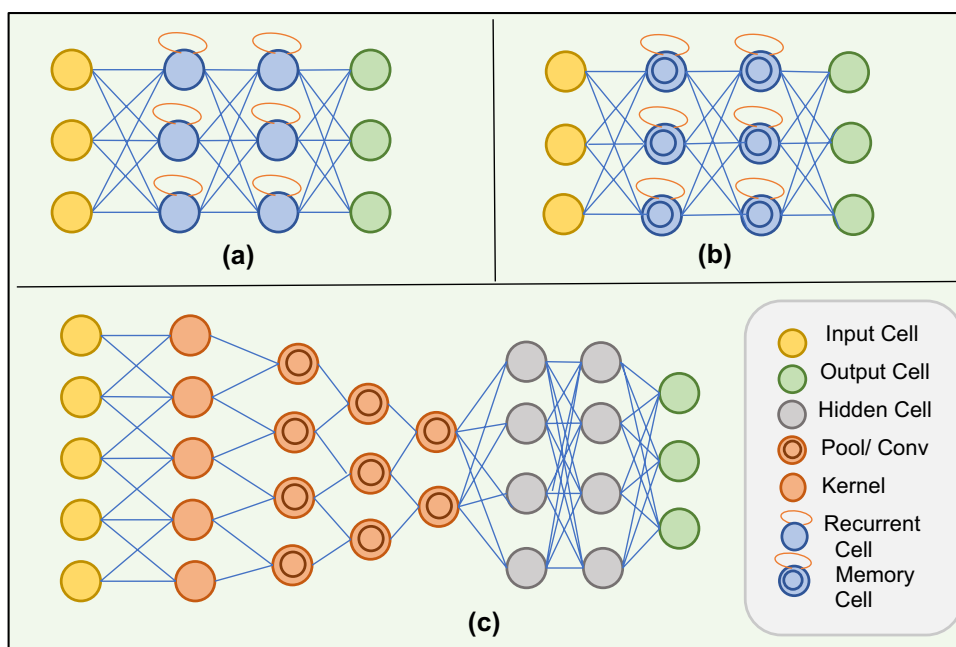
creates a compressed latent space representation. An autoencoder consists of i) an encoder that creates an intermediate representation of the input data, and ii) a decoder that regenerates the input as output. It is a self-supervised dimensionality reduction technique that generates its own labels from the training data and creates a lossy compression. A variation of AE is used in [27] that uses Multimodal Variational AE to capture the correlation between text and visual.

ii. *Generative Adversarial Network (GAN):* GAN is a class of unsupervised DL technique that consists of i) a generator that learns the generation of new sample data with the same statistics as training data, and ii) a discriminator that ties to classify the sample as true or fake. The discriminator is updated after every epoch to get better at classifying the samples, and the generator is updated to efficiently generate more believable samples. It is used widely for manipulating images and generating DeepFakes. To detect deepfake [125] detects and extracts a fingerprint that represents the Convolutional Traces (CT) left by GANs during image generation.

iii. *Transformer:* The Transformer-based networks [127] have come into existence in the past few years and have shown tremendous results for various NLP tasks. It aims to solve sequence-to-sequence tasks and handles long-range dependencies. To compute representations of its input and output, it relies on self-attention without using sequence aligned RNNs or CNNs. A transformer network consists of the encoder stack and the decoder stack that have the same number of

units. The number of encoder and decoder units act as hyperparameter. In addition to the self-attention and feed-forward layers that are present in both encoder as well as decoder, the decoders also have one more layer of Encoder-Decoder Attention layer to focus on the appropriate parts of the input sequence. BERT[57], RoBERT, ALBERT are some of the widely used transformer-based network that has been applied successfully in FND [26, 128, 129]. These networks are pre-trained and can be fine-tuned for various NLP tasks. Various transformer-based word embeddings are introduced in Sect. 3.2.

iv. *Boltzmann Machine:* It is a type of recurrent neural network where the nodes make binary decisions and are present with certain biases. Several Boltzmann Machines (BM) can be stacked together to make even more sophisticated systems such as a Deep Boltzmann Machine (DBM). These networks have more hidden layers compared to BM and have directionless connections between the nodes. The task of training is to find out how these two sets of variables are actually connected to each other. For the large unlabelled dataset, a DBM incorporates a Markov random field for layer-wise pretraining and then provides feedback to previous layers. Restricted Boltzmann Machine (RBM) shares a similar idea as Encoder-Decoder, but it uses stochastic units with particular distribution instead of deterministic distribution. RBM plays an important role in dimensionality reduction, classification, regression and many more which is used for feature selection and feature extraction. [130] pre-

Fig. 8 Generative Models (a) Auto Encoder (b) Generative Adversarial Network (c) Transformer. [131]
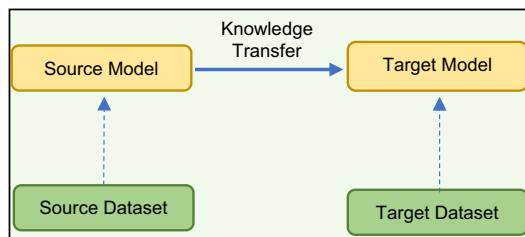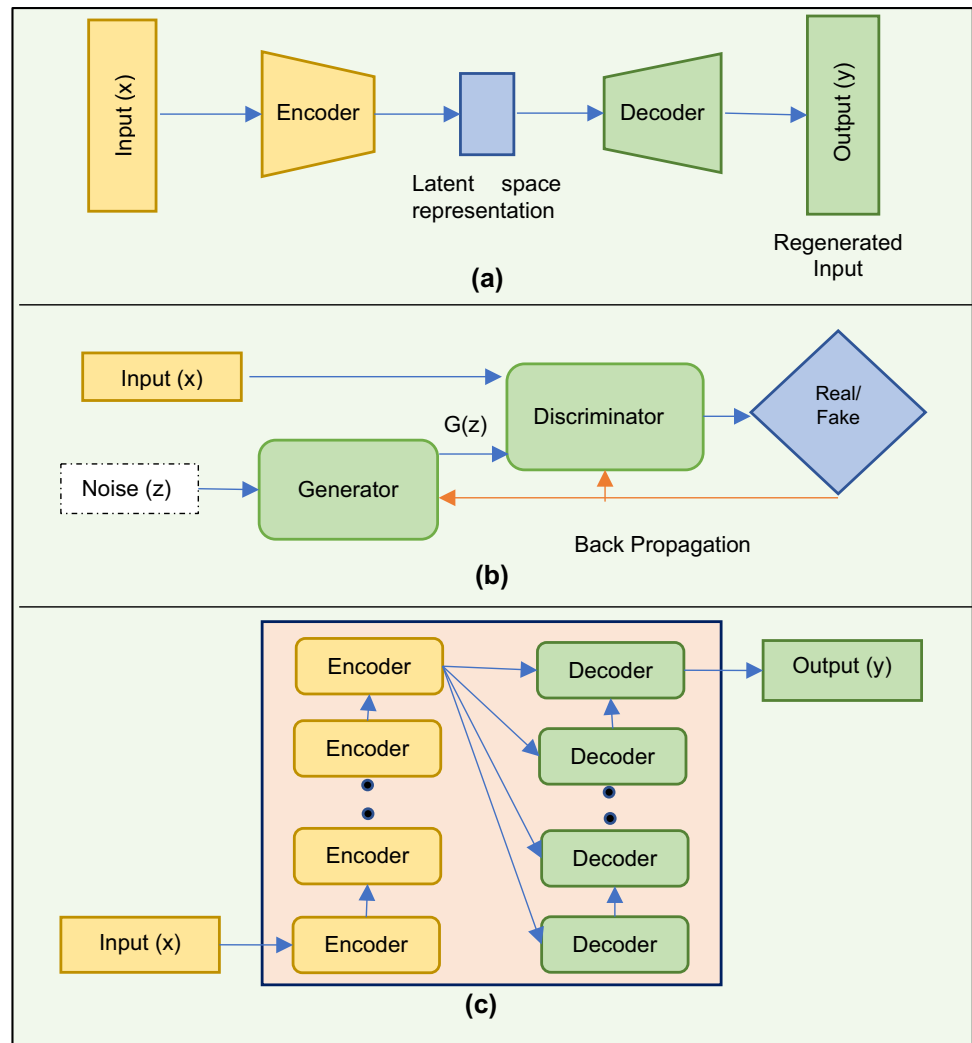

Fig. 9 Transfer learning

sents a Deep Boltzmann Machine based multimodal deep learning model for fake news detection.

Among various generative models (Fig. 8) that have been used widely to solve a vast domain of problems, Generative Adversarial Network (GAN), Auto Encoder (AE), Transformer-based network are widely used and have also shown promising results in the field of FND.

3.2 Transfer learning and pre-trained models

Transfer learning is a machine learning technique that leverages and applies the weights of a model that is trained on one task to some other related tasks with different datasets for improving efficiency as shown in Fig. 9. This approach can be applied in one of the two ways (i) using the pre-trained model as feature extraction, or (ii) fine-tuning a part of the model. The first variant is directly applied without changing the weights of the pre-trained model e.g., using Word2Vec in the NLP task. For the second variant, fine-tuning is done by trial-and-error experiments. For two different tasks around 50% of fine-tuning can be considered and if the tasks are very similar fine-tuning of the last few layers can be done. The nature and amount of fine-tuning needed take time and effort to explore depending on the nature of the task. Using Transfer Learning is beneficial when the target dataset is significantly smaller than the source dataset, as the model can learn features even with less training data without overfitting. Also, such a model exhibits better efficiency and

requires a lesser training time than a custom-made model. As a result, this idea finds vast usage in the field of Computer Vision and Natural Language Processing.

A. *Transfer learning with image data:* Earlier with smaller datasets in the picture, the ML models were effectively used for computer vision tasks. But with the increase in the amount of data available we have seen a shift toward DL-based models that have proven to be very efficient in handling recognition tasks using these huge datasets. Effectively, there are many models available for Computer Vision tasks, this section gives a brief overview of some of these models. Table 4 provides a comparison of pre-trained image models. Many of these models like the VGG, ResNet-50 Inception V3, and Xception models are pre-trained on the ImageNet (contains 1.28 million images divided among 1,000 classes) for object detection tasks.

B. *AlexNET***:** Convolutional Neural Networks (CNNs) have typically been the model of choice for object recognition since they are powerful, easy to train, and control. AlexNET [132], a Deep convolutional network, consists of eight layers (having five convolutional layers and three fully-connected layers), and Rectified Linear Unit (ReLU) function as it does not suffer from the issue of vanishing gradient. It also combats overfitting by employing drop-out layers, in which a link is deleted with a probability of $p = 0.5$ during training. Apart from this, it allows a multi-GPU training environment that helps to train a larger model and even reduces the training time. AlexNet is a sophisticated model that can achieve high accuracies even on huge datasets however, its performance is compromised if any of the convolutional layers are removed.

C. *VGG Model*: The VGGNet (Visual Geometry Group Network) is a CNN model with a multilayered operation and is pre-trained on the ImageNet dataset. VGG [54] is available in two variants with 16 and 19 weight layers namely VGG-16 and VGG-19 respectively. These models are substantially deeper than the previous models and are built by stacking convolution layers but the model's depth is limited because of an issue called diminishing gradient which makes the training process difficult. To reduce the number of parameters in these very deep networks, a 3×3 convolution filter is used in all layers with stride set to 1. The model uses fixed 3×3 sized kernels that can reproduce all of Alexnet's variable-size convolutional kernels (11×11, 5×5, 3×3).

D. *GoogLeNet***:** GoogLeNet (Inception V1) [56] was the first version of the GoogLeNet architecture, which was further developed as Inception V2 and Inception V3. While larger kernels are preferable for global features distributed over a vast area of the image, while smaller kernels detect area-specific features that are scattered across the image frame efficiently, and hence choosing a set kernel size is a challenging task. To resolve the problem of recognition of a variable-sized feature, Inception employs kernels of varied sizes. Instead of increasing the number of layers in the model, it expands it by including several kernels of varied sizes within the same layer. The Xception architecture is a modification of the Inception architecture that uses depth-wise separable convolutions instead of the usual Inception modules.

E. *ResNET:* To address the issue of diminishing gradient in VGG models, the ResNET (Residual Network) [55] model was developed. The primary idea is to use shortcut connections to build residual blocks to bypass blocks of convolutional layers. CNN models can get deeper and deeper using Resnet models. There are many variations for Resnet models but Resnet50 and ResNet101 are used mostly.

