

Application of Artificial Intelligence Techniques to Detect Fake News: A Review

Penerapan Teknik Kecerdasan Buatan untuk Mendeteksi Berita Palsu: Sebuah Tinjauan

Muthnna Nihad Al-Tameemi
Muthana.altmeme.28@gmail.com

Abstract. General Background: The rapid expansion of social media platforms and online news has led to an increase in fake news, impacting the accuracy and reliability of disseminated information. **Specific Background:** As machine learning excels at handling large datasets and pattern recognition, it has become a prominent tool in fake news detection by identifying signs of misinformation through textual content, social context, and network structures. **Knowledge Gap:** Although recent advances exist, challenges persist in handling different data types and optimizing model performance, as well as ensuring accuracy across languages and platforms. **Aims:** This article aims to review the latest advancements in fake news detection through machine learning, focusing on various techniques like data mining, deep learning, natural language processing (NLP), ensemble learning, transfer learning, and graph-based approaches. **Results:** The findings show that each technique has specific advantages and limitations; for example, deep learning uncovers complex linguistic features, while ensemble learning increases accuracy through combined models, and graph-based techniques leverage social network analysis but lack linguistic context. **Novelty:** The review identifies emerging trends, including hybrid models that combine NLP, graph-based, and deep learning approaches for broader, multi-perspective detection, as well as optimization techniques using precomputed data and pre-trained algorithms. **Implications:** The discussion highlights the potential for these combined approaches to enhance fake news detection accuracy and efficiency on social media, suggesting that integrating these methods with quantum computing advancements in NLP could further improve model performance. In summary, the article provides a comprehensive analysis of the machine learning landscape in fake news detection and recommends future directions for developing more scalable, effective solutions.

Keywords – Fake News Detection, Machine Learning, Social Media, Natural Language Processing, Graph-Based Techniques

Abstrak. Latar Belakang Umum: Perluasan platform media sosial dan berita daring yang pesat telah menyebabkan peningkatan berita palsu, yang memengaruhi akurasi dan keandalan informasi yang disebarluaskan. **Latar Belakang Khusus:** Karena pembelajaran mesin unggul dalam menangani kumpulan data besar dan pengenalan pola, pembelajaran mesin telah menjadi alat yang menonjol dalam pendeteksian berita palsu dengan mengidentifikasi tanda-tanda misinformasi melalui konten tekstual, konteks sosial, dan struktur jaringan. **Kesenjangan Pengetahuan:** Meskipun ada kemajuan terkini, tantangan tetap ada dalam menangani berbagai jenis data dan mengoptimalkan kinerja model, serta memastikan akurasi di seluruh bahasa dan platform. **Tujuan:** Artikel ini bertujuan untuk meninjau kemajuan terbaru dalam pendeteksian berita palsu melalui pembelajaran mesin, dengan fokus pada berbagai teknik seperti penambangan data, pembelajaran mendalam, pemrosesan bahasa alami (NLP), pembelajaran ansambel, pembelajaran transfer, dan pendekatan berbasis grafik. **Hasil:** Temuan menunjukkan bahwa setiap teknik memiliki kelebihan dan keterbatasan tertentu; misalnya, pembelajaran mendalam mengungkap fitur linguistik yang kompleks, sementara pembelajaran ansambel meningkatkan akurasi melalui model gabungan, dan teknik berbasis grafik memanfaatkan analisis jaringan sosial tetapi tidak memiliki konteks linguistik. **Kebaruan:** Tinjauan ini mengidentifikasi tren yang muncul, termasuk model hibrida yang menggabungkan pendekatan NLP, berbasis grafik, dan pembelajaran mendalam untuk deteksi multiperspektif yang lebih luas, serta teknik pengoptimalan menggunakan data yang telah dihitung sebelumnya dan algoritme yang telah dilatih sebelumnya. **Implikasi:** Pembahasan ini menyoroti potensi pendekatan gabungan ini untuk meningkatkan akurasi dan efisiensi deteksi berita palsu di media sosial, yang menunjukkan bahwa mengintegrasikan metode ini dengan kemajuan komputasi kuantum dalam NLP dapat lebih meningkatkan kinerja model. Singkatnya, artikel ini memberikan analisis komprehensif tentang lanskap

pembelajaran mesin dalam deteksi berita palsu dan merekomendasikan arah masa depan untuk mengembangkan solusi yang lebih terukur dan efektif.

