A Review of Fake News Detection Models: Highlighting the Factors Affecting Model Performance and the Prominent Techniques Used

Suhaib Kh. Hamed[1]*, Mohd Juzaiddin Ab Aziz[2], Mohd Ridzwan Yaakub[3]
Center for Software Technology and Management (SOFTAM)-Faculty of Information Science and Technology, University Kebangsaan Malaysia (UKM), Bangi 43600, Selangor, Malaysia[1, 2]
Center for Artificial Intelligence Technology (CAIT)-Faculty of Information Science and Technology, University Kebangsaan Malaysia (UKM), Bangi 43600, Selangor, Malaysia[3]

***Abstract*—In recent times, social media has become the primary way people get news about what is happening in the world. Fake news surfaces on social media every day. Fake news on social media has harmed several domains, including politics, the economy, and health. Additionally, it has negatively affected society's stability. There are still certain limitations and challenges even though numerous studies have offered useful models for identifying fake news in social networks using many techniques. Moreover, the accuracy of detection models is still notably poor given we deal with a critical topic. Despite many review articles, most previously concentrated on certain and repeated sections of fake news detection models. For instance, the majority of reviews in this discipline only mentioned datasets or categorized them according to labels, content, and domain. Since the majority of detection models are built using a supervised learning method, it has not been investigated how the limitations of these datasets affect detection models. This review article highlights the most significant components of the fake news detection model and the main challenges it faces. Data augmentation, feature extraction, and data fusion are some of the approaches explored in this review to improve detection accuracy. Moreover, it discusses the most prominent techniques used in detection models and their main advantages and disadvantages. This review aims to help other researchers improve fake news detection models.**

***Keywords*—Fake news detection; social media; data augmentation; feature extraction; multimodal fusion**

I. Introduction

Social media platforms are now the main source of news consumption for many around the world. Unlike traditional media, social media networks support the rapid spread of posts to a wide audience in a short time and without validation restrictions and costs [1]. This contributes to fake news dissemination on social platforms [2]. Research has indicated that many people have trouble distinguishing between real and fake news. This issue is not related to a specific age or gender and does not depend on education [3]. Researchers observed that fake news spread 70% more than real news [4]. Recent studies have stated that fake news dissemination on social media has become a current-day issue that has attracted global attention that needs intervention and an immediate halt to its spread [5] because this issue is creating social panic and economic unrest [6]. Fake news detection is a difficult challenge. Therefore, the research community has paid much attention to this issue. It is considered one of the modern fields [7], and research in this field is still developing, but constantly increasing. This needs more improvement and exploration of upcoming directions in research to enhance fake news detection methods [1]. However, fake news detection overlaps with several domains [8]. Therefore, this matter has become of interest to many researchers from different disciplines and areas [9]. Based on previous studies, fake news models face many difficulties due to the distinct attributes of this issue These challenges include the lack of standard datasets, their small size, or their imbalanced distribution, which affects detection models' performance. Another issue to be highlighted is how to deal with social media data, the features used, and the improvement of techniques for extracting these features. In addition, developing methods for the fusion of features and making decisions. Although there are many review articles discussing several aspects of fake news detection, most merely review the techniques used in detection models. In addition, they categorize features by type, or group datasets based on labels or domains. What distinguishes this review is that it discusses the challenging aspects that still affect fake news detection models. It also discusses the possibility of increasing detection accuracy by providing future suggestions for improving the techniques used. This review investigates three key and critical aspects of fake news detection studies, namely datasets, extracted features, and data fusion. These aspects affect the accuracy of fake news detection models. Studies published in the following well-known databases and digital libraries in the academic field (Web of Science, ACM Digital Library, Springer Link, IEEE Explore, Science Direct) in English for the period from 2017 to 2023, which dealt with the three aspects referred to above, were covered. The main contributions of this review are as follows:

- Provide an overview of fake news, its types, the impact of its spread, and the role of social networks in disseminating news.
- Highlight the critical parts of the fake news detection model and the most serious limitations related to them.
- Investigate the methods used in several fields that have provided promising results and future suggestions for improvement.

The rest of this review article is arranged as follows: Section II summarizes relevant studies that discuss significant aspects of fake news detection. Section III provides an overview of fake news. Section IV shows the role of social media in disseminating news. Section V reviews the main modules that make up fake news detection models. Section VI investigates the main detection model techniques and their challenges. Section VII discusses the prominent techniques used in Detection models. Section VIII suggests future directions to improve detection models. Section IX concludes this review.

II. Related Works

There are some reviews presented that deal with fake news detection studies. Some reviewed certain aspects, and others investigated others. Cardoso Durier da Silva, Vieira [10] presented a review of studies that dealt with the use of Machine Learning (ML) techniques or Natural Language Processing (NLP) methods in detecting fake news on social networks. While Sharma, and Qian [11] investigated the causes of fake news propagation and ways to reduce it, as well as analyzing the characteristics of standard datasets used in relevant studies. According to Pathak, Mahajan [12], previous studies employing Machine Learning (ML) and Deep Learning (DL)-based models using supervised, unsupervised, and hybrid approaches to rumor detection were reviewed. In addition, they presented in their review the standard datasets used in rumor detection studies and discussed the features used. Vishwakarma and Jain [13] presented a concise review of the methods and datasets used in research on fake news detection. In addition, they categorized fake news in terms of its types. They also investigated the features used in these detection models which are textual content features, and image-based features. In a different context, Zhou and Zafarani [14] published a review examining techniques for identifying fake news from four perspectives: the misinformation it contains, the writing style, the patterns of dissemination, and the reliability of the source. They also highlighted fundamental theories from different fields related to fake news dissemination. In another direction, De Beer and Matthee [15] presented a systematic review discussing the approaches used to detect fake news. These methods included the language, topic-agnostic, machine learning, hybrid, and knowledge-based approach. While Alam, Cresci [16] reviewed misinformation detection research papers divided down by content-based features, including image, speech, video, and network and temporal information. They also mentioned certain difficulties with multimodal detection models. Ansar and Goswami [17] published a comprehensive review of characterizing fake news identification from a data science point of view. They covered the different types of fake news, and the different features used in detection models. They also presented the most significant standard datasets currently available in this area. Studies of COVID-19 misinformation detection were also investigated as a case study. Considering that the issue of detecting fake news is a classification problem, Li and Lei [18] produced a review article that classified research related to fake news detection based on DL based on the data structures used in news classification which are text classification, graph classification, and hybrid classification. An overview of fake news detection studies was presented by Swapna and Soniya [19]. This overview was categorized according to features utilized in these models, such as linguistic and semantic, style, and visual features. Meanwhile, Hu, Wei [20] provided a thorough analysis of DL-based fake news detection techniques that consider many aspects like content, social context, and external knowledge. In their review, several widely used datasets and pertinent studies were presented, in addition to suggesting future work.

III. Fake News Overview

Fake news is perceived as among the most severe dangers to journalism, freedom of expression, and autonomy. It has been proven in studies that in comparison to authentic news, fake news on social media gets more retweets and shares, especially political news [21]. This issue has reduced public confidence in governments, including the controversial “Brexit” referendum, as well as the divisive 2016 U.S. presidential election [22]. The strongest emphasis was placed on the reach of fake news in the crucial months of the 2016 U.S. presidential election movement. Fake news as a term was chosen by the Oxford Dictionary in 2016 as the international word of the year [23]. This nation's economy is susceptible to fake news, which is linked to the fluctuating stock market and big deals. As an example, fake news claimed that US President Barack Obama was injured in an explosion, which led to the erasure of $130 billion worth of shares [24]. In the case of fake news's ability to gain public trust, psychological and social elements play a significant part. They reinforce fake news distribution. It has been proven that people become less reasonable and show vulnerability when authenticity and fabrication are differentiated. At the same time, they are burdened with fake news. Based on research conducted on 1,000 participants in more than 100 experiments in social psychology and communications, a slightly higher human capability of identifying falsehood was recorded compared to possibility with common precision degrees ranging from 55% to 58% and a mean precision of 54% [25]. In the case of news where truth and objectivity are expected, gaining public confidence is easier. People also believe fake news after being exposed to them repeatedly (validity satisfies [26] or if the news satisfies their desirability bias [27], is in line with preexisting principles, bias in confirmation [28] or viewpoints (selective exposure [29]). In some cases, peer pressure could “control” people’s perspectives and conduct (e.g., the bandwagon impact [30]).