B. *Transfer learning with Language data:* Word Embeddings use vector representations of words to encode the relationships between them. The pre-trained word vector is related to the meaning of the word and is one of the most effective ways to represent a text since it intends to learn both the syntactic and semantic meaning of the given word. Figure 10 provides the classification of Word Embeddings that is broadly divided into four categories depending upon (i) whether these can preserve the context or not as Context-independent and Context-dependent word embeddings, (ii) whether the underlying architecture of the model is RNN-based or Transformer based, (iii) the level at which the encoding is produced and (iv) whether the underlying task is supervised or unsupervised. An overview and comparative analysis of

Table 4 Pretrained Image Models

Network	Author(s), Year	Salient Features	Parameters	FLOP	Top 5 Accuracy
AlexNET	Krizhevsky et al. (2012)	Deeper	62 M	1.5B	84.70%
VGGNet	Simonyan et al. (2014)	Fixed-size kernel	138 M	19.6B	92.30%
Inception	Szegedy et al. (2014)	Wider parallel kernel	6.4 M	2B	93.30%
ResNET	He et al. (2015)	Shortcut connections	60.3 M	11B	95.51%

the different types of word embedding techniques categorized under Traditional, Static, and Contextualized word embeddings is provided in [133]. Pretrained Word Embedding is a type of Transfer Learning approach where embeddings learned in one task on a larger dataset are utilized to solve another similar job. These are highly useful in NLP tasks in a scenario where the training data is sparse and number of trainable parameters are quite large. [134] studies the utility of employing pretrained word embeddings in Neural Machine Translation (NMT) from a number of perspectives.

Some of the widely used text embeddings are discussed below (Fig. 10):

i. *One-hot vector representation:* It is one of the first and simplest word embeddings. It represents every word as an $\mathbf{R}^{|V| \times 1}$ vector with all 0 s and only a single 1 at the index of that particular word in the sorted English language, where |V| represents the vocabulary. Though one-hot vectors are easy to construct, but these are not a good choice to represent a large corpus of words as it does not capture the similarity between the words in the corpus.

ii. *Word2Vec:* Word2Vec is developed by Google and is trained on Google News Dataset. It is one of the widely used pre-trained word embeddings, takes text corpus as input, and generates word vectors as output. It is a Shallow Neural Network architecture that uses only one hidden layer in its feed-forward network. It learns vector representations of words after first constructing a vocabulary from the training text input. Distance tools like cosine similarity are used for finding the nearest words for a user-specified term. Depending on how the embeddings are learned, the Word2Vec model can be categorized into one of two approaches: Continuous Bag-of-Words (CBOW) model that learns the target word from the surrounding words, and the Skip-gram model that learns the surrounding words given the target word.

iii. *GloVe* is an unsupervised learning technique that generates word vector representations by leveraging the relationship between the words from Global Statistics. The training produces linear substructures of the word vector space, which are based on aggregated global word-word co-occurrence statistics from a corpus. The basic premise of the model is that ratios of word-word co-occurrence probabilities can contain some form of meaning. A co-occurrence matrix shows the frequency of occurrence of a pair of words together.

iv. *BERT* (Bidirectional Encoder Representations from Transformer) [57] has an advancement over Word2Vec and generates dynamic word representations based on the context in which the word is being used rather than generating fixed representation like Word2Vec. A polysemy word e.g., bank, can have multiple embeddings depending on the context in which the word is being used. This has brought context-depend-

Fig. 10 Taxonomy of Word Embeddings
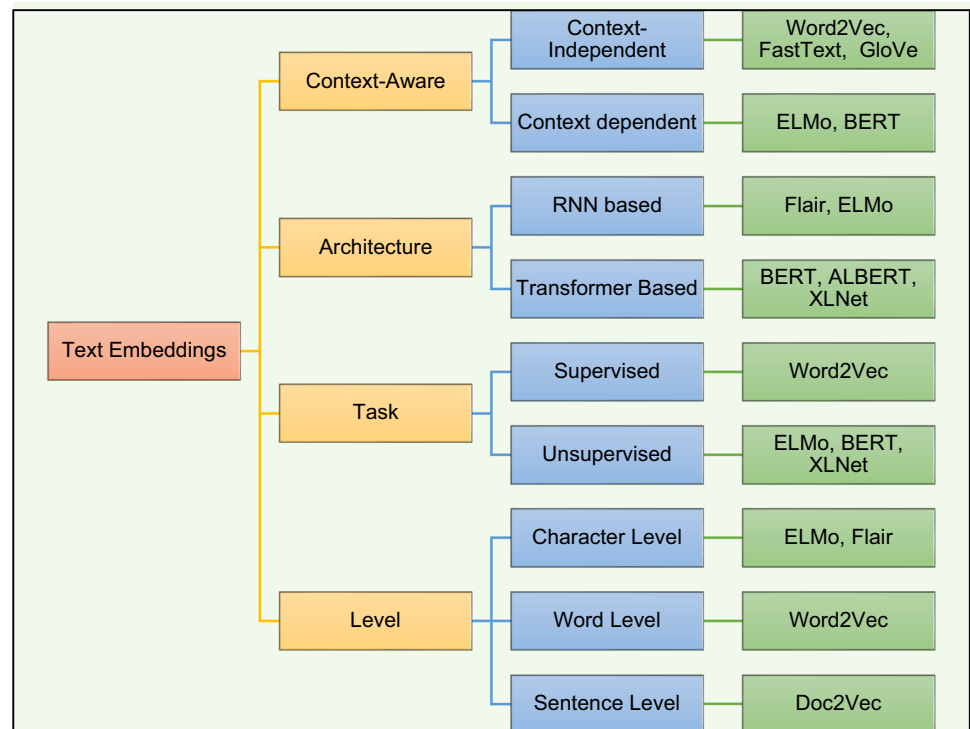
ent embeddings into the mainstream in present times. BERT is a pre-trained bidirectional transformer-based contextualized word embedding that can be fine-tuned as per the need. Since its introduction, many variants like RoBERTa (Robustly Optimized BERT Approach) [135] and Albert (A Lite BERT) [136] have been introduced to further enhance the state-of-the-art in language representations.

v. *ELMo:* Unlike traditional word-level embeddings like Word2Vec and GLoVe that have the same vector for a given word in the vocabulary, the same word can have distinct word vectors under varied contexts in the case of ELMo representation like the representation of BERT. The ELMo vector assigned to a word is a function of the complete input sentence containing that word.

vi. *XLNet* learns unsupervised language representations based on a novel generalized permutation language modeling aim. It fuses the bidirectional facility of BERT with the autoregressive technology of Transformer-XL.

Table 5 provides a comparison between various embeddings and highlights the advantages and disadvantages of each one of them.

3.3 Deep learning frameworks and libraries

Over the past few years, various detection methods have been proposed for solving the issue of fake news and rumors on online social media. Researchers are constantly working in these domains to find effective solutions and techniques. Deep learning is one of the several techniques that has become increasingly popular in solving problems in various domains. Neural networks such as CNN, RNN, LSTM are becoming increasingly popular. Although, using deep learning techniques complex tasks are performed easily compared to the machine learning counterparts but, successfully building and deploying them is a challenging task. Training a deep learning model takes a little longer when compared with traditional models but testing can be done rapidly. The deep learning frameworks are developed with an intention to accelerate and simplify the processing of the model. These frameworks combine the implementation of contemporary DL algorithms, optimization techniques, with infrastructure support. Figure 11 gives an overview of various DL/ML tools that are widely used to simplify the research problems.

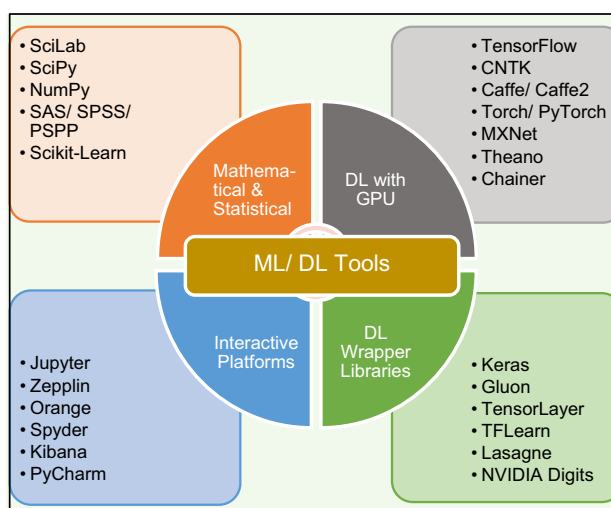Some of the ML/DL frameworks and libraries are discussed below:

Fig. 11 Overview of ML/DL frameworks and libraries

Table 5 Comparison of various Text Embeddings

Embedding	Advantage	Weakness
Word2Vec	Consume much less space than one-hot encoded vectors Maintain semantic representation of word Capable of capturing multiple degrees of similarity between words using simple vector arithmetic	Can't handle OOV words No shared representation is used at subword level Scaling to new languages requires separate embedding matrices
GloVe	Can handle Out-Of-Vocabulary words	Gives random vectors to OOV words which confuses the model in long run
BERT	Creates contextualized vectors Learns representations at a "subword" (also called WordPieces) level	Computationally intensive Neglects dependency present between the masked positions Suffers from the pretrain-finetune inconsistency
ELMo	Generates contextualized word embeddings Can handle Out-Of-Vocabulary words	Complex Bi-LSTM structure makes train and embedding generation very slow Representing long-term context dependencies becomes difficult
XLNet	Provides autoregressive pretraining Enables bidirectional learning by maximizing the expected likelihood over all permutations of the factorization order	XLNet is pre-trained to capture long-term dependencies but can underperform on short sequences XLNet is generally more resource-intensive and takes longer to train and to infer compared to BERT

i. *TensorFlow:* TensorFlow is an open-source, end-to-end framework for deep learning published under the Apache 2.0 license. It is designed for large-scale distributed training and testing that can run on a single CPU system, GPUs, mobile devices as well as on distributed systems. It provides a large and flexible ecosystem of tools and libraries that allows developers to quickly construct and deploy ML applications. It uses data flow graphs to perform numerical computations. Model building and training with TensorFlow employs high-level Keras API and is quite simple as it offers multiple levels of abstraction, allowing researchers to choose the level of abstraction that best suits their needs.

ii. *Torch and PyTorch:* Torch is one of the oldest frameworks which provides a wide range of algorithms for deep machine learning. It provides a multi-GPU environment. Torch is used for signal processing, parallel processing, computer vision, NLP etc. Pytorch is an open-source Python version of Torch, which was developed by Facebook in 2017 and is released under the Modified BSD license. PyTorch is a library for processing tensors. It is an alternative to NumPy to use the power of GPUs and other accelerators. It contains many pre-trained models and supports data parallelism. It is one of the widely used Machine learning libraries, along with TensorFlow and Keras. It is particularly useful for small projects and prototyping.

iii. *Caffe and Caffe2:* Caffe is a general deep learning framework that is based on C + +launched with speed, and modularity in mind. It was developed by Berkeley AI Research (BAIR). Caffe is designed primarily for speed and includes support for GPU as well as Nvidia's Compute Unified Device Architecture (CUDA). It performs efficiently on image datasets but doesn't produce similar results with sequence modeling. Caffe2, open-sourced by Facebook since April 2017, is lightweight and is aimed towards working on (relatively) computationally constrained platforms like mobile phones. Caffe2 is now a part of PyTorch.

iv. *MXNet:* Apache MXNet is a flexible and efficient deep learning library suited for research prototyping and production. It works on multiple GPUs with fast context switching. MXNet contains various tools and libraries that enable tasks involving computer vision, NLP, time series etc. GluonCV and GlutonNLP are libraries for computer vision and NLP modeling, respectively. The Parameter Server and Horovod support enable scalable distributed training and performance optimization in research and production.

v. *Theano:* Theano is a python-based deep learning library developed by Yoshua Bengio at Université de Montréal in 2007. The latest version of Theano, 1.0.5, is Python 3.9 compatible. Theano is built on top of NumPy and helps to easily define, evaluate, and optimize mathematical operations involving multi-dimensional arrays. It can be run on the CPU or GPU, providing smooth and efficient operation. It offers its users with extensive unit-testing which aids in code debugging. Keras, Lasagne, and Blocks are open-source deep libraries built on top of Theano.

vi. *Chainer:* Chainer is a robust and flexible deep learning framework that supports a wide range of deep networks (RNN, CNN, RvNN etc.). It supports CUDA computation and uses CuPy to leverage a GPU computation. Parallelization with multiple GPUs is also possible. Code debugging with Chainer is quite easy. It provides two DL libraries i) ChainerRL that implements a variety of deep reinforcement algorithms, and ii) ChainerCV which is a Library for Deep Learning in Computer Vision.

vii. *Computational Network Toolkit (CNTK):* The Microsoft Cognitive Toolkit (CNTK) is an open-source toolkit, since April 2015, for providing commercial-grade distributed deep learning services. It is one of the first DL toolkits that supports the Open Neural Network Exchange (ONNX) format for shared optimization and interoperability. The newest release of CNTK, 2.7., supports ONNX v1.0. With CNTK neural networks are represented as a series of computational steps using a directed graph. It allows the user to easily develop and deploy various NN models such as DNN, CNN, RNN etc. CNTK can either be included as a library in Python, and C + +code, or can be used as a standalone ML/DL tool through BrainScript (its own model description language).

viii. *Keras:* Keras is Python wrapper library for DL written in Python, and runs on top of the ML/DL platforms like TensorFlow, CNTK, Theano, MXNet and Deeplearning4j. Given the underlying frameworks, it runs on Python 2.7 to 3.6 on both GPU as well as on CPU. It was launched with a prime focus on facilitating fast experimentation and is available under the MIT license.

ix. *TFLearn:* TFlearn is a modular and transparent deep learning library built on top of Tensorflow and facilitates and speed-up experimentations using multiple CPU/GPU environment. All functions are built over tensors and can be used independently of TFLearn.

x. *TensorLayer*: TensorLayer is a deep learning and reinforcement learning library built on top of TensorFlow framework. Other TensorFlow libraries including Keras and TFLearn hide many powerful features of TensorFlow and provide only limited support for building and training customized models.