Kata Kunci – *Deteksi Berita Palsu, Pembelajaran Mesin, Media Sosial, Pemrosesan Bahasa Alami, Teknik Berbasis Grafik*

I. INTRODUCTION

Fake news refers to intentionally false or misleading information presented as if it were true and presented as news. This can include fabricated news stories, doctored images or videos, and even legitimate news stories that are taken out of context or presented with misleading headlines or captions. The goal of fake news is often to deceive people, generate clicks or advertising revenue, or influence public opinion [1]

Fake news has become a significant concern in today's digital age. With the rise of social media and online news platforms, fake news has become a pervasive problem, spreading misinformation and propaganda to millions of people worldwide. The consequences of fake news can be severe, causing confusion, polarization, and even violence. In recent years, fake news has been linked to major events such as the Brexit vote [2], the 2016 US presidential election [3], and the COVID-19 pandemic [4]. In 2016, the Oxford Dictionary defined post-truth as the word of the year [5]. In each case, false information and conspiracy theories were shared widely, leading to public mistrust in institutions and damaging consequences. Last, the new open AI tool, ChatGPT4, is helping this automatic fake news generation due to its generative component and has used fake articles that were never written by journalists like The Guardian [6]. This tool is likely to become our new source of truth and research, so it is extremely worrying that fake news is already there.

After analyzing some of the consequences that fake news has in our society and the consequences that misinformation has had, plus the potential it has in future crises, it is more than obvious that being able to detect fake news is critical. Before the social media era, newspapers and journals were in charge of the information, and are nowadays too, but their power has been reduced, especially in some age ranges. According to [7], journalists must rigorously check facts, confirm information with multiple sources, and ensure the accuracy of their reporting before publication. But in a "culture of immediacy" where we consume information so fast, this checking takes a lot of time and effort. Several organizations provide rumor tools such as [8]. However, since the amount of news and information is so big, they do not tackle the issue in a satisfactory way.

Given the challenges posed by fake news, especially the fast pace of fake news generation, the spreadability, and the high number of this type of news, there has been a growing interest in developing automated methods for detecting it. Machine learning techniques, in particular, have shown great promise in detecting fake news due to their ability to handle large data volumes. Another promising feature that machine learning algorithms bring to this field is the idea of learning from past data and hidden patterns. Once fake news is detected, a news article with a similar spreadability pattern can be identified later on without having to do all the time-consuming research. Many journals are exploring these techniques and have had relative success detecting fake news as [9] explained for The Guardian.

In this paper, we review the state of the art in fake news detection from a machine learning techniques perspective, but we do not consider other traditional techniques. We first provide an overview of the problem of fake news, discussing its causes, consequences, and challenges. We review recent research on machine learning-based approaches for detecting fake news, including deep learning, natural language processing, and graph-based methods. Finally, we discuss future research directions and the potential impact of fake news detection on society.

The structure of this paper is as follows. Section 2 presents the results in three different subsections. Section 2.1 explores different machine learning techniques that have been used in detecting fake news fields. Section 2.2 presents data types and classification of the papers according to the different data sources they use. Section 3 provides the literature review, including a discussion of the methods used, the papers selected, and the results. In Section 4, the papers are analyzed and discussed in depth. Finally, in Section 5, the conclusions are given.

II. METHODS

Machine Learning Techniques

According to [10], machine learning is defined as a subfield of artificial intelligence that focuses on the development of algorithms and models that allow computer systems to automatically learn and improve from experience without being explicitly programmed. Machine learning techniques are designed to analyze and extract patterns from data, allowing the system to recognize and understand complex relationships and make predictions or decisions based on that knowledge. These techniques often involve the use of statistical methods and algorithms to

train models on labeled or unlabeled data, enabling the system to generalize and make accurate predictions or classifications on new, unseen data.

Ref. [11] defined two main fields inside machine learning: supervised learning and unsupervised learning. Supervised learning involves training a model using labeled data where the input variables (features) are associated with known output labels. The goal is to learn a mapping function that can predict the labels for new, unseen data. Unsupervised learning deals with learning from unlabeled data where the objective is to discover hidden patterns, structures, or relationships within the data. Depending on the data type, different machine learning algorithms are applied. Last, there is a subfield mentioned by [11] named semi-supervised learning that lies between supervised and unsupervised learning, which utilizes a combination of labeled and unlabeled data to train models.

These are not the only techniques that can be used to resolve the fake news problem in large amounts of data but are the ones this article will be focusing on. As proposed by [12], metaheuristics can also be used to resolve these challenges.

Deep Learning Techniques

Deep learning techniques can be defined as a subset of machine learning methods that are inspired by the structure and function of the human brain, specifically artificial neural networks with multiple layers. These techniques are capable of automatically learning hierarchical representations of data, enabling them to extract complex patterns and make accurate predictions [13].

Deep learning is a trendy technique that uses artificial neural networks with multiple layers to learn and identify patterns in data. It has been widely used for fake news detection by analyzing the linguistic features of the news articles.

Natural Language Processing Techniques (NLP)

According to [14], an NLP technique refers to a computational method or algorithm that enables computers to process, analyze, and understand human language. These techniques encompass a wide range of tasks, including but not limited to text classification, information extraction, sentiment analysis, machine translation, and question answering. NLP techniques often involve the use of statistical models, machine learning algorithms, and linguistic resources to tackle the inherent complexities and ambiguities of natural language.

