applied sciences

MDPI

Article

Exploring the Role of Emotions in Arabic Rumor Detection in Social Media

Hissa F. Al-Saif * and Hmood Z. Al-Dossari

Computer Science and Information Systems, University of King Saud, Riyadh 11421, Saudi Arabia; hzaldossari@ksu.edu.sa
* Correspondence: halsaif@su.edu.sa

Abstract: With the increasing reliance on social media as a primary source of news, the proliferation of rumors has become a pressing global concern that negatively impacts various domains, including politics, economics, and societal well-being. While significant efforts have been made to identify and debunk rumors in social media, progress in detecting and addressing such issues in the Arabic language has been limited compared to other languages, particularly English. This study introduces a context-aware approach to rumor detection in Arabic social media, leveraging recent advancements in Natural Language Processing (NLP). Our proposed method evaluates Arabic news posts by analyzing the emotions evoked by news content and recipients towards the news. Moreover, this research explores the impact of incorporating user and content features into emotion-based rumor detection models. To facilitate this investigation, we present a novel Arabic rumor dataset, comprising both news posts and associated comments, which represents a first-of-its-kind resource in the Arabic language. The findings from this study offer promising insights into the role of emotions in rumor detection and may serve as a catalyst for further research in this area, ultimately contributing to improved detection and the mitigation of misinformation in the digital landscape.

Keywords: social media; rumor detection; sentiments; emotions analysis; user-credibility; Arabic NLP; pre-trained language model

Citation: Al-Saif, H.F.; Al-Dossari, H.Z. Exploring the Role of Emotions in Arabic Rumor Detection in Social Media. *Appl. Sci.* **2023**, *13*, 8815. https://doi.org/10.3390/app13158815

Academic Editor: Chilukuri K. Mohan

Received: 17 May 2023
Revised: 22 July 2023
Accepted: 26 July 2023
Published: 30 July 2023

1. Introduction

Social media platforms generate a vast quantity of data and have become a popular medium for information transmission and real-time news. According to a recent Pew Research study [1], about half of U.S. adults currently obtain news from social media. Twitter is among the most popular social media platforms in Arabic-speaking nations. Launched in 2006, Twitter has grown significantly in both its user base and content in recent years, with currently 330 million active users, making it one of the most important platforms for news posting, sharing, and dissemination [2]. Notably, social media platforms allow users to easily and freely share whatever they choose in a publicly visible way, opening up unparalleled communication possibilities [3]. However, they lack adequate control and authority over their content, which can lead to the dissemination of misleading information either unintentionally or intentionally, the latter with the aim of deceiving users [4]. This situation has potential negative impacts on politics, the economy, and health services. Additionally, the dissemination of rumors can lead to panic and anxiety in society. For example, the COVID-19 outbreak resulted in a surge of false information on social media, influencing people's opinions and decision-making and arguably hindering efforts to manage the disease. Recent studies indicate that fake news is a global issue due to the well-documented effects of rumor dissemination [5,6]. Moreover, nations are experiencing a climate inundated with deceptive information in social media, which is indicative of a situation that may have adverse effects on citizens' political involvement [7]; for example, inaccurate information played a significant role in the outcome of the 2016 U.S. presidential

election [8,9]. The manual approach to detecting trending false rumors on social media through evaluating the substance of news reports and verifying primary sources is a time-consuming and laborious process [5,6], leading to low-quality labeling, particularly given the high volume of rumors circulating every day. Additionally, manual systems are often unable to detect rumors in a timely enough manner to mitigate their impact [10]. Consequently, there is a need to investigate automated solutions to this problem.

Social research studies have suggested that news items that evoke strong emotions, such as anger and anxiety, are more likely to spread virally on social media platforms compared to less inflammatory items [11]. This phenomenon motivates the spreaders of false rumors to intensify the emotional content of their posts, which attracts greater public attention. This research focuses on examining the impact of emotions on rumor detection by extracting emotions from two sources: source emotions (i.e., the source of news content) and audience emotions (i.e., those who receive the information disseminated on social media) using popular lexicons and pre-trained models that support the Arabic language. Additionally, the study evaluates the effectiveness of these features in different contexts, such as different cultures or languages. Figure 1 presents examples of rumor and non-rumor posts from Twitter, along with related comments.