Table 6 shows some popular DL frameworks (such as Keras, Caffe, PyTorch, TensorFlow, etc.) along with their

Table 6 Comparison of popular Deep Learning Frameworks

Software	Platform	Written in	Interface	Open MP support	Open CL support	CUDA support	RNN	CNN	Has pre-trained Models
TensorFlow	Windows, Linux, macOS, Android	Python, C++, CUDA	C/C++, R, Python (Keras), Java, JavaScript	✗	via SYCL support	✔	✔	✔	✔
PyTorch	Windows, Linux, macOS, Android	C/C++, Python, CUDA	Python, C++	✔	Via separately maintained package	✔	✔	✔	✔
Caffee	Linux, macOS, Windows	C++	Python, C++ MATLAB,	✔	Under development	✔	✔	✔	✔
Theano	Cross-platform	Python	Python (Keras)	✔	Under development	✔	✔	✔	Through Lasagne's model zoo
Chainer	Linux, macOS	Python	Python	✗	✗	✔	✔	✔	✔
MXNet	Linux, AWS macOS, iOS, Windows, Android, JavaScript	Small C++ core library	C++, Python, MATLAB, JavaScript, Scala, Perl, R	✔	On roadmap	✔	✔	✔	✔
Microsoft Cognitive Toolkit (CNTK)	Linux, Windows, macOS (via Docker on roadmap)	C++	Python (Keras), C++ Command Line	✔	✗	✔	✔	✔	✔

comparative analysis. These frameworks and libraries are implemented in Python, are task-specific and allow researchers to develop tools by offering a better level of abstraction

Several machine learning and deep learning frameworks have emerged in the last decade, but their open-source implementations appear to be the most promising for several reasons: (i) openly available source codes, (ii) a large community of developers, and, as a result, a vast number of applications that demonstrate and validate the maturity of these frameworks.

3.4 Review of state-of-the-art multimodal frameworks

With the rapid expansion of social media platforms, news content has evolved from traditional text-only articles to multimedia articles involving images and videos that carry richer information. Multimodal articles have the power to engage more readers as compared to the traditional text-only articles as the images and videos attached to these articles make them more believable. Most of the online users get affected by such information, unknowingly spread the misinformation, and become a part of this whole vicious network.

Traditionally, the great majority of methods for identifying false news have focused solely on textual content analysis and have relied on hand-crafted textual features to do so. However, with the growing quantity of articles on the Internet that include visual information and the widespread usage of social media networks, multimodal aspects are becoming increasingly important in understanding the overall intent of the content in a better way.

Given the contents of a news claim $\mathbb{C}$ with its text set T and image set I, the task of multimodal fake news detector is to determine whether the given claim can be considered as true or fake, i.e., to learn a prediction function $\mathcal{F}(\mathbb{C}) \rightarrow 0, 1$ satisfying:

$$\mathcal{F}(\mathbb{C}) = \begin{cases} 1, & \text{if } \mathbb{C} \text{ is confirmed to be fake} \\ 0, & \text{otherwise} \end{cases}$$

The following figure, Fig. 12, presents a general framework that depicts various channels present in a multimodal fake news detection (MFND) framework. The framework illustrates how the features are extracted individually and then merged to detect the credibility of the claim.

Some multimodal FND frameworks, apart from fusing textual and image data, also evaluate the similarity between
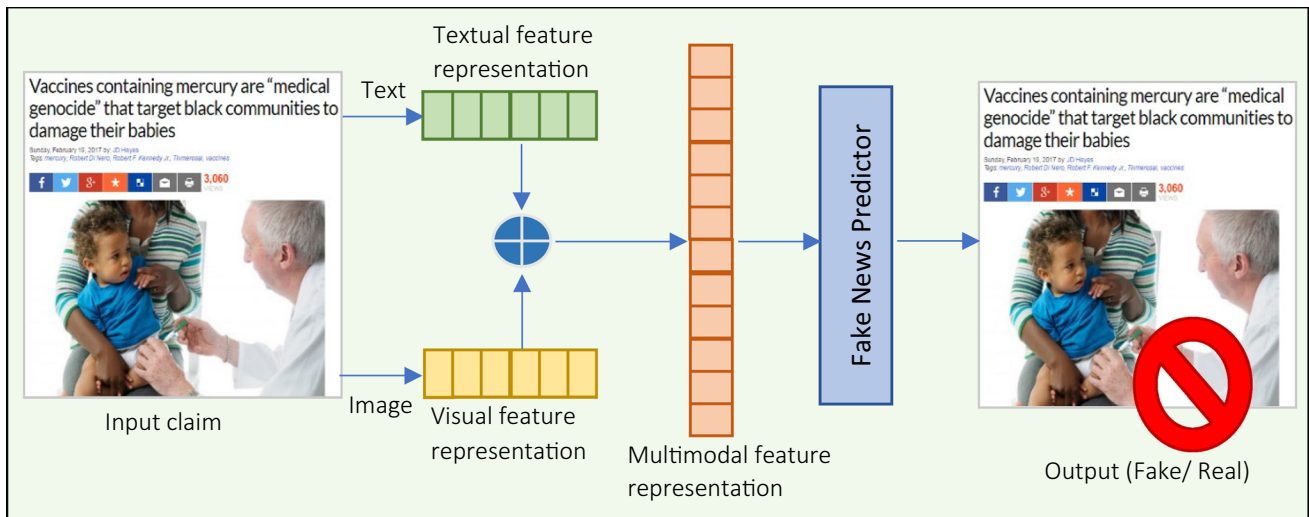
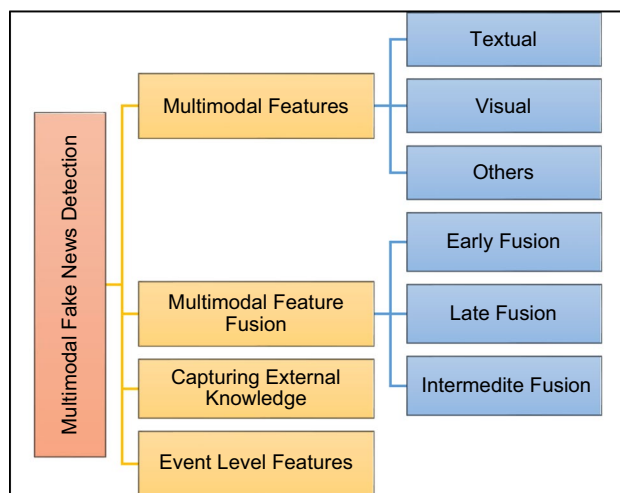**Fig. 12** A general framework for Multimodal Fake News Detection

Fig. 13 Various components of Multimodal Fake News Detection

the two [97], or have used external knowledge and event-specific information to check the credibility of a given news article. Others have also used social context including user data[25], propagation structure, sentiments[98], and other auxiliary information to effectively combat and detect fake news.

Figure 13 depicts the taxonomy of a MFND framework by focusing on the various techniques used in processing the individual channels.

A. *Multimodal Features:* With the increasing use of social media, a shift from all text to a multimodal news article can be seen. Now, the news articles comprise images and videos along with the text. The models for MFND have used two different channels for handling the text and image data. The section below summarizes how different models have exploited various techniques for the same.

i. *Text channel:* To capture the context from textual data, the researchers have used various word embeddings and pre-trained model, as discussed in Sect. 3.2. Word2Vec is one of the popular word embedding that is used by [4, 25, 27, 91]. But as word2vec can't handle out-of-vocabulary words, researchers have exploited Glove, BERT, XLNet and other embeddings instead [26, 28, 92, 95, 96, 98, 110, 129]. FND-SCTI [4] considers the hierarchical document structure and uses Bi-LSTM at both word-level (from word to sentence) and sentence-level (from sentence to document) to capture the long-term dependencies in the text. Ying et al. [92] proposed a multi-level encoding network to capture the multi-level semantics in the text. The model KMAGCN [99] proposed by Qian et al. captures the non-consecutive and long-range semantic relations of the post by modeling it as a graph rather than a word sequence and proposes a novel adaptive graph convolutional network handle the variability in the graph data.

ii. *Visual channel:* The Visual sub-network uses the article's visual information as input to generate the post's weighted visual features. The image is initially resized (usually to 224×224 pixel size), after which it is placed into a pre-trained model to extract the image features. The visual channel in the framework captures the manipulation in image data using pre-trained models. VGG-19 is the most widely used model, apart from this VGG-16, ResNet50 are also utilized. [97] uses image2sentence to represent news images by

generating image captions. [110] uses Image forensic techniques—Noise Variance Inconsistency (NVI) and Error Level Analysis (ELA) for the identification of manipulated images. [26, 96] uses the bottom-up attention pre-trained ResNet50 model to extract region features for every image attached with the article. As an article sometimes comes with multiple images, Giachanou et al. in [94] propose visual features that are extracted from multiple images.

iii. Apart from the visual and textual channels some of the researchers have also focused on other aspects of a social media post that might be helpful in detecting fake news. Cui et al. [98] proposed an end-to-end deep framework, named SAME that incorporates user sentiment extracted from users' comments (with VADER -sentiment prediction tool) with the multimodal data. Experiments on PolitiFact and GossipCop shows F1 score of 77% (approx.) and 80%(approx.) which is better than the baseline methods. User profile, network, and propagation features are another set of vectors that are highly exploited [25].

B. *Multimodal Feature Fusion:* Dealing with multimedia data comes with an intrinsic challenge of handling the data of varied modalities while keeping intact the correlation between them. In a multimodal social media post, finding the correlation between text and image is an important step in identifying a fake post. There are three techniques (Fig. 14) that are widely used for multimodal data fusion.

i. *Early fusion/ Data-level Fusion* first fuses the multimodal features and then applies the classifier on the combined representation; The data-level fusion of multimodal features starts with feature extraction of unimodal features and after analysis of the different unimodal feature vectors these features are combined into a single representation. With early fusion as the features are integrated from the start, a true multimedia feature representation is extracted. There are various methods like principal component analysis (PCA), canonical correlation analysis (CCA), independent vector analysis (IVA) and independent component analysis (ICA) which are used to accomplish this task [137]. One of the main disadvantages of this approach is the complexity to fuse the features into a common joint representation. Also, rigorous filtering of data is needed to make a common ground before fusion which poses a challenge if the dataset available is already limited in number.

ii. *Late fusion/ Decision-level Fusion* combines the results obtained from different classifiers trained on different modalities; Late fusion also starts with extracting the unimodal features. But in contrast to early fusion, the late fusion approach learns semantic concepts separately from each unimodal channel and different models are available to determine the optimal approach to combine each of the independently trained models. It is based on the ensemble classifier technique. This method gives the flexibility to concatenate the input data streams that significantly varied in terms of the number of dimensionality and sampling rate. Fusing the features at the decision-level is expensiveness in terms of the learning effort as separate models are employed for each modality. Furthermore, the fused representation requires an additional layer of the learning stage. Another disadvantage is the potential loss of correlation in fused feature space.

iii. *Intermediate Fusion* allows the model to learn a joint representation of modalities by fusing different modalities representations into a single hidden layer. Intermediate fusion changes input data into a higher level of representation (features) through multiple layers and allows data fusion at different stages of model training. Each individual layer uses various linear and non-linear functions to learn specific features and generates a new representation of the original input data.