It has been used to analyze the semantic and syntactic features of news articles to detect fake news.

The field of NLP has significantly evolved thanks to BERT, which introduced the ability to capture contextualized word representations and understand the nuances of language, which, in turn, has led to improvements in various downstream tasks, such as sentiment analysis, named entity recognition, and machine translation. Ref. [15] have investigated what linguistic information BERT learns during pre-training to understand how it captures syntax and structure in language. Ref. [16] have focused on syntactic transfer, which involves training a model in one language and applying it to another, exploring the transferability of syntactic knowledge across languages.

Ref. [17] provide a detailed review of challenges in natural language processing (NLP)-based online fake news detection. Ref. [18] propose a transformer-based method (a similar approach to [16]) for detecting fake news in multilingual contexts, with a specific focus on languages with limited resources.

Ensemble Learning

“Ensemble methods involve constructing a set of base learners from the training data, and then combining their predictions at test time to create a single, stronger learner”. [11]. This machine learning approach combines multiple individual models, called base learners, to make more accurate predictions or decisions. Each base learner in an ensemble is trained independently, and their outputs are combined to generate a final prediction or decision. The idea behind ensemble learning is that by aggregating the predictions of multiple diverse models, the overall performance can be improved, often surpassing the performance of any individual model.

It has been used for fake news detection by combining the outputs of different algorithms to improve the detection accuracy. They have the ability to resolve the weaknesses of individual models and reduce bias.

Transfer Learning

Ref. [19] describes the limitation of traditional machine learning because it needs training data and testing data to have the same input feature space and the same data distribution. When there is a difference in data distribution between the training data and test data, the results of a predictive learner can be degraded. In certain scenarios, obtaining training data that matches the feature space and predicted data distribution characteristics of the test data can be difficult and expensive. Therefore, there is a need to create a high-performance learner for a target domain trained from a related source domain. This is the motivation for transfer learning.

Transfer learning is used to improve a learner from one domain by transferring information from a related domain. We can draw from real-world nontechnical experiences to understand why transfer learning is possible.

This technique allows the use of pre-trained models for new tasks. It has been used for fake news detection by using pre-trained models for NLP tasks and fine-tuning them for fake news detection.

Graph-Based Techniques

A graph-based technique is an approach that leverages graph structures to represent and analyze data where the data are represented as nodes or vertices connected by edges. Graph-based techniques have gained significant attention due to their ability to capture and exploit the inherent relationships and dependencies within complex datasets. By modeling data as graphs, these techniques enable the exploration of connectivity patterns, community structures, and network properties.

Ref. [20] argues that machine learning and machine learning on graphs are both problem-driven disciplines that seek to build models that can learn from data in order to solve particular tasks. The usual categories of supervised and unsupervised are not necessarily the most informative or useful when it comes to graphs. The following four types of problems are defined: node classification, relation prediction, clustering and community detection, and graph classification, regression, and clustering.

Graph-based algorithms are a trendy technique that represents news articles as graphs and analyzes the relationships between the nodes in the graph to detect fake news. These algorithms have shown promising results for fake news detection.

Data Types

Ref. [21] classifies the data types used in fake news detection on social media as follows: content-based data, which refer to the textual and visual content of the news, including news headlines, text, images, and videos; social context data, which are social media data associated with the news, including user profiles, follower–friend relationships, activity logs, and comments; and, network structure data, which name the structural information of the social media network, including user–user relationships and interactions.

Different machine learning approaches are more suited to different types of data. For example, content-based data can be effectively analyzed using natural language processing techniques such as word embeddings and recurrent neural networks. Another popular NLP approach to detect anomalies in unstructured text is Word2vec [22]. Social context data can be used to build user profiles and detect patterns in social networks, which can be used to identify fake news sources. Network structure data can be analyzed using graph-based approaches, such as community detection and centrality measures.

Understanding the different types of data used for fake news detection and selecting appropriate machine learning techniques can lead to more accurate and efficient detection of fake news on social media.

In addition to the data types, there is also extensive research generating different datasets to compare all the fake news techniques. Ref. [23] introduce the FA-KES dataset, which is a collection of fake news related to the Syrian war. Ref. [24] provide a span detection dataset detecting opinion spam and fake news through text classification techniques. The dataset described in [25] is designed for studying fake news in the context of the COVID-19 pandemic.

Last, from a linguistic point of view, the domain to which the fake news detection task is applied will have large implications. User-generated content is very different from journalistic texts.

Literature Review

Characteristics of the Literature Review

The purpose of this paper is to collect the state of the art when it comes to detecting fake news with machine learning techniques. In order to do so, the following engines were used: Google Scholar and ResearchRabbit. The search terms were [fake news detection, social media, data mining, deep learning, natural language processing, ensemble learning, transfer learning, and graph-based approach]. The papers that met the following criteria were selected: use of machine learning techniques to solve problems in fake news and written in English.