A. Fake News Types

"Fake news" refers to news items that have been published but contain misleading information to deceive readers intentionally [7] for malicious purposes [31]. Literature shows that there are several types of fake news, as illustrated in Fig. 1. These are rumor, disinformation, misinformation, hoax, and clickbait [7]. A rumor is an unconfirmed or unsupported statement, and it spreads like wildfire [32]. Disinformation is misleading information deliberately posted to deceive people, while misinformation is inaccurate information that is unintentionally shared [7]. When a user publishes false information with malicious intent, it falls under the category of disinformation [33]. It is the users' lack of knowledge of a

particular topic or field that causes misinformation to be circulated [7]. One category of fake news is a hoax whose purpose is to intentionally mislead the reader with ill-intention. This includes defrauding users and losing money [34]. According to psychological studies, clickbait is one of the forms of fake news that draws in readers by evoking their interest in learning more about the catchy news headline. It also encourages them to click [35]. The purpose of clickbait is to redirect readers to fake websites to increase traffic to websites containing ads. It is a type of attention-grabbing title that might not reflect the content of the article [34].

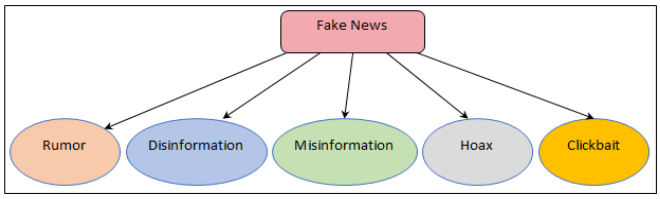

Fig. 1. Types of fake news.

B. Fake News Consequences

Since human society began, fake information has existed. But with the change and recent technological advancements in the global media environment, fake news circulation is growing. Fake news may cause severe consequences in political, social, economic, and health domains. Fake news comes in many forms. Fake news greatly affects the way information shapes our view. We make crucial decisions based on information. Based on the information we hear, we build an opinion about a situation or a collection of people. Contrived, misleading, twisted, or fraudulent information online prevents us from making the right decisions [36]. Fake news has the following main effects:

- Impact on citizens: Rumors about particular people can have a major effect. These individuals can face online abuse. In addition, they can be subjected to insults and threats that may have far-reaching negative effects. Individuals should not directly believe disinformation posted on social media and not make premature judgments about others based on this misleading information.
- Impact on health: People turn to the Internet for health news. Health-related fake news can affect people's lives. This problem has become the issue of the times. In the past few years, misinformation targeting health has had a serious negative impact. As a result, and based on effective lobbying by health organizations, doctors, and health advocates, many social media companies have had to change their policies to prevent and restrict misinformation spread.
- Impact on finances: Currently, fake news is a serious problem in industry and commerce. Fraudulent businessmen publish deceptive information or reviews to increase revenues. Stock prices may drop due to false information. It can destroy a company's reputation. Fake news also impacts customers' expectations. False news can breed unscrupulous commercial practices.
- Impact on democracy: Because fake news was so influential in the US presidential election, the fake news problem has received considerable media attention. Fake news has become a major issue threatening democracy, therefore, its spread must be stopped [37].

IV. Social Media Platforms

In recent years, rapid technological development, particularly in the mobile phone sector has made social media networks like Facebook, Twitter, and Sina Weibo accessible. These platforms have become an integral part of people's daily lives [7]. Social media is now a potent tool for all types of journalism, including sports, medical, and political reporting [15]. Instead of watching traditional media, most people now spend their time on social media to connect, gather knowledge, and share it [7]. Social media is used by many people to post or share news or information. This is because, unlike traditional media, news distribution through networks is real-time and quick, there are no costs involved, and there are no restrictions imposed on validation [4]. For example, in 2012 in the U.S., about 49% of users shared news on social media platforms. The Pew Research Center issued a report in 2016 stating that more than 62% of users daily receive their news from social networking pages [4], while in 2018, a report indicated that two-thirds of adults in the U.S. received their news from these pages [38]. The fact that social media is used on a variety of devices has significantly expanded the amount of data available [15]. Furthermore, it is worthwhile to refer to the issue of the language used in social media. This is because social media users come from all cultural, academic, and age backgrounds. Therefore, many posts on social media contain linguistic mistakes, use acronyms, are written in slang, or include obscene words. Fig. 2 shows a post containing slang and acronyms.

Fig. 2. A colloquial post on a facebook page.

There are drawbacks to these platforms despite their advantages. These platforms are misused by people or organizations to propagate false information for malicious purposes. This may be for financial gain, extremist hatred, or manipulation of people's minds for political reasons. It may also be intended to form biased opinions for electoral purposes [39]. The negative effects of social media as a result of fake news foreshadow a real danger that negatively affects individuals and society. This requires providing models for detecting fake news and limiting its spread [40].

V. Components of the Fake News Detection Model

In this section, the critical parts that make up the fake news detection model and affect its performance and how these elements relate to each other to carry out this task are reviewed. The detection model has three main components: a dataset, features, and a model based on a supervised classifier.

A. Dataset Used in Fake News Detection Models

Fake news detection studies demonstrated that no benchmark dataset is currently available that offers resources for extracting all crucial features. Fake news circulates on social media in various temporal patterns than real news. The dataset is the most significant component of fake news detection, and any model's effectiveness depends on it [34]. The larger size of the dataset [41], more diverse [42], more feature-rich [43], and low-noise [44], leads to improving the model's performance and increases its accuracy in identifying fake news [40]. Many researchers utilize fact-checking websites for data collection [45], because gathering data on fake news is time-consuming [46]. There are several significant challenges facing data gathering. These challenges include the ability to create a large volume of data in high quality, as well as ease of access without privacy restrictions. In addition to the data annotation process [7]. Data collection resources used by the researchers included reliable resources [44] such as government websites and websites of reputable media organizations, which contain information based on the facts [47]. According to researchers Subramani, Michalska [48], collecting, annotating, and labeling data is a laborious process that takes time, cost, and effort. The results are a medium-sized dataset that includes domains not previously investigated and considered acceptable by researchers. Some researchers are compelled to manually gather, annotate, and label data to produce a standard dataset that is checked for quality and reliability by specialists [48]. It should be underlined that facts, not opinions or feelings, must serve as the foundation for gathering ground truth data [44]. Additionally, it is essential to pre-process the dataset to eliminate extraneous data [42]. Future research would take less time, effort, and cost if standard datasets were created for previously unstudied topics, or expanded by adding more information [49]. This would benefit the research community. [50].

The researchers note that the model's accuracy is affected by the balance or imbalance in the dataset structure [51]. When compared to models that use a balanced dataset, models that use an imbalanced dataset produce higher results during training (a bias for one category over another) [52]. This means that these models are biased. Therefore, models that use a balanced dataset perform better [53]. Fig. 3 shows the most significant factors affecting the detection accuracy of fake news detection models.

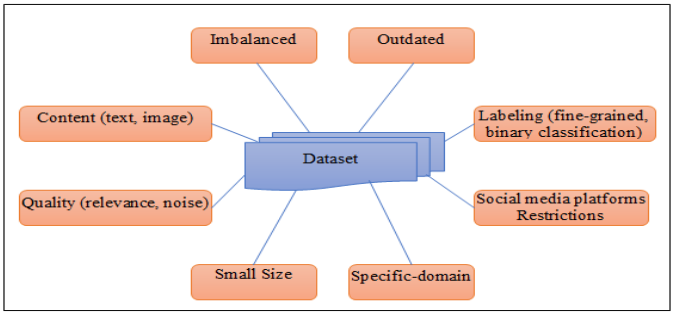

Fig. 3. The main factors affecting detection model accuracy.