Several research works in the literature have come up with various techniques, [24] presents a neuron-level attention mechanism for aligning visual features with a joint representation of text and social context, and as a result, greater weights are assigned to visual neurons with semantic meanings related to the word. FND-SCTI proposed in [4] uses a hierarchical attention mechanism to put an emphasis on the important parts of the news article. Giachanou et al. in [94] propose the use of cosine similarity to find the image-text similarity between the title and image tags embeddings. To preserve semantic relevance and representation consistency across different modalities [98] uses an adversarial mechanism. To filter out the noise and highlight the image regions that are strongly related to the target word, [16] uses a word-guided visual attention module. To learn complementary inter-dependencies among textual and visual features [26, 92, 96, 99, 129] uses multiple co-attention layers, hierarchical multi-modal contextual attention network, feature-level attention mechanism, blended attention module and multimodal cross-attention network respectively. The proposed the Crossmodal Attention Residual Network (CARN) in [111] can selectively extract information pertaining to a target modality from another source modality while preserving the target modality's unique information. Another model SAFE [97], jointly learns the text and image features and

**Fig. 14** General schemes for multimodal fusion (a) Early Fusion, (b) Late Fusion, (c) Intermediate Fusion
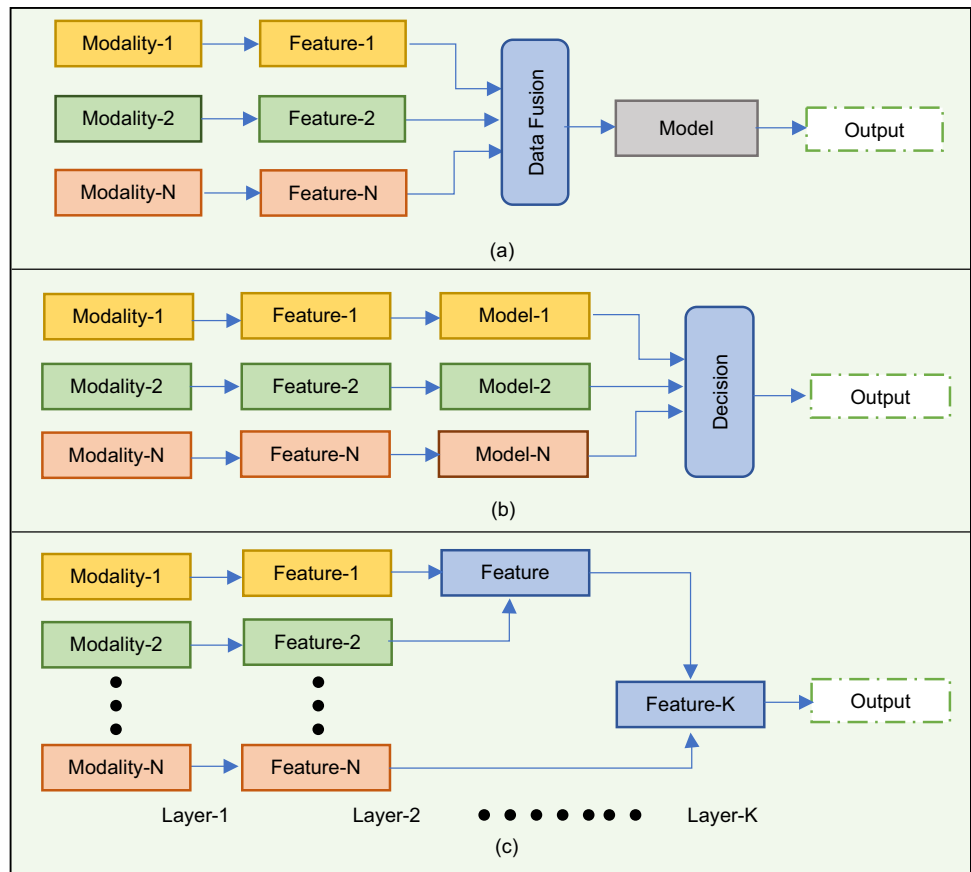
also learns the similarity between them to evaluate whether the news is credible or not (Fig. 14).

III. *Capturing External Knowledge channel:* The Knowledge module tries to capture background knowledge from a real-world knowledge graph to supplement

Fig. 15 Illustration of Knowledge Distillation process [16]
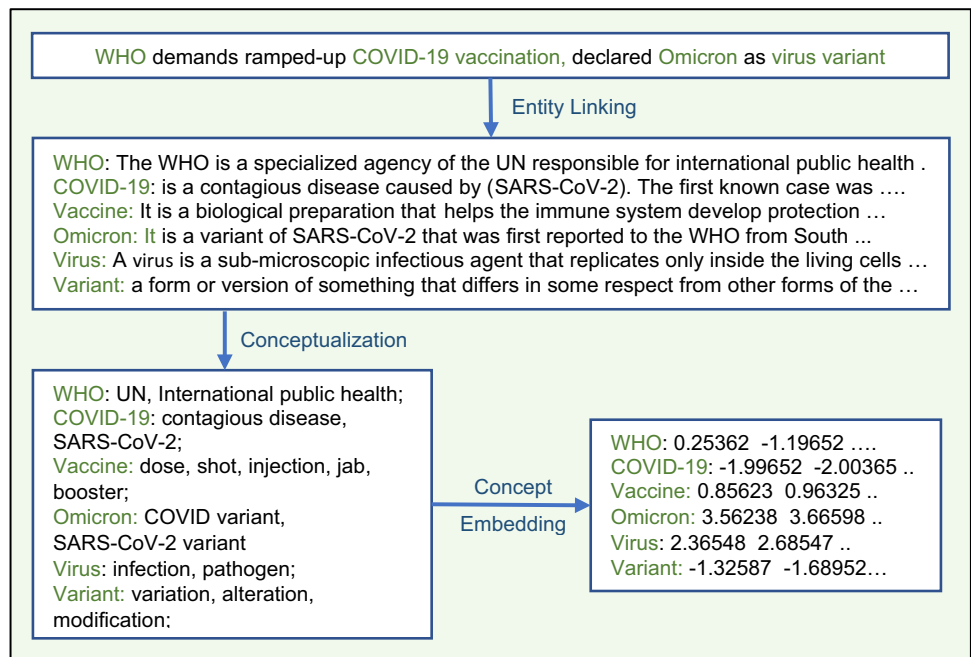
the semantic representation of short text posts. An illustration of the knowledge distillation process is presented in Fig. 15. Given a post, [16, 99] utilize the entity linking technique to associate the entity mentions in the text with pre-defined entities in a knowledge graph. Rel-Norm [138], Link Detector [139], and STEL [140] are widely used entity linking techniques. To gain the conceptual data of each distinguished entity from an existing knowledge graph, YAGO [141, 142], and Probase [143] can be exploited.

IV. *Event-level Features:* Individual microblog posts are short and have very limited content. An event generally contains various related posts relevant to the given claim. Detecting fake news at the event level comprises of predicting the veracity of the whole event instead of an individual post. Most of the existing MFND frameworks work at post level [4, 27, 28, 92, 95], and learn event-specific features that are not useful in predicting the unseen future events and suffer from generalization. But [15, 16] intends to accurately categorize the post into one of the K events based on the multimodal feature representations. [96] incorporates topic memory module that captures topic-wise global feature and also learns post representation shared across topics. The Event Adversarial Neural Network (EANN) model proposed in [15] captures the dissimilarities between different events using an event discriminator. The role of the event discriminator is to eliminate the event-specific features and learn shared transferable features across various events. This model is evaluated on two multimedia datasets extracted from Twitter and Weibo and shows an accuracy of 71% and 82% respectively which is better than the baseline models. Ying et al. offers a unique end-to-end Multi-modal Topic Memory Network (MTMN) [115] that captures event-invariant information by merging post representations shared across global latent topics features to address real-world scenarios of fake news in newly emerging posts.

Table 7 gives the comparative analysis of various existing state-of-the-art deep learning-based multimodal fake news detection models focusing on the techniques used for processing individual channels and the methods used for concatenating the features for presenting a combined feature space. The table also discusses about the future perspective. In addition to this, Table 8 provides a detailed analysis of the experimental setup of the DL-based MFND frameworks. These tables provide some ideas about how one can practically approach the problem of fake news.

4 Data collection

Before starting the data collection process, the developers must decide upon the size of the dataset, news domain (e.g., entertainment, politics etc.), media (e.g., texts, images, videos, etc.), type of disinformation (e.g., fake news, propaganda, rumors, hoaxes, etc.) in advance. Various datasets for the task of FND have been developed by various studies and these vary in terms of the news domain, size, type of misinformation, content type, rating scale, language, and media platform.

[44] has laid down the requirements for fake news detection Corpus (Fig. 16). The study in [70] has introduced and divided the requirements of FND dataset into four categories, namely Homogeneity requirements, Availability requirements, Verifiability requirements, Temporal requirements (Fig. 17).

This section describes various data collection and annotation strategies apart from the datasets that are available for the given task.

4.1 Existing datasets

Several datasets are available for FND and related tasks like LIAR, CREDBANK, FEVER but most of these are text-only data. There are only a few datasets that have text along with visual data. The authors in [70] have systematically reviewed and done a comparative analysis of twenty-seven popular FND datasets by providing insights into existing dataset. Table 9 gives an overview of the multimodal datasets that are available and are widely used in the study.

Despite the fact that multimodal fake news datasets are available, but these datasets still have some shortcomings. The above table clearly shows that the multimodal datasets that are available are coarse-grained (have 2 or 3 labels only). These datasets fail to acknowledge fine-grained labeling that may be found in datasets like, LIAR [106], but these datasets are unimodal and can't be used in a multimodal setting.

Even though FakeNewsNet [120] dataset is one of the newest benchmark datasets that contains content, social context as well as socio-temporal data, but is not available as whole and only subsets of dataset can be retrieved using APIs provided. This is due to the fact that the dataset uses Twitter data for capturing user engagement, and so is not entirely publicly accessible according to license regulations.

While some datasets incorporate data from various domains, the existing multimodal datasets focus on limited domains. One of the main reasons that contribute to the compromised performance is due to the fact that the existing datasets only focus on specific topics like politics, entertainment etc. and hence have domain-specific word

Table 7 Review of literature of existing multimodal fake news frameworks

Model, Ref	Contribution	Feature (s) used	Feature Extraction		Multi-modal Fusion	Activation function	Future Scope
			Text	Visual			
Att-RNN [25]	Fuses the textual and visual features along with the social context using attention mechanism	M, SC	Word2Vec	VGG-19	Attention network	Sigmoid, Softmax, ReLU, Tanh	To improve the proposed model's performance
EANN [15]	Uses event discriminator to discover event-specific data to enhance the detection efficiency on new events	M, ES	Text-CNN	VGG-19	Dense layer (Concatenation)	Softmax, ReLU	To improve the fusion network
SAME [98]	Fuses the multimodal information along with user sentiment using adversarial learning	M, SC	GloVe	VGG-19	Adversarial network	Softmax, ReLU	Early detection
MKEMN [16]	The model gathers the event-invariant features shared between different events and captures the external knowledge connections for effective news verification.	M, EK, ES	GloVe	VGG-19	CNN	Softmax, Tanh	Use memory network to exploit the rumor propagation information
MVAE [27]	Model discovers correlations across the modalities leveraging VAE	M	Word2Vec	VGG-19	Concatenation	Tanh, Softmax	Utilizing tweet propagation and user characteristics.
SpotFake [28]	The prime novelty of the proposed model is the use of pre-trained language model BERT	M, SC	BERT	VGG-19	Dense layer (Concatenation)	Sigmoid, ReLU	improvement on longer length articles
SpotFake+ [95]	The prime novelty of the proposed model is the use of the pre-trained language model XLNet	M	XLNet	VGG-19	Dense layer (Concatenation)	Sigmoid, ReLU	Incorporate meta-level feature modalities
MTMN [96]	Address early detection by fusing features shared by various topics with global features of latent topics and modeling intra-modal and inter-modal data in a combined framework	M, ES	BERT	ResNET50	Blended Attention network	Softmax	Explore effective ways to learn background knowledge

Table 7 (continued)