The following methodology was used: search for all those terms in Google Scholar and discard the ones that did not fit our criteria. After this, we used ResearchRabbit to expand and look into the papers that reference these initial papers and apply the filtering criteria again. This concluded with a set of 11 papers between 2017 and 2022. The following table summarizes it.

Materials

In Table 1 we can see a summary of all the selected articles. A total of 11 articles were selected. The articles were classified according to two different criteria: data type and the machine learning algorithm. When it comes to the data type, besides the state-of-the-art article, which is the one that creates the classification, we have at least two articles per data type. The classification is not strict, so an article can have more than one data type included. Refs. [17,26–29] are based on content data: they analyze the text of the news to determine whether it is fake or not. Some others, such as [30,31], leverage user profile and relationship information to enrich the dataset while they explore the content

of the news as well. Last, refs. [32,33] leverage pure network data. Most of the articles are based on content-based data (8 out of 10).

When it comes to machine learning techniques, we have explored five techniques with two papers per technique. Refs. [30,34] explore deep learning techniques, refs. [17,26] explore natural language processing, refs. [27,31] focus on ensemble learning techniques, refs. [28,29] explain transfer learning, while [32,33] explore graph-based techniques. Once again, ref. [21] is a state-of-the-art paper that does not focus on any specific technique and analyzes the state of the literature in 2017.

Table 1. Summary of the analyzed papers

Articles	Data Source	Machine Learning Technique	Date
Shu et al., 2017 [21]	This paper creates the categorization	Social media data mining	2017
Thota et al., 2018 [34]	Content-based data	Deep learning	2018
Monti et al., 2019 [30]	Network data + content + social	Deep learning	2019
Hirlekar and Kumar, 2020 [17]	Content-based data	Natural language processing	2020
Oshikawa et al., 2018 [26]	Content-based data	Natural language processing	2018
Ahmad et al., 2020 [31]	Content-based + social context data	Ensemble learning	2020
Agarwal and Dixit, 2020 [27]	Content-based data	Ensemble learning	2020

Table 1. Cont.

Articles	Data Source	Machine Learning Technique	Date
Saikh et al., 2020 [28]	Content-based data	Transfer learning	2020
Tida et al., 2022 [29]	Content-based data	Transfer learning	2022
Chandra et al., 2020 [32]	Network data	Graph-based	2020
Gangireddy et al., 2020 [33]	Network data	Graph-based	2020

III. RESULT AND DISCUSSION

Table 2 represents the summary of each of the machine learning techniques presented in each of the papers. The main line of work represents where the techniques belong in the previous machine learning technique categorization. Inside each of these fields, there are several techniques that the paper applied. The detection approach is the summary of the paper—the one-liner that relates each of the papers from a machine learning perspective and answers the question: How does this paper detect fake news?

Table 2. Summary of the machine learning techniques used in each paper.

Articles	Main Line of Work	Techniques Used	Detection Approach
Shu et al., 2017 [21]	Social media data mining	Preprocessing, feature extraction, and classification algorithms	Social media data mining
Thota et al., 2018 [34]	Deep learning	Deep neural networks (e.g., CNNs, RNNs)	Neural network-based
Monti et al., 2019 [30]	Deep learning	Graph representations, feature extraction	Geometric deep learning
Hirlekar and Kumar, 2020 [17]	Natural language processing	Text preprocessing, sentiment analysis, and topic modeling	Natural language processing-based
Oshikawa et al., 2018 [26]	Natural language processing	Various natural language processing techniques	Natural language processing-based
Ahmad et al., 2020 [31]	Ensemble learning	Combination of machine learning algorithms	Ensemble learning
Agarwal and Dixit, 2020 [27]	Ensemble learning	Novel ensemble methods or variation	Ensemble learning
Saikh et al., 2020 [28]	Transfer learning	Transfer learning using source domain knowledge	Transfer learning
Tida et al., 2022 [29]	Transfer learning	Fine-tuning BERT model	Unified training
Chandra et al., 2020 [32]	Graph-based	Graph modeling, community analysis	Graph-based
Gangireddy et al., 2020 [33]	Graph-based	Unsupervised graph-based methods	Graph-based

Based on the classification in Table 2, the articles are grouped into similar categories or approaches.

From a data mining and deep learning perspective, ref. [21] present a comprehensive analysis of various data mining techniques and their application in identifying fake news on social media. This article focuses on utilizing data mining techniques to detect fake news specifically on social media platforms. It involves preprocessing the data, extracting relevant features, and applying classification algorithms to identify fake news patterns within social media data. This study emphasizes the importance of feature engineering and classification algorithms for accurate detection. Similarly, using deep learning as a base [34] focuses on the application of deep learning methods. It utilizes deep neural networks, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to automatically learn complex patterns and representations in the data. The focus is on leveraging neural network-based techniques to detect and classify fake news. These approaches leverage the inherent patterns and representations in textual data to achieve high detection accuracy.

Both [30,34] leverage techniques of the deep learning field, but the way data are preprocessed for each of the approaches is very different.