B. Features Used in Fake News Detection Models

AI-based models for fake news detection rely on some key features, such as content-based, network-based, or user-based. In any case, using all of these features may not increase detection model accuracy. Depending on the nature of the issue, one or more features may be used. The results of the study by Kim, Kim [54] indicated that rumor detection accuracy using user features was the most poor of all. In contrast, rumor detection accuracy using content-only features was significantly higher than utilizing all features at once. According to these researchers, the experiments demonstrated that propagation features and user features were insufficient to identify rumors. Generally, news content (linguistics and visual information) is used as features in news identification [36]. Psychologists say people prefer articles with engaging visuals paired with text and believe them. When an article contains multimedia elements such as images instead of just text, it reaches more users [19]. Therefore, while textual content is essential for news verification, visual content also has a vital function in detecting fake news [17].

C. Fake News Classifiers Based on Supervised Learning

The performance of any model normally relies on the classifiers employed [43], the hyperparameter tuning [48], and the dataset used [34]. The diversity of AI techniques used in fake news detection models in previous research encouraged us to break up related studies based on supervised learning methods applied to detect fake news. Using labeled examples, supervised models learn from a variety of features [20]. Therefore, we list below the previous studies according to their use of ML and DL methods.

1) Machine learning-based models: Many previous research articles have employed Machine-Learning (ML) techniques to identify fake news. Aldwairi and Alwahedi [9] presented a study to detect fake news using a logistic classifier. They scraped their data from web pages. They also applied information gain and correlation attribute methods to rank the attributes according to how they relate. In all their experiments, the accuracy of the results exceeded 99%, and it was unclear whether they were training or test results. While Bhutani, Rastogi [55] proposed a model using Random Forest (RF) to identify fake news by employing sentiment analysis-based features. They utilized TF-IDF with cosine similarity for feature representation. The model was trained and tested on three datasets, each separately. However, there was a disparity in the results between training and testing, which means the model was overfitted. Therefore, the three datasets were combined, and based on them; the proposed model was trained and tested on this merged dataset. The AUC measure was 84.30%. In the same direction, Varshney and Vishwakarma [45] presented a two-stage model, data collection and classification. The data collection stage retrieves claims about statements from rumor-debunking web pages. Based on content similarity to claims, the data classification stage extracts features from web statements that have been retrieved. Claims are classified as real news or fake news by the RF classifier based on content- and sentiment-

based features. Faustini and Covões [56] used KNN, RF, NB, and SVM machine learning-based classifiers to identify fake news. SVM and RF outperformed other classifiers, according to experiments. They used five small datasets, and the number of examples in each dataset ranged from 137 instances, as in the btvlifestyle dataset, to 8,981 instances, as in the TwitterBR dataset. Regarding the feature set, the researchers applied the bag-of-words method in their experiments, as well as the DCDistance algorithm to reduce the dimensionality of the features. Although some of the results from some of the datasets examined may be encouraging, the majority of the results are poor. Kim, Kim [54] proposed an Ensemble Solution (ES) based on Soft Voting which consists of RF, XGBoost, and Multilayer perception. In order to determine which ML model was most suitable for creating an ES model, they examined several. To detect rumors, the latter model used the features of content, network, and users, and their model provided the highest F1 score of 79%.

2) Deep learning-based models: For identifying rumors, researchers Rath, and Gao [32] have developed a believability concept-based model that uses the LSTM classifier. Believability is determined based on the level of trust between Twitter users. They formed their dataset by merging two datasets, Twitter15 and Twitter16. LINE embedded the user in the network according to its reply and retweet. Their model provided an accuracy of 73.80%. Li, Hu [57] presented a detection model to identify fake news based on CNN at multiple levels. This model extracts semantic information from articles and represents it based on word and sentence levels. A pre-trained Word2Vec model was utilized for feature vectorization. In addition, they used the TFW method to calculate the weights of sensitive words to improve classification accuracy. They used five small datasets and combined them into two datasets. Their proposed model provided an accuracy of 88.80% and 90.10% on the Weibo and NewsFN datasets, respectively. Furthermore, Alkhodair, Ding [31] created a model that uses the LSTM classifier by plugging it in parallel with the Word2Vec embedding model. The model is constantly updated with new tweets so that the classifier could detect rumors about upcoming topics on Twitter based only on the text of the tweet. The highest score recorded for the suggested model was 79.50% based on the F1 score on the PHEME dataset. In their study, Braşoveanu and Andonie [8] examined several DL-based models to identify fake news using a variety of features from the datasets, such as textual, relational, and meta-features. Textual features were represented by a pre-trained Glove model. The two datasets used are Politifact and Liar, which are small and imbalanced. The results were unsatisfactory for most models with less than 50% accuracy. The highest result was an accuracy of 52.40% on the Politifact dataset and an accuracy of 64.90% on the Liar dataset. This was using the CapsNet model with the attention mechanism using all features. The researchers Guo, Xu [58] developed a CNN-based rumor detection model using Transfer Learning (TL). This proposed model was trained on the small YELP-2 dataset. By copying the basic model's parameter values and then re-adjusting them after the proposed model had been trained on a small Five Breaking News (FBN) dataset to prevent the negative transfer, the TL method addressed the issue of a limited training dataset. Their fine-tuned model produced an 82.50% result based on the F1 score. Regarding Gadek and Guélorget [59], they presented an interpretable text classifier built on CNN architecture with the Class Activation Maps (CAM) method. The classifier identifies fake news in two distinct stages by applying two different datasets. Text analysis is used in the first classification and emotion in the second classification. For their experiments, the Kaggle fake news and Signal-Media datasets were utilized. For the second dataset, they applied an under-sampling method to balance it. To extract the features, a pre-trained FastText model was applied to consider punctuation. The result of detecting fake news relying on textual features was 91.8% of the F1-score, while the result of identifying fake news using sentiment analysis features was 68.3% of the F1-score. While Kaliyar, Goswami [60] proposed a fake news detection model built on Multi-layer DNN. Features based on news textual content and social context were used to identify fake news and were represented by the tensor factorization technique. They conducted their experiments using the BuzzFeed and PolitiFact datasets, and the results on both datasets were 88.37 % based on F1-score. By capturing the connections between rumors and comments on a particular topic, the researchers' Lin and Chen [40] built a Feed Neural Networks (FNN) based model for rumor detection with multi-layer transformer encoding blocks and one fully connected layer. An attention mechanism has been utilized in transformer encoding blocks to improve model performance. They used two datasets Weibo and PHEME for training and testing, and their model provided an accuracy of 84.1% on the PHEME dataset. Consequently, DL is superior to Machine Learning, because of its capacity to extract high-dimensional features [61], automated feature extraction, little reliance on data pre-processing, and improved accuracy [36].