Model, Ref	Contribution	Feature (s) used	Feature Extraction		Multi-modal Fusion	Activation function	Future Scope
			Text	Visual			
SAFE [97]	Produces joint representation of textual and visual features of an article and uses cosine function to measure the similarity between them	M	Text-CNN	Text-CNN (image2sentence)	Cosine Similarity	Softmax, ReLU	Incorporate user and network information
[94]	proposes visual features that are extracted from multiple images, additionally, cosine similarity of the title and image tags embeddings is calculated to find the image-text similarity	M	BERT	VGG-16	Attention network	Softmax	Improve the performance of the proposed model
MCAN [129]	Proposes multiple co-attention layers that fuse and learn inter-modality relations	M	BERT	VGG-19	Multiple co-attention layers	ReLU	To extend the fusion with the co-attention network to fake news diffusion.
HMCAN [26]	Propose a multi-modal contextual attention network that takes data from different modalities which complement one another	M	BERT	ResNET50	Attention network	Softmax	Explore an effective way to exploit visual data and utilize auxiliary information
KMAGCN [99]	The model represents posts as graphs instead as word sequences to capture long-range non-consecutive semantic relations and leverages knowledge concepts along with multimodal information	M, EK	Adaptive graph convolutional network	VGG-19 (128-D)	Feature-level attention mechanism	Softmax, ReLU	To improve the proposed model's performance
CARMN [91]	The model keeps unique properties intact while reducing the noise induced while fusing different modalities	M	Word2Vec (32-D)	VGG-19	Multichannel CNN	Softmax, ReLU	Event-level multimodal fake news

Table 7 (continued)

Model, Ref	Contribution	Feature (s) used	Feature Extraction		Multi-modal Fusion	Activation function	Future Scope
			Text	Visual			
FND-SCTI [4]	Fuses multi-modal data along with an image-augmented text representation in a multi-task setting	M	Word2Vec	VGG-19	Hierarchical attention network, VAE	tanh, Softmax	Bot detection using user characteristics
- [110]	The proposed system consists of four independent parallel networks with individual predictions that are merged with the max voting ensemble method	M	GloVe	Image caption (Caption-Bot)	Ensemble with max voting	Softmax, Tanh, ReLU	Incorporating better image forensic techniques
MMCN [92]	Fuses the text and image embeddings by considering inter-modal relationships using a multi-modal cross-attention network	M	BERT	ResNET50	Cross Attention network	Softmax	To explore effective ways to utilize background knowledge

M Multimodal, *SC* Social Context, *EK* External Knowledge, *ES* Event Specific features

Table 8 Experimental Setup of Multimodal Fake News Detection Models

Model	Ref.	Dataset	Batch size	Learning rate	Dropout	Epochs	Optimizer	Loss function	Performance Evaluation
Att-RNN	[25]	Twitter16, Weibo	128	–	–	100	Stochastic gradient descent	Cross Entropy	Acc- ~78%, ~68%
EANN	[15]	Twitter15, Weibo	100	–	–	100	–	Cross Entropy	Acc- ~71%, ~82%
SAME	[98]	FakeNewsNet	128	0.001	0.5	–	RMSprop	Adversarial, Hybrid similarity, Cross entropy	Acc- ~77%, ~80%
MKEMN	[16]	Twitter15, PHEME	128	–	–	–	–	Cross Entropy	Acc- ~86%, ~81%
MVAE	[27]	Twitter15, Weibo	128	0.00001	–	300	Adam	VAE Loss	Acc- ~74%, ~82%
SpotFake	[28]	Twitter15, Weibo	256	0.0005, 0.001	0.4	–	Adam	–	Acc- ~72%, ~80%
SpotFake+	[95]	FakeNewsNet	–	–	0.4	–	–	–	Acc- ~84%, ~85%
SAFE	[97]	FakeNewsNet	–	–	–	–	–	Cross Entropy	Acc- ~87%, ~83%
–	[94]	FakeNewsNet	32	0.00005	0.2	60	Adam	–	F1 score- ~76%
MCAN	[129]	Twitter16, Weibo	–	–	–	100	Adam	Cross Entropy	Acc-~80%, ~89%
HMCAN	[26]	Weibo, Twitter15, PHEME	256	0.001	–	150	Adam	Cross Entropy	Acc- ~85%, ~89%, ~88%
KMAGCN	[99]	Weibo, Twitter15, PHEME	128	0.01	–	300	Adam	Cross Entropy	Acc- ~84%, ~78%, ~86%
CARMN	[91]	Twitter16, Weibo	150	–	–	150	Adam	Cross Entropy	Acc- ~74%, ~85%
FND-SCTI	[4]	Twitter15, Weibo	128	0.00001	–	300	Adam	VAE Loss	Acc- ~75%, ~83%
–	[110]	AllData, Kaggle datasets	32	–	–	40	Adam	Cross Entropy	Acc- ~95%, ~95%, ~95%
MMCN	[92]	Weibo, PHEME	64, 256	0.001	–	150	Adam	Cross Entropy	Acc- ~87%, ~87%
MTMN	[96]	Weibo, PHEME	256	0.001	–	200	Adam	Cross Entropy	Acc- ~88%, ~88%

usage, whereas a real news stream typically covers a wide range of domains.

4.2 Dataset collection and annotation

Social media platforms are one of the best ways to reach the masses. Everyday a huge volume of data is circulated, but as most of the data goes unchecked and so the proliferation of fake news on these platforms becomes easy. Many researchers have used these platforms to collect data and create their datasets. But as only a few datasets are available which are suited to our requirement, this section gives a clear understanding of the collection of data from online platforms. Broadly speaking there are two ways of data collection and sampling from online platforms, namely (i) top-down approach, and (ii) bottom-up approach.

The top-down approach involves a collection of information and posts keeping some keywords under consideration. This approach is useful for debunking long-standing rumors that are already known, the data is crawled from fact-checking websites like Snopes, PolitiFact etc. [59] uses this approach for data collection. But, on the contrary, if

emerging news articles are to be collected the bottom-up approach comes in handy and it collects all the relevant posts and articles in a given time frame. [12, 145] have used the bottom-up approach.

Collecting reliable datasets for FND is not a trivial task and requires the fact checking of news to label and rate the news items (as binary or multiclass rating scale). As the ground truth is already known if data is collected using the top-down approach, no further annotation is needed. But, in the case of the bottom-up approach annotation is necessary and can be performed in one of the three ways (1) manual expert-oriented fact checking; (2) automated fact checking via knowledge graphs and other web sources; or (3) crowd-sourced fact checking.

Web APIs are one of the simplest ways to access, collect, and store data from social networking networks, and they usually come with documentation that explains how to get the required data. An application can use the APIs to request data using a set of well-defined methods. For example, retrieving all the data posted by some particular user or all the posts containing a specific keyword from a social media platform APIs are used. Twitter and Sina Weibo are two key

Table 9 Multimodal Fake News datasets

Dataset	Year of release	Statistics	Domain	Contents	Labels	Collected from	Used in
Twitter 15 [144]	2015	361 (I) 7032 (F) 5008 (R)	Posts related to 11 events	Text, visual	2	Twitter	[4, 15, 26–28, 99]
Twitter 16 [89]	2016	413 (I) 9596 (F) 6225 (R)	Posts related to 17 events	Text, visual	2	Twitter	[25, 91, 111, 129]
Weibo [25]	2016	9528 (I) 4749 (F) 4779 (R)	Crawl the verified false rumor posts from May, 2012 to Jan, 2016	Text, visual	2	Weibo (Non-rumor tweets are verified by Xinhua News Agency, an authoritative news agency in China)	[4, 15, 25–28, 91, 91, 99]
PHEME [12]	2016	2672 (I) 1972 (F) 3830 (R)	9 different events, which include 5 cases of breaking news	Tweet, conversational threads	3	Twitter	[16, 92, 96, 99]
ALLData [100]	2018	20,015 (I) 11,941 (F) 8074 (R)	2016 US Presidential elections	The title, text, image, author and website	2	Fake and real news scraped from 240 websites and authoritative news websites, i.e., the New York Times, Washington Post, etc. respectively	[100, 110, 111]
FakeNewsNet [120]	2019	19,200 (I) 5367 (F) 17,222 (R)	Politics, Entertainment	Text, image url, conversational threads, location, and timestamp of engagement	2	Content is crawled from PolitiFact, GossipCop, E! online; For user engagements Twitter API is used	[94, 95, 97, 98]

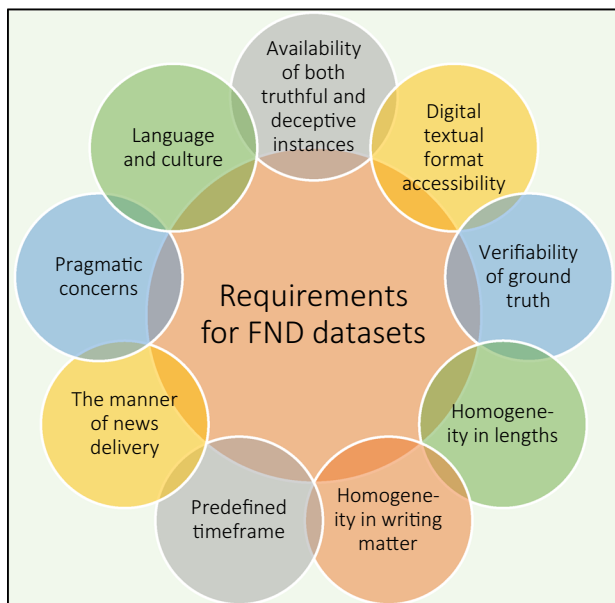Note: I—Total Number of Images, F—Number of Fake claims, R—Number of Real claims

Fig. 16 Requirements for fake news detection datasets defined by [44]

platforms that have been widely explored for studying fake news. [61, 146] have provided and given an overview of several ways of data collection from social media sites. Twitter provides REST APIs to allow users to interact with their service but the user should have a developer account. Tweets based on particular topics can be extracted in real-time from Twitter API using the Tweepy library. Similarly, Sina Weibo provides an API to help users access their data. Apart from this, the data can be also scaped from any website using web crawlers. Beautiful Soup is one of the Python libraries for pulling data out of HTML and XML files.

5 Open issues and future direction— discussion

In this section, we discuss several challenging problems of fake news detection (Fig. 18). To avoid any further spread of fake news on social media, it is particularly important

Fig. 17 Requirements for fake news detection datasets as defined by [70]

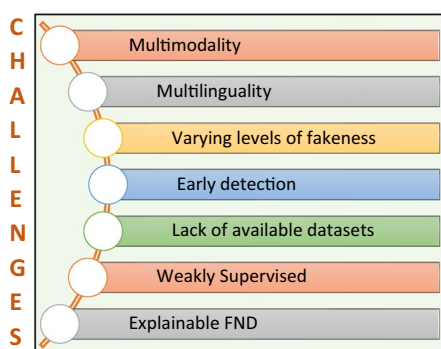| **Homogeneity Requirements** |
| • homogeneity of news length: news artiles in the dataset should be of comparable lengths; |
| • homogeneity in news domain: the text corpus should be aligned with the news domains; |
| • homogeneity in type of disinformation: the texts in the dataset must be aligned with the type of misinformation |
| **Availability Requirements** |
| • availability of fake and real news: dataset should be balanced in terms of the composition of True and fake articles |
| • textual format availability: texts and textual transcriptions of audio and video should be available |
| • public availability: the dataset should be openly available |
| • multilingual: the availability of news items in multiple languages; |
| **Verifiability Requirement** |
| • verifiability of ground truth: the ability to identify whether the news is genuine or fabricated; |
| **Temporal Requirements** |
| • belongs to a predefined timeframe: the dataset should be collected within a set time window. |

Fig. 18 Fake news detection challenges

to identify fake news at an early stage. Because getting the ground truth and labeling the false news dataset is time-consuming and labor-intensive, investigating the problem in a weakly supervised scenario, i.e. with few or no labels for training, becomes essential. In addition, there has been relatively little research into the multimodal nature of fake news on social media. It's also important to understand why a piece of news is labeled as fake by machine learning models because the resulting explanation can provide fresh data and insights that remain hidden when using content-based models.