Ref. [30] explores the application of geometric deep learning techniques for fake news detection on social media. It involves representing social media data as graphs and extracting relevant features from these graphs. Geometric deep learning methods are then employed to analyze the graph structure and identify fake news.

Ref. [32] adopts a graph-based approach to detect fake news within online communities. It involves constructing graph models that represent the relationships between users and content and analyzing the community structure within these graphs. The focus is on utilizing graph-based techniques to identify patterns and anomalies indicative of fake news. Both [30,32] model the relationships between users and their interactions while [33] focus on unsupervised methods and graph-based methods that do not rely on labeled data for training. The approach involves analyzing the structure and characteristics of the graph to identify suspicious patterns and detect fake news.

Regarding the challenges in natural language processing (NLP) [17,26] address the specific challenges and opportunities in utilizing NLP techniques for fake news detection. These challenges include the detection of subtle linguistic cues, the identification of context-specific features, and the handling of noisy and unstructured text data. The surveyed approaches propose various NLP-based solutions, such as sentiment analysis, topic modeling, and stance detection, to effectively tackle these challenges. Ref. [17] focus more on preprocessing the text, conducting sentiment analysis, and applying topic modeling to uncover patterns indicative of fake news. They propose different action types depending on the problem each case is trying to solve and the quality of the insight (novelty or uncover

patterns are not always fake news), while [26] cover a wider range of NLP techniques that have been employed in this domain.

Refs. [27,31] propose the use of ensemble learning techniques for fake news detection. These approaches combine multiple classifiers or models to improve the overall performance and robustness of the detection system. While [31] focus on utilizing ensemble learning techniques to detect and classify fake news, ref. [27] introduce novel ensemble methods or variations that aim to improve the accuracy and robustness of the detection process, specifically designing an ensemble method for fake news.

Last, the article by [28] focuses on applying transfer learning techniques to fake news detection. It utilizes knowledge and models pre-trained on a source domain to improve the detection performance on the target domain. The approach involves transferring learned representations from the source domain to enhance the detection of fake news. While [29] propose a unified training process specifically for fake news detection. It focuses on fine-tuning a BERT (Bidirectional Encoder Representations from Transformers) model, which is a state-of-the-art language model, to improve its performance in detecting fake news. The approach involves a unified training pipeline specifically designed for fake news detection.

Discussion

In this section, we will discuss the results. There is a summary table, Table 3, with the pros and cons of each technique.

Table 3. Summary of the pros and cons of each technique.

Articles	Main Line of Work	Pros	Cons
Shu et al., 2017 [21]	Social media data mining	<ul style="list-style-type: none"> - Large volumes of data - User behavior analysis 	<ul style="list-style-type: none"> - Data quality and availability - Difficulty distinguishing patterns - Computation
Monti et al., 2019 and Thota et al., 2018 [30,34]	Deep learning	<ul style="list-style-type: none"> - Learning complex patterns - Semantic and contextual information - Relational + graph-based features 	<ul style="list-style-type: none"> - Labeled data - Computation - Interpretability
Hirlekar and Kumar, 2020 and Oshikawa et al., 2018 [17,26]	Natural language processing	<ul style="list-style-type: none"> - Leveraging linguistic patterns - Enabling sentiment analysis, topic modeling, and textual analysis techniques 	<ul style="list-style-type: none"> - Only textual context - Language ambiguity and context understanding
Agarwal and Dixit, 2020 and Ahmad et al., 2020 [27,31]	Ensemble learning	<ul style="list-style-type: none"> - Combination of multiple models - Mitigating individual limitation - Reducing bias 	<ul style="list-style-type: none"> - Computation - Complex system
Saikh et al., 2020 and Tida et al., 2022 [28,29]	Transfer learning	<ul style="list-style-type: none"> - Leveraging pre-trained models - Reducing the need for labels - Simplifying the training pipeline 	<ul style="list-style-type: none"> - Availability of similar pre-trained models - Fine-tuning
Chandra et al., 2020 and Gangireddy et al., 2020 [32,33]	Graph-based	<ul style="list-style-type: none"> - Relational structure - Insights about communities 	<ul style="list-style-type: none"> - Data quality - Computation - Lack of textual context

Regarding social media data mining, this approach is planned for leveraging the vast amount of data that are available on social media. It also allows for user behavior analysis and analysis of the network dynamic and can be very insightful when it comes to explainability. It would explain the expected behavior of specific users and how the networks and users evolve. However, even if it is designed to be able to ingest very big data volumes, the quality, availability (if you do not work in a specific social media), and reliability of the data are an issue for this type of technique. As the data may not be ideal, and taking into account that we are working with user data, the algorithms may struggle to distinguish new information or a post that went viral from fake news. Last, the fact that they deal with such a big volume of data makes the training of an algorithm computationally very expensive.