VI. Limitations of Fake News Detection Models

Based on the weak results of Wang's [62] research and the use of high dropout values in the proposed model trained on a fine-grained dataset of 12,800 examples, it can be concluded that the detection model was overfitted and high dropout values were used to eliminate overfitting. Ghanem, and Rosso [63] used an imbalanced fine-grained FNC-1 dataset in their research. They did not address the imbalanced dataset, as their model produced poor results. There is a variation in the results of the Kumar, Asthana [64] study, as the test results were very high for most models on a small test set. However, after testing these trained models on a larger volume of the same dataset from PolitiFact the results were poor. Note that the PolitiFact dataset has the same features as the one used to train the models. This means that their proposed models fell into an overfitting problem, which prevents them from generalizing to the new data. In addition, the FakeNewsNet dataset is multimodal and visual features were excluded from their

research. Based on the weak results of Shu, and Mahudeswaran's [65] experiments, it is clear to us that the lack of effective representation of features in detecting fake news, which was represented as a one-hot encoded vector, was one of the reasons for these results. In addition, one of the datasets used, although feature-rich, is small in size, which can lead to overfitting that reduces the generalizability of the model when tested on test data. Raza and Ding [66] used two multimodal datasets in their research, but visual features were not utilized in detecting fake news. In addition, the imbalanced dataset was addressed by the under-sampling method, as the omitted examples may include relevant attributes. In addition, no pre-trained word embedding model was used to enrich the model with features. All of these issues contributed to decreasing the detection model's performance. Elhadad, Li [67] used TF-IDF and N-Gram techniques to extract features and the use of such techniques may lead to the loss of many attributes, including neglecting to capture semantic relationships between words, and this is evident from the research results. Segura-Bedmar and Alonso-Bartolome [68] used the CNN model for feature extraction and did not employ a pre-trained model for visual feature extraction. Also, extracted textual and visual features were directly fused. In addition, they indicated that their model suffered from misclassification due to an imbalanced dataset. If advanced techniques were used, the results would be better. Also, the model presented by Singhal, Shah [69] fused extracted features based on simple concatenation. It was possible to increase accuracy if more focus was placed on data fusion. Kalra, Kumar [70] indicated that their model performed poorly because of the imbalanced fine-grained dataset used. Their model was overfitted and a dropout layer was added after each layer. Moreover, fusing multimodal features from different models directly loses many attributes.

VII. The Prominent Techniques Used in Detection Models

This section highlights some methods and techniques used in other fields. These techniques provided outstanding results, which can be used in fake news detection models to improve these models and increase their accuracy.

A. Dataset Augmentation Techniques

Deep Learning-based models are computationally costly and need properly labeled data for high performance. Enhancing the model's performance requires big data to identify the most features. Many researchers utilize fact-checking web pages for data collection, [44], because gathering a dataset related to fake and real news is time-consuming [46]. The ability to provide a big volume of data, its relevance to the research topic, its high quality, rich in features, and its ease of access without privacy limitations, particularly with social media data, are some of the most significant issues facing the data collection process [6]. Moreover, manually labeling this data is labor intensive [71]. The small size of standard datasets or they are available in large sizes but of poor quality, poses the biggest obstacle to developing and evaluating any model's efficacy in identifying fake news [72]. Deep learning-based approaches used for fake news detection require a lot of training data. The size of the data has an impact on the model's accuracy. This is why, the larger the dataset, the better the accuracy of the model [73]. The data augmentation method can create more data from the original existing data. This process leads to increases in the model's accuracy, without human effort to collect data and save time where it is difficult to collect more real data. The data augmentation process is one solution to reduce the occurrence of overfitting and underfitting during the training phase. This is done by increasing the dataset size with synthetically labeled data [74]. The method of data augmentation plays a vital role in the success of DL models, as this augmentation can lead to better detection accuracy for models when using these large datasets [75]. Data augmentation techniques have been widely used in computer vision, and have provided impressive results in image classification [76], in particular using the Generative Adversarial Network (GAN) method. GAN is an effective data augmentation method which is a type of deep network used to generate new examples [77]. In recent years, a trend has emerged for tackling NLP problems via natural language generation models such as LeakGAN. This is a modified GAN to deal with text. This is completely unsupervised or semi-supervised learning for data generation [78].

1) Generative Adversarial Network (GAN): Deep learning-based generative models are called GANs. Their architecture is made up of a generator model for producing new instances and a discriminator model for detecting whether the instances created by the generator model are real or fake. Adversarial networks are frequently utilized to produce images that match observed samples. The generator model creates new images that mimic the original image using features derived from training data. Whether the created image is fake or real is predicted by the discriminator model. In detail, a vanilla GAN is made up of two networks that cooperate during training: Generator and Discriminator as illustrated in Fig. 4. Generator: This network produces images with the same structure as the training set of images when a vector of random values is presented as input. Discriminator: This network attempts to classify observations as "real" or "fake" based on batches of images that include observations from the training set and images created by the generator. The generator output is directly connected to the discriminator input. The generator utilizes the discriminator classification as a signal by using a backpropagation process to update its weights. [79].

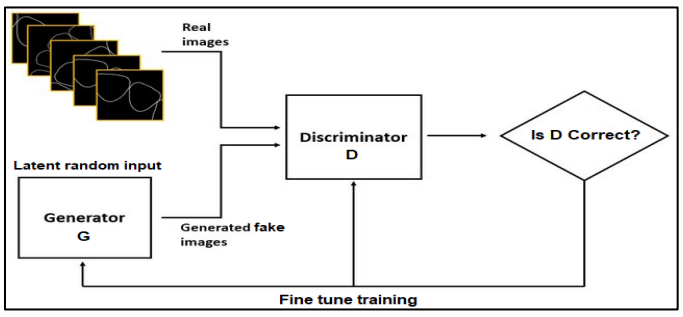

Fig. 4. The basic architecture of GAN.

In a semi-supervised environment, GANs are used to train discriminators and are very effective in generative modeling, which reduces the need for human intervention in data labeling. GANs are also helpful when data contain underrepresented samples or classes. GANs can only generate synthetic data if their foundation is a set of continuous numbers. The GAN technique has been successfully used mostly in image processing to produce real image samples. However, despite the fact that several prominent GANs models have also been suggested for image inpainting, synthesized images still have pixel errors or color inconsistencies throughout the image generation process. These errors are typically called fake textures [80].

2) LeakGAN: Unfortunately, there are two issues with using GAN in NLP to produce sequences. First off, GAN struggles to directly generate sequences of discrete tokens, like sentences, as it is built for producing real-valued, continuous data. For this reason, GAN begins with random sampling before moving on to a deterministic transform controlled by model parameters. The score/ loss for a complete sequence can only be provided by GAN after it has been formed; for a partially generated sequence, it is difficult to reconcile the present performance with the expected score for the entire sequence in the future [81]. However, because all NLP models are based on discrete variables like words, letters, or bytes, GANs cannot be used with NLP data. Novel strategies for training GANs on textual data are needed [36]. Some promising models that address this problem have been presented, such as LeakGAN which handles the problem of generating long text [82]. LeakGAN is a novel algorithmic framework proposed by Guo, Lu [82] that addresses both sparsity and non-informative problems related to previous GAN versions. LeakGAN is an innovative approach that builds on recent developments in hierarchical reinforcement learning. It delivers more information from the discriminator to the generator. A hierarchical generator G has been introduced as shown in Fig. 5, and it comprises a high-level MANAGER module and a low-level WORKER module. Mediation is performed by an LSTM called MANAGER. It gets generator D's high-level feature representation for each step, such as the CNN feature map, and utilizes it to create the WORKER module's guiding objective for that timestamp. Since D maintains its information and plays an adversarial game, it is not supposed to give G access to that information. As a result, it is called a leak of information from D. The WORKER then takes final action in the current state by combining the LSTM output and the goal embedding. This is given the goal embedding created by the MANAGER. This is done by encoding the currently generated words with another LSTM first. Therefore, the guiding signals from D are available to G both at the end in the form of scalar reward signals and during the generation process in the form of a goal embedding vector to help G improve. The discriminator evaluates the created sentence in an adversarial manner once the generator generates the following word. The main innovation is that, in contrast to traditional adversarial training, the discriminator communicates its internal state (feature) during the process to direct the generator more frequently and informatively. Consequently, LeakGAN achieved significant performance gains when generating longer sentences.