There are several challenges associated with the FND problem. After extensive study and evaluation of the literature that is available and has been discussed in Sect. 2 and Sect. 3, the following research gaps have been identified.

i. *Multimodality:* With the rise of social media, there has been a change from all text to multimodal news articles. Images and videos are now included in news pieces, to complement the textual content. As

a result, a model that can work with a multi-modal dataset must be built. Although several models have been established to combat such a problem, there has been very little research on the multimodal nature of fake news on social media, as working in a multimodal setting is difficult in and of itself. While some of the offered approaches have been successful in detecting fake news, they still face the challenge of determining the relationships between multiple modalities.

ii. *Multi-linguality:* Most of the existing works in this domain focus on detecting fake news in the English language, very limited work is done on multi-lingual and cross-lingual fake news detection models. Social media platforms are used by a huge population all over the world and hence are not limited to the usage of one language. Also, collecting and annotating fake articles in foreign languages is difficult and time-consuming.

iii. *Varying levels of fakeness:* The majority of existing fake news detection methods tackle the problem from a binary standpoint. However, in practice, a piece of news might be a mixture of factual and false statements. As a result, it's critical to divide fake news into several categories based on the degree of deception. Nevertheless, for multiclass fake news detection, the classifier needs to offer better discriminative power and be more robust as the boundary between classes becomes more intricate as the number of classes increases.

iv. *Early detection:* Fake news or rumor has a very negative impact on the people and society at large. This kind of propaganda can even tarnish the image of a person or organization, so it becomes very crucial to track down a piece of fake news or rumor at an early stage. Fake news early detection aims to curb fake sto-

ries at an early stage by giving prompt signals of fake news during the diffusion process.

v. *Lack of available datasets:* The data source is also a challenge. Unstructured data contains a lot of unnecessary data and junk values that can degrade the algorithm's performance. The number of datasets in the subject of fake news is quite restricted, and only a few are available online. The multi-modal dataset present in this problem domain is scarce. A comparison of major benchmark multimodal fake news datasets is shown in the Table 9 in the above section.

vi. *Weakly supervised fake news detection:* Labels are predicted with low or no supervision labels in a weakly supervised environment. Because getting the ground truth of false news is time-consuming and labor-intensive, it's crucial to investigate fake news detection in a weakly supervised scenario, that is, with few or no labels for training.

vii. *Explainable fake news detection:* The problem of detecting fake news has yielded promising breakthroughs in recent years. However, manually classifying fake news is quite subjective, and there is a vital missing aspect of the study that should explain why a specific item of news is considered to be fake by providing the reader web-based proof. Traditionally, it used manual methods to verify the news content's validity with a variety of sources.

6 Conclusion

In this paper, we have presented an overview of contemporary state-of-the-art techniques and approaches to resolve the issue of detecting fake news on social media platforms with a focus on multimodal context. Our review is primarily concentrated on five key aspects. First, the paper provides a clear definition of Fake News and distinctions between various related terms with an appropriately defined taxonomy of the fake news detection techniques. While surveying we found out that limited work has been done on the multimodal aspect of the news content. Secondly, various DL models, frameworks, libraries, and transfer learning approaches that are widely used in the literature have also been emphasized with TensorFlow being the one that is widely used. Third, we have provided an impression of various state-of-the-art techniques to perform fake news detection on social media platforms using deep learning approaches considering the multimodal data. The review shows that CNN-based models are widely used for handling the image data and the RNN-based models are used for preserving the sequential information present in the text. Additionally, various modifications of the attention network are used to preserve the correlation

between text and image. These models take English as their primary language for detection and lack in processing the multi-lingual data which is prevalent with the use of social media. Fourth, the review sheds light on various data collection sources and data extraction techniques. Since this field of research is quite novel, there are only a few multimodal datasets that are available for this particular task. Finally, we have provided some insights to open issues and possible future directions in this area of research and found out that handling the multimodal data while maintaining the correlation becomes a challenge.

Declarations

References

1. S. Singhania, N. Fernandez, and S. Rao, "3HAN: A Deep Neural Network for Fake News Detection Sneha," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10635 LNCS, no. October, pp. 118–125, 2017, https://doi.org/10.1007/978-3-319-70096-0_59.

2. H. Ahmed, I. Traore, and S. Saad, "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques," *First Int. Conf. Intelligent, Secur. Dependable Syst. Distrib. Cloud Environ.*, vol. 10618, pp. 169–181, 2017, https://doi.org/10.1007/978-3-319-69155-8.

3. H. Karimi and J. Tang, "Learning hierarchical discourse-level structure for fake news detection," *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, pp. 3432–3442, 2019, https://doi.org/10.18653/v1/n19-1347.

4. Zeng, J., Zhang, Y., Ma, X.: Fake news detection for epidemic emergencies via deep correlations between text and images. Sustain. Cities Soc. **66**, 102652 (2021). https://doi.org/10.1016/j.scs.2020.102652

5. R. Oshikawa, J. Qian, and W. Y. Wang, "A survey on natural language processing for fake news detection," *Lr. 2020 - 12th Int. Conf. Lang. Resour. Eval. Conf. Proc.*, pp. 6086–6093, 2020.

6. S. Yoon *et al.*, "Detecting incongruity between news headline and body text via a deep hierarchical encoder," *33rd AAAI Conf. Artif. Intell. AAAI 2019, 31st Innov. Appl. Artif. Intell. Conf. IAAI 2019 9th AAAI Symp. Educ. Adv. Artif. Intell. EAAI 2019*, pp. 791–800, 2019, https://doi.org/10.1609/aaai.v33i01.3301791.

7. Bodaghi, A., Oliveira, J.: The characteristics of rumor spreaders on Twitter: a quantitative analysis on real data. Comput. Commun. **160**, 674–687 (2020). https://doi.org/10.1016/j.comcom.2020.07.017

8. S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang, "Prominent features of rumor propagation in online social media," *Proc. - IEEE Int. Conf. Data Mining, ICDM*, pp. 1103–1108, 2013, https://doi.org/10.1109/ICDM.2013.61.

9. Rath, B., Gao, W., Ma, J., Srivastava, J.: From Retweet to Believability: Utilizing Trust to Identify Rumor Spreaders on Twitter. Soc. Netw. Anal. Min. **8**(1), 179–186 (2018). https://doi.org/10.1007/s13278-018-0540-z

10. K. Shu, H. R. Bernard, and H. Liu, "Studying Fake News via Network Analysis: Detection and Mitigation," no. January, pp. 43–65, 2019, https://doi.org/10.1007/978-3-319-94105-9_3.

11. Wu, Z., Pi, D., Chen, J., Xie, M., Cao, J.: Rumor detection based on propagation graph neural network with attention mechanism. Expert Syst. Appl. **158**, 113595 (2020). https://doi.org/10.1016/j.eswa.2020.113595

12. Zubiaga, A., et al.: Analysing how people orient to and spread rumours in social media by looking at conversational threads. PLoS ONE **11**(3), 1–29 (2016). https://doi.org/10.1371/journal.pone.0150989

13. K. Shu, S. Wang, and H. Liu, "Understanding User Profiles on Social Media for Fake News Detection," *Proc.—IEEE 1st Conf. Multimed. Inf. Process. Retrieval, MIPR 2018*, pp. 430–435, Jun. 2018, https://doi.org/10.1109/MIPR.2018.00092.

14. K. Shu, X. Zhou, S. Wang, R. Zafarani, and H. Liu, "The role of user profiles for fake news detection," *Proc. 2019 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Mining, ASONAM 2019*, pp. 436–439, 2019, doi: https://doi.org/10.1145/3341161.3342927.

15. Y. Wang *et al.*, "EANN: Event adversarial neural networks for multi-modal fake news detection," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 849–857, Jul. 2018, https://doi.org/10.1145/3219819.3219903.

16. H. Zhang, Q. Fang, S. Qian, and C. Xu, "Multi-modal knowledge-aware event memory network for social media rumor detection," *MM 2019 - Proc. 27th ACM Int. Conf. Multimed.*, pp. 1942–1951, Oct. 2019, https://doi.org/10.1145/3343031.3350850.

17. J. Ma *et al.*, "Detecting rumors from microblogs with recurrent neural networks," *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 2016-Janua, pp. 3818–3824, 2016.

18. K. Popat, S. Mukherjee, A. Yates, and G. Weikum, "Declare: Debunking fake news and false claims using evidence-aware deep learning," *Proc. 2018 Conf. Empir. Methods Nat. Lang. Process. EMNLP 2018*, pp. 22–32, 2020, https://doi.org/10.18653/v1/d18-1003.

19. L. Hu *et al.*, "Compare to the knowledge: Graph neural fake news detection with external knowledge," *ACL-IJCNLP 2021 - 59th Annu. Meet. Assoc. Comput. Linguist. 11th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf.*, pp. 754–763, 2021, https://doi.org/10.18653/v1/2021.acl-long.62.

20. J. Ma, W. Gao, Z. Wei, Y. Lu, and K. F. Wong, "Detect rumors using time series of social context information on microblogging websites," *Int. Conf. Inf. Knowl. Manag. Proc.*, vol. 19–23-Oct-, no. October, pp. 1751–1754, 2015, https://doi.org/10.1145/2806416.2806607.

21. Jiang, J., Wen, S., Yu, S., Xiang, Y., Zhou, W.: Rumor Source Identification in Social Networks with Time-Varying Topology. IEEE Trans. Dependable Secur. Comput. **15**(1), 166–179 (2018). https://doi.org/10.1109/TDSC.2016.2522436

22. Kwon, S., Cha, M., Jung, K.: Rumor detection over varying time windows. PLoS ONE **12**(1), 1–19 (2017). https://doi.org/10.1371/journal.pone.0168344

23. N. Ruchansky, S. Seo, and Y. Liu, "CSI: A hybrid deep model for fake news detection," *Int. Conf. Inf. Knowl. Manag. Proc.*, vol. Part F1318, pp. 797–806, Nov. 2017, https://doi.org/10.1145/3132847.3132877.

24. Shin, J., Jian, L., Driscoll, K., Bar, F.F.F.: The diffusion of misinformation on social media: Temporal pattern, message, and source. Comput. Human Behav. **83**, 278–287 (2018). https://doi.org/10.1016/j.chb.2018.02.008

25. Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo, "Multimodal fusion with recurrent neural networks for rumor detection on microblogs," *MM 2017 - Proc. 2017 ACM Multimed. Conf.*, pp. 795–816, Oct. 2017, https://doi.org/10.1145/3123266.3123454.

26. S. Qian, J. Wang, J. Hu, Q. Fang, and C. Xu, "Hierarchical Multi-modal Contextual Attention Network for Fake News Detection," *SIGIR 2021 - Proc. 44th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, no. 1, pp. 153–162, 2021, https://doi.org/10.1145/3404835.3462871.

27. D. Khattar, M. Gupta, J. S. Goud, and V. Varma, "MvaE: Multimodal variational autoencoder for fake news detection," *Web Conf. 2019 - Proc. World Wide Web Conf. WWW 2019*, no. May, pp. 2915–2921, May 2019, https://doi.org/10.1145/3308558.3313552.

28. S. Singhal, R. R. Shah, T. Chakraborty, P. Kumaraguru, and S. Satoh, "SpotFake: A multi-modal framework for fake news detection," *Proc.—2019 IEEE 5th Int. Conf. Multimed. Big Data, BigMM 2019*, pp. 39–47, Sep. 2019, https://doi.org/10.1109/BigMM.2019.00-44.

29. Shah, D., Zaman, T.: Finding rumor sources on random trees. Oper. Res. **64**(3), 736–755 (2016). https://doi.org/10.1287/opre.2015.1455

30. Shelke, S., Attar, V.: Source detection of rumor in social network—a review. Online Soc. Netw Media **9**, 30–42 (2019). https://doi.org/10.1016/j.osnem.2018.12.001

31. D. Kr, "On Rumor Source Detection and Its," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 1, no. 840, pp. 110–119, 2017, doi: https://doi.org/10.1007/978-3-319-54472-4.

32. Sam Spencer and R. Srikant, "Maximum likelihood rumor source detection in a star network."

33. Xu, F., Sheng, V.S., Wang, M.: Near real-time topic-driven rumor detection in source microblogs. Knowledge-Based Syst. **207**, 106391 (2020). https://doi.org/10.1016/j.knosys.2020.106391

34. K. Popat, S. Mukherjee, J. Strötgen, and G. Weikum, "Where the truth lies: Explaining the credibility of emerging claims on the web and social media," *26th Int. World Wide Web Conf. 2017, WWW 2017 Companion*, pp. 1003–1012, 2017, https://doi.org/10.1145/3041021.3055133.

35. C. Cai, L. Li, and D. Zeng, "Detecting social bots by jointly modeling deep behavior and content information," *Int. Conf. Inf. Knowl. Manag. Proc.*, vol. Part F1318, pp. 1995–1998, 2017, https://doi.org/10.1145/3132847.3133050.

36. M. Orabi, D. Mouheb, Z. Al Aghbari, and I. Kamel, "Detection of Bots in Social Media: A Systematic Review," *Inf. Process. Manag.*, vol. 57, no. 4, p. 102250, 2020, https://doi.org/10.1016/j.ipm.2020.102250.