When it comes to deep learning or learning complex patterns in data, these algorithms have very big potential. This is great for this case scenario since it could learn semantic and contextual information about the data that are provided to the algorithm. In the case of geometrical deep learning, there is also the potential for learning relational and graph-based features. However, it is quite hard to get the labeled data that are needed to train these systems, and this could be very expensive since automated labeling does not necessarily fit all the cases here and human-labeled data (and the challenges that these techniques propose) may need to be addressed. On the other hand, these systems are also computationally expensive to train. And finally, the interpretability of neural networks is a known challenge. We may get good results but not be able to explain why or what is going on behind the scenes.

Natural language processing is an emerging field in artificial intelligence. This technique allows us to identify various linguistic patterns behind fake news. It allows us to do sentiment analysis, topic modeling, and textual analysis that may easily explain whether a new being is classified as fake or not. However, we lose a lot of contextual information if this is the only technique used because we only focus on the text. We may be losing data, user-topic relationships, other multimedia content, etc. Moreover, there is the challenge of textual ambiguity—slang, irony, and sarcasm are examples of text that this type of model may find challenging to classify.

When it comes to ensemble learning, since it is a combination of techniques, the main benefit is that it resolves the weaknesses of individual models and outputs a better result. It also helps reduce bias and includes diverse perspectives since it does not rely on a single algorithm. However, this huge advantage comes with a big computational cost and a high complexity of the system. If training one algorithm is expensive, training six is way more expensive. When it comes to the complexity of the system you need to take care of multiple algorithms, their preprocessing, training, and online inference. This is the reason why often if the performance boost is not that big, a type of technique loses in the performance complexity trade-off.

Transfer learning is a promising technique when it comes to reducing complexity. Since it leverages pre-trained models, it simplifies the training pipeline and reduces the need for labels. It can also potentially improve the performance of the model through fine-tuning, but that process is quite costly most of the time. This technique relies on two principles that are quite hard to find when it comes to fake news detection: the availability of a suitable pre-trained model for fake news detection and the similarity of final data with the pre-trained data.

Finally, when it comes to graph-based approaches, they leverage the graph structure of social networks. This is really interesting when it comes to fake news because they can provide additional insights, such as the spread or the influence of fake news in different communities, which can be as valuable as precision in some scenarios. However, constructing this graph, besides being computationally expensive, is significantly related to the data quality, so it may not always be possible or it may not have a good enough quality. Moreover, this approach lacks the textual context of the data and only relies on the reaction of the community to analyze the news, which can lead to misleading conclusions.

There are two clear tendencies that can be inferred from the current state of the art: model combination and model optimization.

On the one hand, as shown in Table 3, the combination of graph-based methods and deep learning, textual data, or natural language preprocessing is a trend in this field. The combination of techniques allows us to have full context, from the natural language perspective and the graph perspective. It is likely that this hybrid approach will have better performance and cover more use cases compared with focusing on one single technique.

On the other hand, almost all algorithms are computationally expensive. This means that there is a lot of potential in optimizing these techniques. Since a lot of data are needed and all the iterations are expensive, any precomputed data, cached data, or leveraging a pre-trained algorithm for a specific section could significantly improve the machine learning system, albeit sometimes with a trade-off of the algorithm's accuracy.

Detecting fake news poses formidable challenges due to subtle linguistic cues, contextual ambiguity, and adversarial tactics employed by creators. The integration of multimodal analysis for non-textual content, robustness against adversarial attacks, and preprocessing techniques to manage noisy, unstructured data are crucial aspects. Furthermore, the scarcity of labeled data necessitates approaches like transfer learning. Addressing these challenges requires a comprehensive, multidimensional strategy that combines advanced natural language processing techniques, machine learning models capable of understanding context, and continuous adaptation to evolving deceptive tactics. An example of this tactic is provided in [35].

CONCLUSION

The reviewed articles on fake news detection have provided valuable insights into the advancements and challenges in this field. By exploring various techniques, including data mining, deep learning, natural language

processing, ensemble learning, transfer learning, and graph-based methods, researchers have made significant progress in identifying and combating fake news on social media platforms.

In conclusion, the discussion highlighted the pros and cons of various techniques for fake news detection. Social media data mining offers insights into user behavior and network dynamics but faces challenges with data quality, availability, and computational requirements. Deep learning has the potential to learn complex patterns and semantic information but is limited by the availability of labeled data, computational expenses, and interpretability. Natural language processing allows for linguistic pattern identification but loses contextual information and struggles with textual ambiguity. Ensemble learning resolves weaknesses of individual models but comes with high computational costs and system complexity. Transfer learning simplifies the training pipeline but relies on suitable pre-trained models and the similarity of data. Graph-based approaches provide insights into the spread and influence of fake news but require high-quality data and lack textual context.