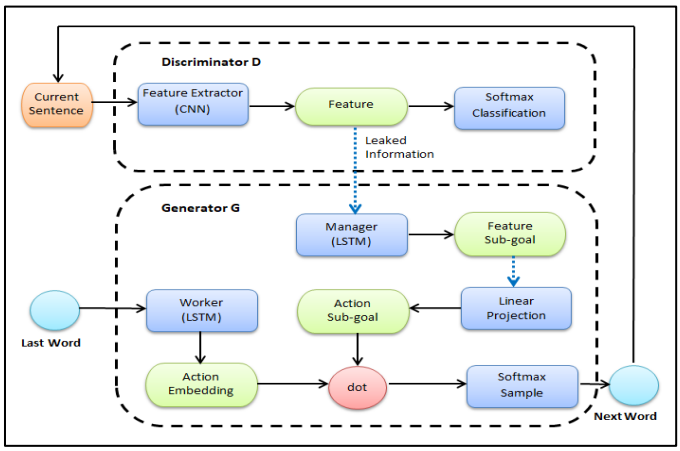

Fig. 5. The LeakGAN architecture.

B. Features Extraction Models

Feature extraction extracts a collection of features, known as a feature vector. This maximizes the prediction rate with the fewest number of elements and produces a similar feature set for several instances of the same symbol [83]. Thus, there is a need for some efficient techniques, such as vectorization, also referred to as word embeddings in the NLP field, and pre-trained image models in the image processing field. In NLP, feature extraction converts each text into a numerical representation in a vector [84], which is the initial stage in training a deep learning model. The word is considered the fundamental building block of documents in NLP [85]. When working with natural languages, the words in the document must be represented grammatically and semantically to obtain the intended results [86]. One of the common and effective techniques used in deep learning models for feature extraction is word embedding. Each word in the text string is transformed into a vector with n dimensions using embedding models, where n is the dimension of the embedding [64]. Word embedding models are distributed feature representations that are dense, low-dimensional, and well-suited to natural language tasks [87], based on deep learning [58]. Based on grammatical and semantic similarity, these models represent words and distribute them in vectors. They also take the word's relations to other terms in the document into account. Words with comparable meanings are represented by low-dimensional vectors [88]. The value of the distance between the two embedding vectors means how close the words are to each other according to the relationship between them [73]. For example, the terms "anxiety" and "depression" are semantically related because they fall under the same class relating to mental health [89] and also the terms "bad" and "good" are closely embedded for the same reason [90]. Word embedding models have proven to be remarkably effective in a variety of NLP tasks [73], including sentiment analysis, text classification, machine translation, and question-answering. This is according to earlier studies [90]. One such effective embedding model is

the BERT which was produced by Google. It performed well in classifying, composing, and summarizing texts. The BERT model was used to address the OOV word issue and the problem of polysemy in conventional embedding models and the incorporation of contextual information [91]. However, embedding methods have received increasing interest in extracting textual features and have the potential to develop better representations. As opposed to relationships between words, relationships between visual concepts in images are essential for computer vision (CV) tasks, but challenging to capture. Image-based features are a key component in detecting fake news [92]. There are several neural models based on transfer learning for image-based feature extraction such as AlexNet, VGG16, and VGG19.

1) Bidirectional Encoder Representations from Transformers (BERT): BERT was created to determine the relations among words within a sentence. BERT uses a language representation approach that only utilizes the encoder section of the transformer together with semi-supervised learning. In particular, BERT is built on a multi-layer bidirectional transformer encoder, which efficiently captures information from the left and right contexts of a token at each layer simultaneously [93]. In order to perform the pre-training, an unsupervised prediction operation will be executed using a masked language model (MLM) and a sentence-next predictor by BERT. In MLM, context knowledge comes before word prediction [94]. The BERT model often uses sentences broken up into individual tokens as inputs to produce a sequence of them. The BERT model takes context into account from both sides. Instead of processing each word separately, the transformer analyzes each word in connection with every other word in the sentence. Additionally, BERT's self-attention mechanism supports determining sentence keywords. The pre-trained BERT model could be effectively fine-tuned for advanced performance in various Natural Language Processing (NLP) tasks, proving that BERT models are incredibly adaptable. BERT's tokenizer is built on words and sub-words. Therefore, if a word is absent from the original vocabulary, it will be broken down into a series of sub-tokens that when combined will make up the original word. To ensure OOV tokens do not appear and that all vocabulary units are regularly updated and sufficiently trained during training, the remaining new tokens, those associated with more uncommon words, are simply divided into smaller units [95]. The BERT framework consists of two stages: pre-training and fine-tuning. BERT was trained using unlabeled data from English Wikipedia (2,500M words) and the Books Corpus (800M words). There are two types of BERT models, the BERT large model which consists of 24 layers of encoders, and the BERT base model which consists of 12 layers. Also, there are two versions, cased and uncased [96]. With just one additional output layer, the BERT pre-trained model may be adjusted to handle a variety of NLP-based tasks, including text summarization, sentiment analysis, chatbots, and machine translation. Fig. 6 shows the fine-tuning process for the pre-trained BERT.

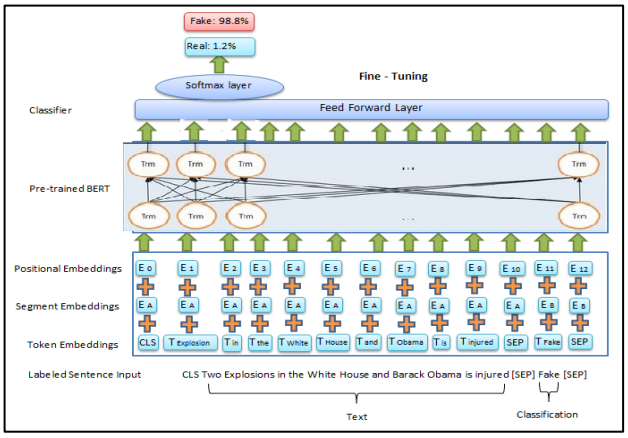

Fig. 6. The fine-tuning of the BERT model.

2) VGG-19 Model: An enhanced variant of the AlexNet architecture is the VGG16-Visual Geometry Group CNN architecture. An improved convolution neural network is implemented by expanding the network depth to 16 or 19 trainable layers. VGG networks in computer vision continue to be favored for many difficult problems [97]. The 143 million parameters of the deep architecture are learned from the ImageNet dataset. VGG receives RGB images in 224 × 224 pixels. The VGG-19 [98] is made up of 19 trainable weight layers, beginning with five stacks of convolutional layers and ending with three fully connected layers (FC) as illustrated in Fig. 7. These convolutional stacking layers carry out the process at each mark, extract image features, and then pass the result to the following layer. The number of filters increases by a factor of two, and all convolution layers employ a 3 by 3 filter size. A max-pooling layer and a Rectified Linear Unit (ReLU) activation function-based layer are followed as a non-linear activation function. Max-pooling layers are applied between each stack of convolution layers after ReLU, using a 2 by 2 kernel filter with 2 strides (pixels).

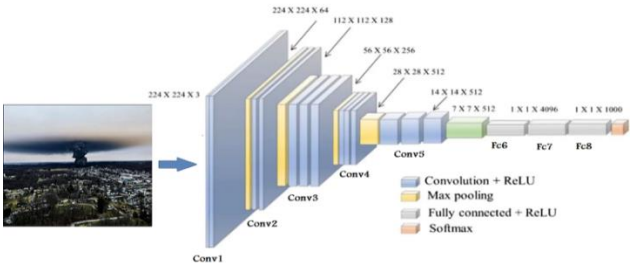

Fig. 7. The standard structure of VGG-19.

Some models, such as AlexNet and VGG-16, have encountered some issues that have been addressed in the improved VGG-19 version, including:

- Model Training: The first completely linked layer will yield a very high number of parameters, according to experiments on the original model VGG-16. This greatly increases the number of calculations and uses up more computational resources which leads to more training time [99].

- Vanishing Gradient (VG): Although network depth is significant, it can be challenging to train a deep network. This is due to the issue of vanishing or exploding gradients that arise when adding more network layers. In addition to gradient problems, if network depths continue to rise, model performance may quickly reach a limit before rapidly declining. Prior to AlexNet, the sigmoid, and tanh activation functions were most frequently utilized. These functions exhibit the VG problem due to their saturation, which makes it challenging for the network to train. AlexNet uses the ReLU activation function, which is immune to VG issues. ReLU aids with vanishing gradient problems, however, because it is unbounded, learned variables may rise excessively [100].
- Model Overfitting: Due to a large number of VGG-16 network parameters, overfitting can easily occur [99].