37. Kudugunta, S., Ferrara, E.: Deep neural networks for bot detection. Inf. Sci. (Ny) **467**, 312–322 (2018). https://doi.org/10.1016/j.ins.2018.08.019

38. G. Gorrell, K. Bontcheva, L. Derczynski, E. Kochkina, M. Liakata, and A. Zubiaga, "RumourEval 2019: Determining rumour veracity and support for rumours," *arXiv*, pp. 69–76, 2018.

39. G. Gorrell *et al.*, "SemEval-2019 Task 7: RumourEval, Determining Rumour Veracity and Support for Rumours," pp. 845–854, 2019, https://doi.org/10.18653/v1/s19-2147.

40. W. Ferreira and A. Vlachos, "Emergent: A novel data-set for stance classification," *2016 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. NAACL HLT 2016 - Proc. Conf.*, no. 1, pp. 1163–1168, 2016, https://doi.org/10.18653/v1/n16-1138.

41. J. Ma, W. Gao, and K. F. Wong, "Detect Rumor and Stance Jointly by Neural Multi-task Learning," *Web Conf. 2018 - Companion World Wide Web Conf. WWW 2018*, pp. 585–593, 2018, https://doi.org/10.1145/3184558.3188729.

42. S. Dungs, A. Aker, N. Fuhr, and K. Bontcheva, "Can Rumour Stance Alone Predict Veracity?," *Proc. 27th Int. Conf. Comput. Linguist.*, pp. 3360–3370, 2018, [Online]. https://aclweb.org/anthology/papers/C/C18/C18-1284/.

43. S. M. Mohammad, P. Sobhani, and S. Kiritchenko, "Stance and sentiment in Tweets," *ACM Trans. Internet Technol.*, vol. 17, no. 3, 2017, https://doi.org/10.1145/3003433.

44. Rubin, V.L., Chen, Y., Conroy, N.J.: Deception detection for news: Three types of fakes. Proc. Assoc. Inf. Sci. Technol. **52**(1), 1–4 (2015). https://doi.org/10.1002/pra2.2015.145052010083

45. F. Yang, X. Yu, Y. Liu, and M. Yang, "Automatic detection of rumor on Sina Weibo," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2012, https://doi.org/10.1145/2350190.235020 3.

46. Boididou, C., Papadopoulos, S., Zampoglou, M., Apostolidis, L., Papadopoulou, O., Kompatsiaris, Y.: Detection and visualization of misleading content on Twitter. Int. J. Multimed. Inf. Retr. **7**(1), 71–86 (2018). https://doi.org/10.1007/s13735-017-0143-x

47. Boididou, C., et al.: Verifying information with multimedia content on twitter: a comparative study of automated approaches. Multimed. Tools Appl. **77**(12), 15545–15571 (2018). https://doi.org/10.1007/s11042-017-5132-9

48. Kakol, M., Nielek, R., Wierzbicki, A.: Understanding and predicting Web content credibility using the Content Credibility Corpus. Inf. Process. Manag. **53**(5), 1043–1061 (2017). https://doi.org/10.1016/j.ipm.2017.04.003

49. M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein, "A stylometric inquiry into hyperpartisan and fake news," *ACL 2018—56th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf. (Long Pap.*, vol. 1, pp. 231–240, 2018, https://doi.org/10.18653/v1/p18-1022.

50. P. Qi, J. Cao, T. Yang, J. Guo, and J. Li, "Exploiting multidomain visual information for fake news detection," *Proc. - IEEE Int. Conf. Data Mining, ICDM*, vol. 2019-Novem, pp. 518–527, 2019, https://doi.org/10.1109/ICDM.2019.00062.

51. X. Zhou and R. Zafarani, "Network-based fake news detection: a pattern-driven approach," *arXiv*, 2019, https://doi.org/10.1145/3373464.3373473.

52. Fire, M., Kagan, D., Elyashar, A., Elovici, Y.: Friend or foe? Fake profile identification in online social networks. Soc. Netw. Anal. Min. **4**(1), 1–23 (2014). https://doi.org/10.1007/s13278-014-0194-4

53. M. Del Vicario, W. Quattrociocchi, A. Scala, and F. Zollo, "Polarization and fake news: Early warning of potential misinformation targets," *arXiv*, vol. 13, no. 2, 2018, https://doi.org/10.1145/3316809.

54. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *3rd Int. Conf. Learn. Represent. ICLR 2015—Conf. Track Proc.*, pp. 1–14, 2015.

55. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016, https://doi.org/10.1109/CVPR.2016.90.

56. C. Szegedy *et al.*, "Going deeper with convolutions," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07–12-June, pp. 1–9, 2015, https://doi.org/10.1109/CVPR.2015.729859 4.

57. J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pretraining of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol.—Proc. Conf.*, vol. 1, no. Mlm, pp. 4171–4186, 2019.

58. Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized autoregressive pretraining for language understanding," *Adv. Neural Inf. Process. Syst.*, vol. 32, no. NeurIPS, pp. 1–18, 2019.

59. Jin, Z., et al.: Novel visual and statistical image features for microblogs news verification. IEEE Trans. Multimed. **19**(3), 598–608 (2017). https://doi.org/10.1109/TMM.2016.2617078

60. K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *arXiv*, vol. 19, no. 1, pp. 22–36, 2017, https://doi.org/10.1145/3137597.31376 00.

61. A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, and R. Procter, "Detection and resolution of rumours in social media: A survey," *arXiv*, vol. 51, no. 2, 2017.

62. J. Cao, J. Guo, X. Li, Z. Jin, H. Guo, and J. Li, "Automatic Rumor Detection on Microblogs: A Survey," vol. 1, no. c, pp. 1–14, 2018, [Online]. http://arxiv.org/abs/1807.03505.

63. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., Ortega-Garcia, J.: Deepfakes and beyond: A Survey of face manipulation and fake detection. Inf. Fusion **64**, 131–148 (2020). https://doi.org/10.1016/j.inffus.2020.06.014

64. P. Meel and D. K. Vishwakarma, "Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities," *Expert Syst. Appl.*, vol. 153, 2020, https://doi.org/10.1016/j.eswa.2019.112986.

65. M. R. Islam, S. Liu, X. Wang, and G. Xu, "Deep learning for misinformation detection on online social networks: a survey and new perspectives," *Soc. Netw. Anal. Min.*, vol. 10, no. 1, 2020, doi: https://doi.org/10.1007/s13278-020-00696-x.

66. X. Zhou and R. Zafarani, "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities," *ACM Comput. Surv.*, vol. 53, no. 5, 2020, https://doi.org/10.1145/3395046.

67. Zhang, X., Ghorbani, A.A.: An overview of online fake news: characterization, detection, and discussion. Inf. Process. Manag. **57**(2), 1–26 (2020). https://doi.org/10.1016/j.ipm.2019.03.004

68. Shu, K., et al.: "Combating disinformation in a social media age", *Wiley Interdiscip*. Rev. Data Min. Knowl. Discov. **10**(6), 1–23 (2020). https://doi.org/10.1002/widm.1385

69. Kandasamy, N., Murugasamy, K.: Detecting and filtering rumor in social media using news media event. Concurr. Comput. Pract. Exp. **33**(20), 1–10 (2021). https://doi.org/10.1002/cpe.6329

70. D'Ulizia, A., Caschera, M.C., Ferri, F., Grifoni, P.: Fake news detection: A survey of evaluation datasets. PeerJ Comput. Sci. **7**, 1–34 (2021). https://doi.org/10.7717/PEERJ-CS.518

71. "fake news - Explore - Google Trends." https://trends.google.com/trends/explore?date=2011-04-16 2021–04–16&q=fake news (accessed Jan. 04, 2022).

72. S. Hangloo and B. Arora, "Fake News Detection Tools and Methods—A Review," *arXiv*, Nov. 2021, [Online]. http://arxiv.org/abs/2112.11185.

73. "How is Facebook addressing false information through independent fact-checkers? | Facebook Help Centre." https://www.facebook.com/help/1952307158131536 (accessed Jan. 04, 2022).

74. "Updating our approach to misleading information." https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information (accessed Jan. 04, 2022).

75. "Introducing Birdwatch, a community-based approach to misinformation." https://blog.twitter.com/en_us/topics/product/2021/introducing-birdwatch-a-community-based-approach-to-misinformation (accessed Jan. 04, 2022).

76. H. Ahmed, I. Traore, and S. Saad, "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10618 LNCS, pp. 127–138, 2017, https://doi.org/10.1007/978-3-319-69155-8_9.

77. Zhou, X., Jain, A., Phoha, V.V., Zafarani, R.: Fake news early detection. Digit. Threat. Res. Pract. **1**(2), 1–25 (2020). https://doi.org/10.1145/3377478

78. Tuna, T., et al.: User characterization for online social networks. Soc. Netw. Anal. Min. **6**(1), 1–28 (2016). https://doi.org/10.1007/s13278-016-0412-3

79. P. Rosso and L. C. Cagnina, "Deception Detection and Opinion Spam," pp. 155–171, 2017, https://doi.org/10.1007/978-3-319-55394-8_8.

80. M. Hardalov, I. Koychev, and N. Preslav, "In Search of Credible News," *Int. Conf. Artif. Intell. Methodol. Syst. Appl.*, pp. 172–180, 2016, https://doi.org/10.1007/978-3-319-44748-3_29.

81. P. Galán-García, J. G. de la Puerta, C. L. Gómez, I. Santos, and P. G. Bringas, "Supervised Machine Learning for the Detection of Troll Profiles in Twitter Social Network: Application to a Real Case of Cyberbullying Patxi," *Log. J. IGPL*, pp. 42–53, 2016, https://doi.org/10.1007/978-3-319-01854-6.

82. O. Ajao, D. Bhowmik, and S. Zargari, "Fake news identification on Twitter with hybrid CNN and RNN models," *ACM Int. Conf. Proceeding Ser.*, pp. 226–230, Jul. 2018, doi: https://doi.org/10.1145/3217804.3217917.

83. Umer, M., Imtiaz, Z., Ullah, S., Mehmood, A., Choi, G.S., On, B.W.: Fake news stance detection using deep learning architecture (CNN-LSTM). IEEE Access **8**, 156695–156706 (2020). https://doi.org/10.1109/ACCESS.2020.3019735

84. Kaliyar, R.K., Goswami, A., Narang, P., Sinha, S.: FNDNet—A deep convolutional neural network for fake news detection. Cogn. Syst. Res. **61**, 32–44 (2020). https://doi.org/10.1016/j.cogsys.2019.12.005

85. Y. Liu and Y. B. Wu, "Early Detection of Fake News on Social Media Through Propagation Path Classification with Recurrent and Convolutional Networks," pp. 354–361.

86. O. Varol, E. Ferrara, F. Menczer, and A. Flammini, "Early detection of promoted campaigns on social media," *EPJ Data Sci.*, vol. 6, no. 1, 2017, https://doi.org/10.1140/epjds/s13688-017-0111-y.

87. Faustini, P.H.A., Covões, T.F.: Fake news detection in multiple platforms and languages. Expert Syst. Appl. **158**, 113503 (2020). https://doi.org/10.1016/j.eswa.2020.113503

88. G. B. Guacho, S. Abdali, N. Shah, and E. E. Papalexakis, "Semi-supervised content-based detection of misinformation via tensor embeddings," *Proc. 2018 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Mining, ASONAM 2018*, no. January, pp. 322–325, 2018, https://doi.org/10.1109/ASONAM.2018.8508241.

89. Boididou, C., et al.: Verifying Multimedia Use at MediaEval 2016. CEUR Workshop Proc. **1739**, 4–6 (2016)

90. P. Zhou, H. Xintong, V. I. Morariu, and L. S. Davis, "Learning Rich Features for Image Manipulation Detection," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1053–1061, 2018, [Online]. Available: http://openaccess.thecvf.com/content_cvpr_2018/papers/Zhou_Learning_Rich_Features_CVPR_2018_paper.pdf%0A, http://dl.acm.org/citation.cfm?doid=3133956.3134027.

91. Song, C., Ning, N., Zhang, Y., Wu, B.: A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks. Inf. Process. Manag. **58**(1), 102437 (2021). https://doi.org/10.1016/j.ipm.2020.102437

92. Ying, L., Yu, H., Wang, J., Ji, Y., Qian, S.: Multi-level Multimodal Cross-attention Network for Fake News Detection. IEEE Access **9**, 132363–132373 (2021). https://doi.org/10.1109/ACCESS.2021.3114093

93. A. Silva, L. Luo, S. Karunasekera, and C. Leckie, "Embracing Domain Differences in Fake News: Cross-domain Fake News Detection using Multi-modal Data," 2021, [Online]. http://arxiv.org/abs/2102.06314.