Two trends emerge from the current state of the art: the combination of techniques and model optimization. Combining graph-based methods with natural language processing or deep learning allows for a more comprehensive approach, leveraging both textual and graph perspectives. This hybrid approach is expected to yield better performance and cover a wider range of use cases. Additionally, due to the computational expenses of these techniques, there is a significant potential for optimization. Leveraging precomputed data, cached data, or pre-trained algorithms for specific sections can improve the overall system, albeit at the cost of some accuracy trade-offs.

Overall, the field of fake news detection continues to evolve, with researchers exploring different techniques and seeking ways to improve performance, efficiency, and interpretability. The combination of approaches and optimization strategies will likely contribute to more effective and scalable solutions in the future.

Due to the relevance of advances made in the NLP field and the high impact they have had in the fake news detection field, it is worth noting there is a new trend in using quantum computing to improve the NLP model's performance. Refs. [36,37] state advances in the field.

REFERENCES

- [1]. B. G. Rannard, "How Fake News Plagued 2017," BBC News, Dec. 31, 2017. [Online]. Available: <https://www.bbc.com/news/world-42487425>. Accessed: Sep. 10, 2023.
- [2]. "Brexit: What You Need to Know about the UK Leaving the EU," BBC News, Dec. 30, 2020. [Online]. Available: <https://www.bbc.com/news/uk-politics-32810887>. Accessed: Sep. 10, 2023.
- [3]. N. Confessore, "Cambridge Analytica and Facebook: The Scandal and the Fallout So Far," The New York Times, Nov. 14, 2018. [Online]. Available: <https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html>. Accessed: Sep. 10, 2023.
- [4]. E. Lawrie and R. Schraer, "Coronavirus: Scientists Brand 5G Claims 'Complete Rubbish,'" BBC News, Apr. 15, 2020. [Online]. Available: <https://www.bbc.com/news/52168096>. Accessed: Sep. 10, 2023.
- [5]. "Oxford Word of the Year 2016," Oxford Languages, Jun. 16, 2020. [Online]. Available: <https://languages.oup.com/word-of-the-year/2016/>. Accessed: Sep. 10, 2023.
- [6]. C. Moran, "ChatGPT Is Making Up Fake Guardian Articles. Here's How We're Responding," The Guardian, Apr. 6, 2023. [Online]. Available: <https://www.theguardian.com/commentisfree/2023/apr/06/ai-chatgpt-guardian-technology-risks-fake-article>. Accessed: Sep. 10, 2023.
- [7]. B. Kovach and T. Rosenstiel, *The Elements of Journalism: What Newspeople Should Know and the Public Should Expect*, 1st ed. New York, NY, USA: Three Rivers Press, 2007.
- [8]. "The Fight Against Disinformation," Spanish Ministry of Foreign Affairs, [Online]. Available: <https://www.exteriores.gob.es/en/PoliticaExterior/Paginas/LaLuchaContraLaDesinformacion.aspx>. Accessed: Sep. 10, 2023.
- [9]. S. O'Brien, "The Battle against Disinformation," BBC, 2019. [Online]. Available: <https://www.bbc.co.uk/blogs/internet/entries/52eab88f-5888-4c58-a22f-f290b40d2616>. Accessed: Sep. 10, 2023.
- [10]. C. M. Bishop, *Pattern Recognition and Machine Learning*, 1st ed. Berlin, Germany: Springer, 2006.
- [11]. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Berlin, Germany: Springer Science & Business Media, 2013.
- [12]. G. Yildirim, "A novel hybrid multi-thread metaheuristic approach for fake news detection in social media," *Applied Intelligence*, vol. 53, pp. 11182–11202, 2022. [CrossRef] [PubMed].
- [13]. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, 1st ed. Cambridge, MA, USA: MIT Press, 2016.
- [14]. D. Jurafsky, *Speech & Language Processing*, 1st ed. Chennai, India: Pearson Education India, 2000.