With the VGG-19 model, some of the previously mentioned problems have been addressed, and it offers higher accuracy than the VGG-16. The VGG-19 network is able to converge after a few iterations because of the implied regularization function of the network depth and the small convolution kernel size [101]. Having a large number of weight layers is the result of the small-size convolution filters in VGG-19, and surely, having more layers results in better performance. Despite this, less trainable variables lead to quicker learning and more resistance to overfitting.

C. Multimodal Fusion Method

Fusion is a crucial area of research in multimodal studies because it combines data from various unimodal data sources into a single, condensed multimodal representation. Fusional representations and multimedia are interrelated [102]. Multimodal fusion has generated a great deal of attention among scholars and widespread concern since it is an efficient method of processing multimodal data acquired [103]. Multimodal fusion is used to gain rich features by integrating various modalities [104]. The current challenge is merging and refining information coming from various modalities. Each modality contributes to varying functions. During the analysis of fusion features, the noise must be removed and relevant information extracted [105]. Multimodal fusion combines features from text and image modalities that must be merged before classification can be performed. Multimodal data fusion can produce extra information that improves the outcome precision. For example, compared to a single-modal CNN-based detection model, a multimodal fusion model for autonomous vehicle detection that fuses features of images from cameras with information from Light Detection and Ranging (LiDAR) sensors can attain noticeably improved accuracy by 3.7% over the previous one [106]. Multimodal fusion methods come in a variety of ways as shown in Fig. 8, including:

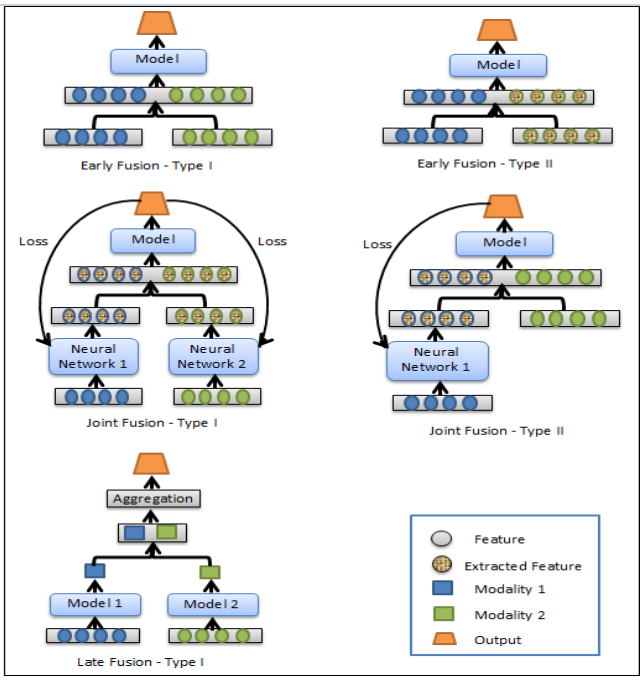

Fig. 8. The types of multimodal fusion.

- Early or Feature-Level Fusion: often referred to as feature-level fusion, it is the task of combining the input of different modalities into a single feature vector before it is presented into a single learning model. There are several techniques to combine input modalities, such as concatenation, pooling, or using a gated unit. Early fusion type I involves combining the original features, whereas type II involves merging extracted features or learned representations from another neural network. The anticipated probabilities are considered as extracted features, making the fusion of features and projected probabilities from several modalities another early type II fusion [107]. Feature-level fusion produced the most effective results for unimodal fusion and reduced processing time [108].
- Joint Fusion or Intermediate Fusion: The technique of combining learned representations of features from the in-between layers of NN with features from other modalities as input to an eventual model is known as joint fusion. The crucial distinction from early fusion is that during training, the loss is returned to the NN for feature extraction. This improves the representation of features for each training iteration. Neural networks are used in joint fusion because they pass on the loss from the prediction model to the feature extraction model. This joint fusion is type I, when feature representations from all modalities are extracted. The feature extraction stage does not always need to be classified as joint fusion for all input features.

- Late or Decision-Level Fusion: Late fusion, often referred to as decision-level fusion, is the task of using predictions from various models to arrive at a final decision. The final decision is typically achieved by using an aggregation function to combine the predictions of various models. Normally, diverse modalities are employed to train individual models. Averaging, weighted voting, majority voting, or a meta-classifier based on each model's predictions are several examples of aggregation functions. Based on the application and input modalities, the aggregation function is typically chosen empirically [107]. This method has the benefit of allowing each modality to learn its features using the best classifier for that modality [108].

VIII. Future Direction

Based on this review, we summarize some of the significant issues in this field that need to be addressed. In addition, we believe there is significant room for further improvement in fake news detection techniques. These issues are as follows:

1) Due to the use of either small or imbalanced datasets, detection models still suffer from significant challenges including underfitting, overfitting, and poor classification that degrade their performance.

2) Image-based features have not been widely used by previous studies in detecting fake news despite their highly critical effect.

3) Despite the vast use of vanilla GANs to generate new samples, they still suffer from some crucial issues, including vanishing gradients, mode collapse, and failure to converge.

4) Although pre-trained word embedding models are efficient at extracting features, they are not able to fully exploit the text's semantic and structural features.

5) Several machine-learning methods have been used to detect fake news. However, lower detection accuracy was provided.

6) The real significance of many modalities cannot be determined by simply concatenating the features. The unique features of each method (text and image) must be preserved while integrating relevant information between the different methods.

IX. Conclusion

This review article presents an overview of fake news, its types, and its consequences. The role of social media platforms in spreading fake news was also discussed. In addition, the most significant factors affecting fake news detection models were highlighted. Among these factors are the dataset, features, and supervised learning classifiers. The critical limitations that still need to be addressed were revealed by reviewing the most promising methods and techniques. These methods provided encouraging results in several areas that can be employed in fake news detection models. Moreover, these techniques were investigated and some of the challenges faced by them were described. This would allow future researchers to improve them and raise fake news detection accuracy.

Acknowledgment

This research was funded by the Universiti Kebangsaan Malaysia under Geran Universiti Penyelidikan (GUP), Grant Code: GUP-2020-088.

References

[1] De Souza, J.V., et al., A systematic mapping on automatic classification of fake news in social media. 2020. 10(1): p. 1-21.

[2] Xu, K., et al., Detecting fake news over online social media via domain reputations and content understanding. 2019. 25(1): p. 20-27.

[3] Atodiresei, C.-S., A. Tănăselea, and A.J.P.C.S. Iftene, Identifying fake news and fake users on Twitter. 2018. 126: p. 451-461.

[4] Habib, A., et al., False information detection in online content and its role in decision making: a systematic literature review. 2019. 9(1): p. 1-20.

[5] Hamed, S.K., M.J. Ab Aziz, and M.R.J.S. Yaakub, Fake News Detection Model on Social Media by Leveraging Sentiment Analysis of News Content and Emotion Analysis of Users' Comments. 2023. 23(4): p. 1748.

[6] Goksu, M. and N. Cavus. Fake news detection on social networks with artificial intelligence tools: systematic literature review. in International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions. 2019. Springer.

[7] Islam, M.R., et al., Deep learning for misinformation detection on online social networks: a survey and new perspectives. Soc Netw Anal Min, 2020. 10(1): p. 82.

[8] Braşoveanu, A.M. and R.J.N.P.L. Andonie, Integrating Machine Learning Techniques in Semantic Fake News Detection. 2020: p. 1-18.

[9] Aldwairi, M. and A.J.P.C.S. Alwahedi, Detecting fake news in social media networks. 2018. 141: p. 215-222.