94. A. Giachanou, G. Zhang, and P. Rosso, "Multimodal multi-image fake news detection," *Proc. - 2020 IEEE 7th Int. Conf. Data Sci. Adv. Anal. DSAA 2020*, pp. 647–654, 2020, https://doi.org/10.1109/DSAA49011.2020.00091.

95. S. Singhal, A. Kabra, M. Sharma, R. R. Shah, T. Chakraborty, and P. Kumaraguru, "SpotFake+: A multimodal framework for fake news detection via transfer learning (student abstract),"

96. Ying, L., Yu, H.U.I., Wang, J., Ji, Y., Qian, S.: Fake news detection via multi-modal topic memory network. IEEE Access **9**, 132818–132829 (2021). https://doi.org/10.1109/ACCESS.2021.3113981

97. X. Z. B, J. Wu, and R. Zafarani, "SAFE : Similarity-Aware Multi-modal Fake," pp. 354–367, 2020, https://doi.org/10.1007/978-3-030-47436-2.

98. L. Cui, S. Wang, and D. Lee, "Same: Sentiment-aware multimodal embedding for detecting fake news," *Proc. 2019 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Mining, ASONAM 2019*, pp. 41–48, Aug. 2019, https://doi.org/10.1145/3341161.3342894.

99. S. Qian, J. U. N. Hu, Q. Fang, and C. Xu, "Knowledge-aware Multi-modal Adaptive Graph Convolutional Networks for Fake News Detection," *ACM Trans. Multimed. Comput. Commun. Appl.*, vol. 17, no. 3, 2021, https://doi.org/10.1145/3451215.

100. Y. Yang *et al.*, "TI-CNN: Convolutional neural networks for fake news detection," *arXiv*, 2018, [Online]. http://arxiv.org/abs/1806.00749.

101. H. Guo, J. Cao, Y. Zhang, J. Guo, and J. Li, "Rumor detection with hierarchical social attention network," *Int. Conf. Inf. Knowl. Manag. Proc.*, pp. 943–952, Oct. 2018, doi: https://doi.org/10.1145/3269206.3271709.

102. C. Shao, G. L. Ciampaglia, O. Varol, K. C. Yang, A. Flammini, and F. Menczer, "The spread of low-credibility content by social bots," *Nat. Commun.*, vol. 9, no. 1, 2018, https://doi.org/10.1038/s41467-018-06930-7.

103. T. N. Nguyen, C. Li, and C. Niederée, "On early-stage debunking rumors on twitter: Leveraging the wisdom of weak learners," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10540 LNCS, pp. 141–158, 2017, https://doi.org/10.1007/978-3-319-67256-4_13.

104. S. Ahmed, K. Hinkelmann, and F. Corradini, "Combining machine learning with knowledge engineering to detect fake news in social networks - A survey," *CEUR Workshop Proc.*, vol. 2350, 2019.

105. E. Ortega-fernández, G. Padilla-castillo, S. L. Carcelén-garcía, and M. Arias-oliva, "fact checking agencies and processes to fight against fake news," pp. 219–228, 2020.

106. W. Y. Wang, "'Liar, liar pants on fire': A new benchmark dataset for fake news detection," *ACL 2017 - 55th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf. (Long Pap.*, vol. 2, pp. 422–426, 2017, https://doi.org/10.18653/v1/P17-2067.

107. A. Zubiaga, M. Liakata, R. Procter, G. Wong Sak Hoi, and P. Tolmie, "Analysing how people orient to and spread rumours in social media by looking at conversational threads," *PLoS One*, vol. 11, no. 3, pp. 1–29, 2016, https://doi.org/10.1371/journal.pone.0150989.

108. Hakak, S., Alazab, M., Khan, S., Gadekallu, T.R., Maddikunta, P.K.R., Khan, W.Z.: An ensemble machine learning approach through effective feature extraction to classify fake news. Futur. Gener. Comput. Syst. **117**, 47–58 (2021). https://doi.org/10.1016/j.future.2020.11.022

109. F. Monti, F. Frasca, D. Eynard, D. Mannion, and M. M. Bronstein, "Fake News Detection on Social Media using Geometric Deep Learning," *arXiv*, pp. 1–15, Feb. 2019, [Online]. http://arxiv.org/abs/1902.06673.

110. Meel, P., Vishwakarma, D.K.: HAN, image captioning, and forensics ensemble multimodal fake news detection. Inf. Sci. (Ny) **567**, 23–41 (2021). https://doi.org/10.1016/j.ins.2021.03.037

The AAAI 2020 - 34th AAAI Conf. Artif. Intell., pp. 13915–13916, 2020, https://doi.org/10.1609/aaai.v34i10.7230.

111. P. Meel and D. K. Vishwakarma, "Multi-modal Fusion using Fine-tuned Self- attention and Transfer Learning for Veracity Analysis of Web Information."

112. V. H. Nguyen, K. Sugiyama, P. Nakov, and M. Y. Kan, "FANG: Leveraging Social Context for Fake News Detection Using Graph Representation," *Int. Conf. Inf. Knowl. Manag. Proc.*, pp. 1165–1174, 2020, https://doi.org/10.1145/3340531.3412046.

113. Wang, Z., Guo, Y.: Rumor events detection enhanced by encoding sentimental information into time series division and word representations. Neurocomputing **397**, 224–243 (2020). https://doi.org/10.1016/j.neucom.2020.01.095

114. S. Singhal, "SpotFake : A Multi-modal Framework for Fake News Detection," 2015.

115. J. Z. Pan, S. Pavlova, C. Li, N. Li, Y. Li, and J. Liu, "Content based fake news detection using knowledge graphs," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11136 LNCS, pp. 669–683, 2018, https://doi.org/10.1007/978-3-030-00671-6_39.

116. C. Boididou, S. Papadopoulos, L. Apostolidis, and Y. Kompatsiaris, "Learning to detect misleading content on Twitter," *ICMR 2017 - Proc. 2017 ACM Int. Conf. Multimed. Retr.*, pp. 278–286, 2017, https://doi.org/10.1145/3078971.3078979.

117. K. Zhou, C. Shu, B. Li, and J. H. Lau, "Evaluating Event Credibility on Twitter," *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, pp. 1614–1623, 2019, https://doi.org/10.18653/v1/n19-1163.

118. K. Shu, S. Wang, and H. Liu, "Beyond news contents: The role of social context for fake news detection," *arXiv*, no. i, pp. 312–320, 2019, https://doi.org/10.1145/3289600.3290994.

119. K. Shu, H. Russell Bernard, and H. Liu, "Studying fake news via network analysis: Detection and mitigation," *arXiv*, pp. 836–837, 2018, https://doi.org/10.1007/978-3-319-94105-9_3.

120. Shu, K., Mahudeswaran, D., Wang, S., Lee, D., Liu, H.: FakeNewsNet: a data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big Data **8**(3), 171–188 (2020). https://doi.org/10.1089/big.2020.0062

121. Y. Dou, K. Shu, C. Xia, P. S. Yu, and L. Sun, "User Preference-aware Fake News Detection," *SIGIR '21 Proc. 44th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, pp. 2051–2055, 2021, https://doi.org/10.1145/XXXXXX.XXXXXX.

122. J. Ma, W. Gao, and K. F. Wong, "Rumor detection on twitter with tree-structured recursive neural networks," *ACL 2018 - 56th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf. (Long Pap.*, vol. 1, pp. 1980–1989, 2018, https://doi.org/10.18653/v1/p18-1184.

123. Q. You, L. Cao, H. Jin, and J. Luo, "Robust visual-textual sentiment analysis: When attention meets tree-structured recursive neural networks," *MM 2016 - Proc. 2016 ACM Multimed. Conf.*, pp. 1008–1017, 2016, https://doi.org/10.1145/2964284.2964288.

124. Nasir, J.A., Khan, O.S., Varlamis, I.: Fake news detection: A hybrid CNN-RNN based deep learning approach. Int. J. Inf. Manag. Data Insights **1**(1), 100007 (2021). https://doi.org/10.1016/j.jjimei.2020.100007

125. Kim, E., Cho, S.: Exposing fake faces through deep neural networks combining content and trace feature extractors. IEEE Access **9**, 123493–123503 (2021). https://doi.org/10.1109/ACCESS.2021.3110859

126. "The mostly complete chart of Neural Networks, explained | by Andrew Tch | Towards Data Science." https://towardsdatascience.com/the-mostly-complete-chart-of-neural-networks-explained-3fb6f2367464 (accessed Apr. 18, 2022).

127. A. Vaswani *et al.*, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 5999–6009, 2017.

128. Yu, H.U.I., et al.: Multi-level multi-modal cross-attention network for fake news detection. IEEE Access **9**, 132363–132373 (2021). https://doi.org/10.1109/ACCESS.2021.3114093

129. Y. Wu, P. Zhan, Y. Zhang, L. Wang, and Z. Xu, Multimodal fusion with co-attention networks for fake news detection, pp. 2560–2569, 2021, https://doi.org/10.18653/v1/2021.findings-acl.226.

130. Jindal, S., Sood, R., Singh, R., Vatsa, M., Chakraborty, T.: NewsBag: a multimodal benchmark dataset for fake news detection. CEUR Workshop Proc. **2560**, 138–145 (2020)

131. J. Feng *et al.*, "Generative adversarial networks based on collaborative learning and attention mechanism for hyperspectral image classification," *Remote Sens.*, vol. 12, no. 7, 2020, https://doi.org/10.3390/RS12071149.

132. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst.*, pp. 1097–1105, 2012, https://doi.org/10.1145/3383972.3383975.

133. C. Wang, P. Nulty, and D. Lillis, "A Comparative Study on Word Embeddings in Deep Learning for Text Classification," *PervasiveHealth Pervasive Comput. Technol. Healthc.*, no. December, pp. 37–46, 2020, https://doi.org/10.1145/3443279.3443304.

134. Y. Qi, D. S. Sachan, M. Felix, S. J. Padmanabhan, and G. Neubig, "When and why are pre-trainedword embeddings useful for neural machine translation?," *NAACL HLT 2018 - 2018 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 2, pp. 529–535, 2018, https://doi.org/10.18653/v1/n18-2084.

135. Y. Liu *et al.*, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," no. 1, 2019, [Online].: http://arxiv.org/abs/1907.11692.

136. Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations," pp. 1–16, 2019, [Online]. http://arxiv.org/abs/1909.11942.

137. Lahat, D., Adalı, T., Jutten, C., Multimodal, C.J.: Multimodal data fusion: an overview of methods, challenges and prospects. Inst. Electr. Electron. Eng. **103**(9), 1449–1477 (2015). https://doi.org/10.1109/JPROC.2015.2460697ï

138. P. Le and I. Titov, "Improving Entity Linking by Modeling Latent Relations between Mentions," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Apr. 2018, vol. 1, pp. 1595–1604, https://doi.org/10.18653/v1/P18-1148.

139. D. Milne and I. H. Witten, "Learning to Link with Wikipedia," 2008.

140. L. Chen, J. Liang, C. Xie, and Y. Xiao, "Short text entity linking with fine-grained topics," *Int. Conf. Inf. Knowl. Manag. Proc.*, pp. 457–466, Oct. 2018, https://doi.org/10.1145/3269206.3271809.

141. F. Suchanek *et al.*, "Yago : A Core of Semantic Knowledge Unifying WordNet and Wikipedia," 2007.

142. Hoffart, J., Suchanek, F.M., Berberich, K., Weikum, G.: YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. Artif. Intell. **194**, 28–61 (2013). https://doi.org/10.1016/j.artint.2012.06.001

143. W. Wu, H. Li, H. Wang, and K. Q. Zhu, "Probase: A Probabilistic Taxonomy for Text Understanding," 2012, Accessed: Dec. 16, 2021. [Online]. http://research.microsoft.com/.

144. C. Boididou *et al.*, "Verifying Multimedia Use at MediaEval 2015," Accessed: Apr. 18, 2022. [Online]. https://github.com/MKLab-ITI/image-verification-corpus/.

145. L. Derczynski *et al.*, "SemEval-2017 Task 8: RumourEval: Determining rumour veracity and support for rumours," Apr. 2017, Accessed: Dec. 30, 2021. [Online]. https://arxiv.org/abs/1704.05972v1.

146. Bondielli, A., Marcelloni, F.: A survey on fake news and rumour detection techniques. Inf. Sci. (Ny) **497**, 38–55 (2019). https://doi.org/10.1016/j.ins.2019.05.035

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.