- [15]. G. Jawahar, B. Sagot, and D. Seddah, "What Does BERT Learn about the Structure of Language?" in Proc. 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 2019, pp. 861-874. [CrossRef].
- [16]. R. Guarasci, S. Silvestri, G. De Pietro, H. Fujita, and M. Esposito, "BERT syntactic transfer: A computational experiment on Italian, French, and English languages," *Computer Speech & Language*, vol. 71, 2022, Art. no. 101261. [CrossRef].
- [17]. V. Hirlekar and A. Kumar, "Natural Language Processing-Based Online Fake News Detection Challenges—A Detailed Review," in Proc. 5th Int. Conf. Communication and Electronics Systems (ICCES), Coimbatore, India, 2020, pp. 1–7. [CrossRef].
- [18]. A. De, D. Bandyopadhyay, B. Gain, and A. Ekbal, "A Transformer-Based Approach to Multilingual Fake News Detection in Low-Resource Languages," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 21, no. 1, pp. 1–20, 2021. [CrossRef].
- [19]. K. H. Weiss, T. M. Khoshgoftaar, and D. Wang, "A Survey of Transfer Learning," *J. Big Data*, vol. 3, no. 9, pp. 1–17, 2016. [CrossRef].
- [20]. W. L. Hamilton, *Graph Representation Learning*, 1st ed. Berlin, Germany: Springer, 2020.
- [21]. K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake News Detection on Social Media," *SIGKDD Explor.*, vol. 19, no. 1, pp. 22–36, 2017. [CrossRef].
- [22]. C. Cavallaro and E. Ronchieri, "Identifying Anomaly Detection Patterns from Log Files: A Dynamic Approach," in *Lecture Notes in Computer Science*, vol. 117, Berlin, Germany: Springer, 2021, pp. 517–532. [CrossRef].
- [23]. F. K. A. Salem, R. A. Feel, S. Elbassuoni, M. Jaber, and M. Farah, "FA-KES: A Fake News Dataset around the Syrian War," in Proc. Int. AAI Conf. Web and Social Media, Munich, Germany, 2019, vol. 13, pp. 573–582. [CrossRef].
- [24]. H. Ahmed, I. Traoré, and S. Saad, "Detecting Opinion Spams and Fake News Using Text Classification," *Security and Privacy*, vol. 1, no. 2, pp. 1–11, 2017. [CrossRef].
- [25]. A. Koirala, "COVID-19 Fake News Dataset," Mendeley Data, vol. 1, 2021. [CrossRef].
- [26]. R. Oshikawa, J. Qian, and W. Y. Wang, "A Survey on Natural Language Processing for Fake News Detection," in *Language Resources and Evaluation*, vol. 10, Berlin, Germany: Springer Science+Business Media, 2018, pp. 6086–6093. [CrossRef]. [Online]. Available: <http://dblp.uni-trier.de/db/conf/lrec/lrec2020.html#OshikawaQW20>. Accessed: Sep. 10, 2023.
- [27]. A. Agarwal and A. A. Dixit, "Fake News Detection: An Ensemble Learning Approach," in Proc. 4th Int. Conf. Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2020, pp. 1–7. [CrossRef].
- [28]. T. B. H. Saikh, A. Ekbal, and P. Bhattacharyya, "A Deep Transfer Learning Approach for Fake News Detection," in Proc. Int. Joint Conf. Neural Networks (IJCNN), Glasgow, UK, 2020, pp. 1–8. [CrossRef].
- [29]. V. S. Tida, S. H. Hsu, and X. Hei, "A Unified Training Process for Fake News Detection Based on Fine-Tuned BERT Model," arXiv preprint, arXiv:2202.01907, 2022.
- [30]. F. Monti, F. Frasca, D. Eynard, and M. Bronstein, "Fake News Detection on Social Media Using Geometric Deep Learning," *ResearchGate*, 2019. [Online]. Available: https://www.researchgate.net/publication/331195263_Fake_News_Detection_on_Social_Media_using_Geometric_Deep_Learning. Accessed: Sep. 10, 2023.
- [31]. I. Ahmad, P. Gao, S. Yousaf, and M. Ahmad, "Fake News Detection Using Machine Learning Ensemble Methods," *Complexity*, vol. 2020, 2020, Art. no. 8885861. [CrossRef].
- [32]. S. Chandra, P. Mishra, H. Yannakoudakis, and E. Shutova, "Graph-Based Modeling of Online Communities for Fake News Detection," *ResearchGate*, 2020. Available online: https://www.researchgate.net/publication/343689274_Graph-based_Modeling_of_Online_Communities_for_Fake_News_Detection (accessed on 10 September 2023).
- [33]. S. R. P. D. Gangireddy, C. Long, and T. Chakraborty, "Unsupervised Fake News Detection," in Proceedings of the HT '20: Proceedings of the 31st ACM Conference on Hypertext and Social Media, New York, NY, USA, 13–15 July 2020. [CrossRef]
- [34]. A. Thota, P. Tilak, S. Ahluwalia, and N. Lohia, "Fake News Detection: A Deep Learning Approach," *SMU Data Science Review*, vol. 1, p. 10, 2018. Available online: <https://scholar.smu.edu/cgi/viewcontent.cgi?article=1036&context=datasciencereview> (accessed on 10 September 2023).
- [35]. H. Che, B. Pan, M. F. Leung, Y. Cao, and Z. Yan, "Tensor Factorization With Sparse and Graph Regularization for Fake News Detection on Social Networks," *IEEE Transactions on Computational Social Systems*, vol. 10, no. 2, pp. 543–554, 2023. [CrossRef]

- [36]. A. Padha and A. Sahoo, "Quantum Enhanced Machine Learning for Unobtrusive Stress Monitoring," in Proceedings of the 2022 Fourteenth International Conference on Contemporary Computing, Noida, India, 4–6 August 2022. [CrossRef]
- [37]. R. Guarasci, G. De Pietro, and M. Esposito, "Quantum Natural Language Processing: Challenges and Opportunities," Applied Sciences, vol. 12, no. 14, p. 5651, 2022. [CrossRef]