[10] Cardoso Durier da Silva, F., R. Vieira, and A.C. Garcia. Can machines learn to detect fake news? a survey focused on social media. in Proceedings of the 52nd Hawaii International Conference on System Sciences. 2019.

[11] Sharma, K., et al., Combating fake news: A survey on identification and mitigation techniques. 2019. 10(3): p. 1-42.

[12] Pathak, A.R., et al., Analysis of techniques for rumor detection in social media. 2020. 167: p. 2286-2296.

[13] Vishwakarma, D.K. and C. Jain. Recent State-of-the-art of Fake News Detection: A Review. in 2020 International Conference for Emerging Technology (INCET). 2020. IEEE.

[14] Zhou, X. and R.J.A.C.S. Zafarani, A survey of fake news: Fundamental theories, detection methods, and opportunities. 2020. 53(5): p. 1-40.

[15] De Beer, D. and M.J.I.S.i.D.A. Matthee, Approaches to identify fake news: a systematic literature review. 2021: p. 13-22.

[16] Alam, F., et al., A survey on multimodal disinformation detection. 2021.

[17] Ansar, W. and S.J.I.J.o.I.M.D.I. Goswami, Combating the menace: A survey on characterization and detection of fake news from a data science perspective. 2021. 1(2): p. 100052.

[18] Li, J. and M.J.P.C.S. Lei, A Brief Survey for Fake News Detection via Deep Learning Models. 2022. 214: p. 1339-1344.

[19] Swapna, H. and B. Soniya. A Review on News-Content Based Fake News Detection Approaches. in 2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS). 2022. IEEE.

[20] Hu, L., et al., Deep learning for fake news detection: A comprehensive survey. 2022.

[21] Vosoughi, S., D. Roy, and S. Aral, The spread of true and false news online. Science, 2018. 359(6380): p. 1146-1151.

[22] Pogue, D., How to Stamp Out Fake News, in Sci Am. 2017. p. 24.

[23] Wang, A.B.J.W.P., Post-truth'named 2016 word of the year by Oxford Dictionaries. 2016. 16.

[24] Rapoza, K.J.F.N., Can ‘fake news’ impact the stock market? 2017.

[25] Rubin, V.L., On deception and deception detection: Content analysis of computer-mediated stated beliefs. Proceedings of the American Society for Information Science, 2010. 47(1): p. 1-10.

[26] Boehm, L.E.J.P. and S.P. Bulletin, The validity effect: A search for mediating variables. 1994. 20(3): p. 285-293.

[27] Fisher, R.J., Social desirability bias and the validity of indirect questioning. Journal of consumer research, 1993. 20(2): p. 303-315.

[28] Nickerson, R.S.J.R.o.g.p., Confirmation bias: A ubiquitous phenomenon in many guises. 1998. 2(2): p. 175-220.

[29] Metzger, M.J., E.H. Hartsell, and A.J. Flanagin, Cognitive dissonance or credibility? A comparison of two theoretical explanations for selective exposure to partisan news. Communication Research, 2020. 47(1): p. 3-28.

[30] Leibenstein, H.J.T.q.j.o.e., Bandwagon, snob, and Veblen effects in the theory of consumers' demand. 1950. 64(2): p. 183-207.

[31] Alkhodair, S.A., et al., Detecting breaking news rumors of emerging topics in social media. 2020. 57(2): p. 102018.

[32] Rath, B., et al., Utilizing computational trust to identify rumor spreaders on Twitter. 2018. 8(1): p. 1-16.

[33] Aldayel, A. and W.J.P.o.t.A.o.H.-C.I. Magdy, Your stance is exposed! analysing possible factors for stance detection on social media. 2019. 3(CSCW): p. 1-20.

[34] Kaur, S., P. Kumar, and P.J.S.C. Kumaraguru, Automating fake news detection system using multi-level voting model. 2020. 24(12): p. 9049-9069.

[35] Zhou, X., et al., Fake news early detection: A theory-driven model. 2020. 1(2): p. 1-25.

[36] Mridha, M.F., et al., A comprehensive review on fake news detection with deep learning. 2021. 9: p. 156151-156170.

[37] Nirav Shah, M., A.J.S.N.A. Ganatra, and Mining, A systematic literature review and existing challenges toward fake news detection models. 2022. 12(1): p. 168.

[38] Sansonetti, G., et al., Unreliable Users Detection in Social Media: Deep Learning Techniques for Automatic Detection. 2020. 8: p. 213154-213167.

[39] Ahmad, I., et al., Fake News Detection Using Machine Learning Ensemble Methods. 2020. 2020.

[40] Lin, L. and Z. Chen, Social rumor detection based on multilayer transformer encoding blocks. J Concurrency Computation: Practice Experience, 2021. 33(6): p. e6083.

[41] Hajek, P., et al., Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining. 2020. 32(23): p. 17259-17274.

[42] Zimbra, D., et al., The state-of-the-art in Twitter sentiment analysis: A review and benchmark evaluation. J ACM Transactions on Management Information Systems, 2018. 9(2): p. 1-29.

[43] Seo, S., et al., Comparative study of deep learning-based sentiment classification. 2020. 8: p. 6861-6875.

[44] Elhadad, M.K., K.F. Li, and F. Gebali, Detecting Misleading Information on COVID-19. IEEE Access, 2020. 8: p. 165201-165215.

[45] Varshney, D. and D.K.J.J.o.A.I. Vishwakarma, Hoax news-inspector: a real-time prediction of fake news using content resemblance over web search results for authenticating the credibility of news articles. 2020: p. 1-14.

[46] Vidgen, B., T.J.J.o.I.T. Yasseri, and Politics, Detecting weak and strong Islamophobic hate speech on social media. 2020. 17(1): p. 66-78.

[47] De Oliveira, N.R., D.S. Medeiros, and D.M.J.I.S.P.L. Mattos, A sensitive stylistic approach to identify fake news on social networking. 2020. 27: p. 1250-1254.

[48] Subramani, S., et al., Deep learning for multi-class identification from domestic violence online posts. 2019. 7: p. 46210-46224.

[49] Singh, R., et al., Deep learning for multi-class antisocial behavior identification from Twitter. 2020. 8: p. 194027-194044.

[50] Al-Sarem, M., et al., Deep learning-based rumor detection on microblogging platforms: a systematic review. 2019. 7: p. 152788-152812.

[51] Suhaimi, N.S., Z. Othman, and M.R. Yaakub. Comparative Analysis Between Macro and Micro-Accuracy in Imbalance Dataset for Movie Review Classification. in Proceedings of Seventh International Congress on Information and Communication Technology: ICICT 2022, London, Volume 3. 2022. Springer.

[52] Eke, C.I., et al., Sarcasm identification in textual data: systematic review, research challenges and open directions. 2020. 53(6): p. 4215-4258.

[53] Kumar, A., et al., Sarcasm detection using multi-head attention based bidirectional LSTM. 2020. 8: p. 6388-6397.

[54] Kim, Y., et al., Do Many Models Make Light Work? Evaluating Ensemble Solutions for Improved Rumor Detection. 2020. 8: p. 150709-150724.

[55] Bhutani, B., et al. Fake news detection using sentiment analysis. in 2019 twelfth international conference on contemporary computing (IC3). 2019. IEEE.

[56] Faustini, P.H.A. and T.F.J.E.S.w.A. Covões, Fake news detection in multiple platforms and languages. 2020. 158: p. 113503.

[57] Li, Q., et al., Multi-level word features based on CNN for fake news detection in cultural communication. 2019: p. 1-14.

[58] Guo, M., et al., An Adaptive deep transfer learning model for rumor detection without sufficient identified rumors. 2020. 2020.

[59] Gadek, G. and P.J.P.C.S. Guélorget, An interpretable model to measure fakeness and emotion in news. 2020. 176: p. 78-87.

[60] Kaliyar, R.K., A. Goswami, and P. Narang, EchoFakeD: improving fake news detection in social media with an efficient deep neural network. Neural Comput Appl, 2021. 33(14): p. 8597-8613.

[61] Alameri, S.A. and M. Mohd. Comparison of fake news detection using machine learning and deep learning techniques. in 2021 3rd International Cyber Resilience Conference (CRC). 2021. IEEE.

[62] Wang, W.Y.J.a.p.a., " liar, liar pants on fire": A new benchmark dataset for fake news detection. 2017.

[63] Ghanem, B., P. Rosso, and F. Rangel. Stance detection in fake news a combined feature representation. in Proceedings of the first workshop on fact extraction and VERification (FEVER). 2018.

[64] Kumar, S., et al., Fake news detection using deep learning models: A novel approach. 2020. 31(2): p. e3767.

[65] Shu, K., et al., FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media. Big Data, 2020. 8(3): p. 171-188.

[66] Raza, S., C.J.I.J.o.D.S. Ding, and Analytics, Fake news detection based on news content and social contexts: a transformer-based approach. 2022. 13(4): p. 335-362.

[67] Elhadad, M.K., K.F. Li, and F. Gebali. A novel approach for selecting hybrid features from online news textual metadata for fake news detection. in International Conference on P2P, Parallel, Grid, Cloud and Internet Computing. 2019. Springer.

[68] Segura-Bedmar, I. and S.J.I. Alonso-Bartolome, Multimodal fake news detection. 2022. 13(6): p. 284.

[69] Singhal, S., et al. Spotfake: A multi-modal framework for fake news detection. in 2019 IEEE fifth international conference on multimedia big data (BigMM). 2019. IEEE.

[70] Kalra, S., et al. Multimodal Fake News Detection on Fakeddit Dataset Using Transformer-Based Architectures. in Machine Learning, Image Processing, Network Security and Data Sciences: 4th International Conference, MIND 2022, Virtual Event, January 19–20, 2023, Proceedings, Part II. 2023. Springer.

[71] Shrivastava, G., et al., Defensive modeling of fake news through online social networks. 2020. 7(5): p. 1159-1167.

[72] Elhadad, M.K., K.F. Li, and F. Gebali. Fake news detection on social media: a systematic survey. in 2019 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM). 2019. IEEE.

[73] Bahad, P., P. Saxena, and R.J.P.C.S. Kamal, Fake news detection using bi-directional LSTM-recurrent neural network. 2019. 165: p. 74-82.

[74] Liu, S., K. Lee, and I.J.K.-B.S. Lee, Document-level multi-topic sentiment classification of email data with bilstm and data augmentation. 2020. 197: p. 105918.

[75] Moreno-Barea, F.J., J.M. Jerez, and L.J.E.S.w.A. Franco, Improving classification accuracy using data augmentation on small data sets. 2020. 161: p. 113696.

[76] Elizar, E., M.A. Zulkifley, and R. Muharar. Scaling and Cutout Data Augmentation for Cardiac Segmentation. in Proceedings of International Conference on Data Science and Applications: ICDSA 2022, Volume 2. 2023. Springer.

[77] Bejani, M.M. and M.J.a.p.a. Ghatee, Regularized deep networks in intelligent transportation systems: A taxonomy and a case study. 2019.

[78] Sun, X., J.J.M.T. He, and Applications, A novel approach to generate a large scale of supervised data for short text sentiment analysis. 2020. 79(9): p. 5439-5459.

[79] Chlap, P., et al., A review of medical image data augmentation techniques for deep learning applications. 2021. 65(5): p. 545-563.

[80] Cha, D. and D. Kim. DAM-GAN: Image Inpainting Using Dynamic Attention Map Based on Fake Texture Detection. in ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2022. IEEE.

[81] Yu, L., et al. Seqgan: Sequence generative adversarial nets with policy gradient. in Proceedings of the AAAI conference on artificial intelligence. 2017.

[82] Guo, J., et al. Long text generation via adversarial training with leaked information. in Proceedings of the AAAI conference on artificial intelligence. 2018.

[83] Kumar, G. and P.K. Bhatia. A detailed review of feature extraction in image processing systems. in 2014 Fourth international conference on advanced computing & communication technologies. 2014. IEEE.

[84] Ahmad, S.R., A.A. Bakar, and M.R.J.I.d.a. Yaakub, A review of feature selection techniques in sentiment analysis. 2019. 23(1): p. 159-189.

[85] Latiffi, M.I.A., et al., Flower Pollination Algorithm for Feature Selection in Tweets Sentiment Analysis. 2022. 13(5).

[86] Wang, J.-H., T.-W. Liu, and X.J.A.S. Luo, Combining Post Sentiments and User Participation for Extracting Public Stances from Twitter. 2020. 10(22): p. 8035.

[87] Deepak, S. and B.J.P.C.S. Chitturi, Deep neural approach to Fake-News identification. 2020. 167: p. 2236-2243.

[88] Vicari, M., M.J.A. Gaspari, and Society, Analysis of news sentiments using natural language processing and deep learning. 2020: p. 1-7.

[89] Subramani, S., et al., Domestic violence crisis identification from facebook posts based on deep learning. 2018. 6: p. 54075-54085.

[90] Batbaatar, E., M. Li, and K.H.J.I.A. Ryu, Semantic-emotion neural network for emotion recognition from text. 2019. 7: p. 111866-111878.

[91] Yenicelik, D., F. Schmidt, and Y. Kilcher. How does BERT capture semantics? A closer look at polysemous words. in Proceedings of the Third BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP. 2020.

[92] AlShariah, N.M., et al., Detecting fake images on social media using machine learning. 2019. 10(12): p. 170-176.

[93] Jwa, H., et al., exbake: Automatic fake news detection model based on bidirectional encoder representations from transformers (bert). 2019. 9(19): p. 4062.

[94] Lochter, J.V., R.M. Silva, and T.A. Almeida. Deep learning models for representing out-of-vocabulary words. in Brazilian Conference on Intelligent Systems. 2020. Springer.

[95] Fernández-Martínez, F., et al., Fine-Tuning BERT Models for Intent Recognition Using a Frequency Cut-Off Strategy for Domain-Specific Vocabulary Extension. 2022. 12(3): p. 1610.

[96] Devlin, J., et al., Bert: Pre-training of deep bidirectional transformers for language understanding. 2018.

[97] Choudhary, A. and A. Arora. ImageFake: An Ensemble Convolution Models Driven Approach for Image Based Fake News Detection. in 2021 7th International Conference on Signal Processing and Communication (ICSC). 2021. IEEE.

[98] Simonyan, K. and A.J.a.p.a. Zisserman, Very deep convolutional networks for large-scale image recognition. 2014.

[99] Wu, S.J.J.o.R., Expression Recognition Method Using Improved VGG16 Network Model in Robot Interaction. 2021. 2021: p. 1-9.

[100] Han, X., et al., Pre-trained models: Past, present and future. 2021. 2: p. 225-250.

[101] Liao, W.-X., et al., Automatic identification of breast ultrasound image based on supervised block-based region segmentation algorithm and features combination migration deep learning model. 2019. 24(4): p. 984-993.

[102] Zhang, C., et al., Multimodal intelligence: Representation learning, information fusion, and applications. 2020. 14(3): p. 478-493.

[103] Che, C., et al., Hybrid multimodal fusion with deep learning for rolling bearing fault diagnosis. 2021. 173: p. 108655.

[104] Zhang, Y., et al., Deep multimodal fusion for semantic image segmentation: A survey. 2021. 105: p. 104042.

[105] Zhang, S., B. Li, and C.J.S. Yin, Cross-Modal Sentiment Sensing with Visual-Augmented Representation and Diverse Decision Fusion. 2021. 22(1): p. 74.

[106] Person, M., et al., Multimodal fusion object detection system for autonomous vehicles. 2019. 141(7).

[107] Huang, S.-C., et al., Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. 2020. 3(1): p. 1-9.

[108] Chandrasekaran, G., et al., Multimodal sentimental analysis for social media applications: A comprehensive review. 2021. 11(5): p. e1415.