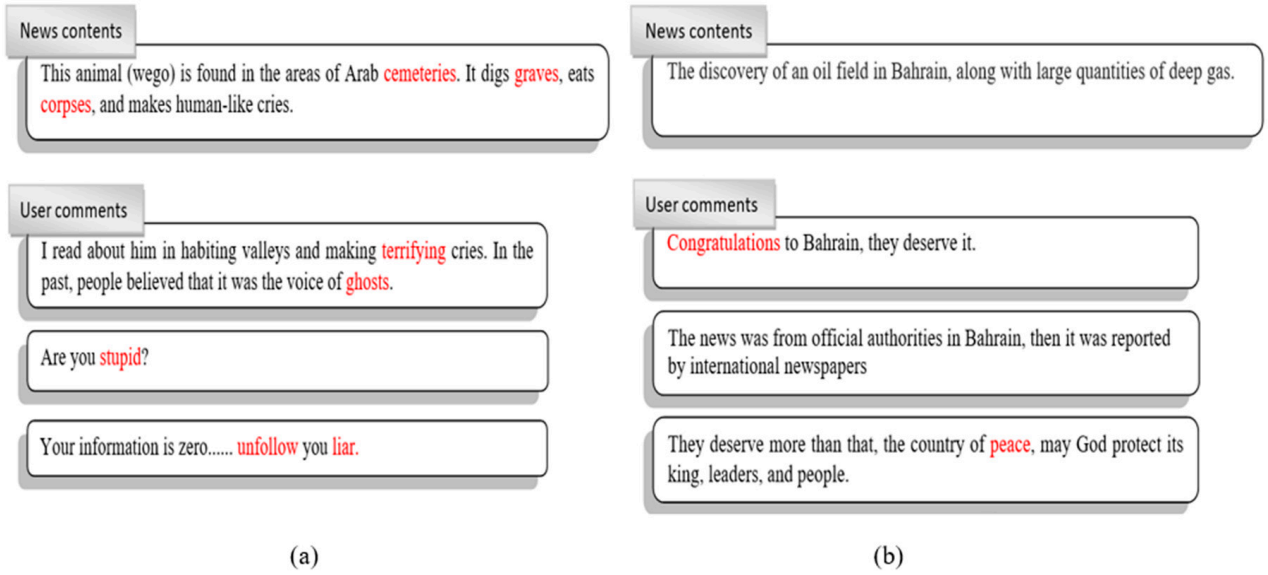

Figure 1. Rumor and no rumor posts from Twitter translated from Arabic into English. (**a**) A rumor post that contains the emotion of fear and evokes emotions such as doubt, anger, and fear in related comments, as illustrated in red words. (**b**) A non-rumor post that contains no emotion and evokes positive emotions.

This research makes a significant contribution to the field of NLP, specifically in the area of rumor detection. Significant progress has been made in the research on rumor detection in several languages, while Arabic studies are scarce and lag behind those in other languages due to the rich vocabulary and numerous dialects of the Arabic language. Moreover, most Arabic research in this area has produced average or poor results due to the difficulty in pre-processing and feature extraction in Arabic compared to other languages, such as English or French. To tackle this problem, we propose a new model and have conducted experiments to assess its performance. The following are the primary contributions of this paper:

First, this work has constructed a dataset for rumor detection from social media in the Arabic language, which will considerably enrich Arabic resources. It is a first-of-its-kind Arabic rumor dataset that consists of related reactions from the public.

Second, emotional features and sentiments are extracted by various lexicons and publicly available pre-trained models trained in the Arabic language and then compared in terms of their ability to detect rumors.

Third, a deep learning framework is proposed in order to detect rumors that utilize emotions. Furthermore, this study investigates the effect of combining user and content features into emotion-based rumor detection algorithms.

Fourth, extensive analyses have been undertaken on the role of emotions in rumor detection from three angles: top contributing emotions, emotion distribution in news and comments, and emotion distribution across different topic areas.

To the best of our knowledge, this work presents the first investigation into the role of emotions in rumor detection in the Arabic language. The results are likely to enrich Arabic NLP and may encourage other researchers to conduct more research in this area. The research seeks to answer the questions listed below:

- RQ1: Are emotional cues and sentiments in news and replies capable of detecting rumors independently, without additional information?
- RQ2: Can emotional cues and sentiments in news and replies improve rumor detection when used as supplementary features for textual content?
- RQ3: How do emotion, sentiments, user, and content features contribute to distinguishing between rumors and true news in Arabic social media?

The rest of this paper is structured as follows. Section 2 presents background on rumors and the key topics related to our research, followed by a review of the relevant literature on rumor detection in Section 3. Section 4 describes the dataset and presents the proposed approach for addressing this issue. Experimental results are presented in Section 5. Finally, Section 6 concludes the study and outlines possible future directions for research in detecting rumors.

2. Background

This study is primarily concerned with exploiting emotional features to alleviate false rumors on social media. To provide a concise and informative background on this research area, this section reviews the concept of false information and its relation to rumors. Following this, we review the pre-trained language models that are linked to advanced techniques in NLP, which are the essential background for our research. Lastly, the task of emotion classification is described, along with an explanation of how this can be used for rumor detection.

2.1. False News and Rumors

The issue of false information has existed since the invention of the first writing instrument [6]. Recent research [3] has recommended using the term "false information" rather than "fake news" because the latter term is inherently political, thus limiting its scope. False information is classified into two types: misinformation and disinformation, depending on whether it causes damage or espouses a particular interest. Misinformation is defined as false information spread without the intention of causing harm to a particular organization or the public generally. It can be caused by a variety of factors, including erroneous labeling (e.g., of people) and lax fact-checking processes, and it can spread quickly to users who are careless about the truth of what they receive or disseminate. Disinformation, on the other hand, refers to erroneous information disseminated with the purpose of confusing, hurting, or deceiving a specific group or the public in general [12,13].

Eight major categories of false information are identified by [14,15], which include fabricated information, biased news, propaganda, conspiracy theories, hoaxes, rumors, clickbait, and satire. Rumors constitute the most popular type of false information and are extensively shared on social media. There are many ways to define rumors. For example, earlier studies have labeled them as false information [16], while recent research [17] has defined them as fake news or unconfirmed news items that have spread widely but require proof to determine their validity. Rumors can be categorized according to their

veracity state (e.g., true, untrue, or unresolved) [18] or their trustworthiness ranking (high or low) [19].

2.2. Pre-Trained Language Models

The field of NLP reached a milestone in language representation models with the emergence of pre-trained language models (PLMs). These language models are trained based on neural networks holding vast data in textual form and fine-tuned on downstream NLP tasks. The advancement in transformer architecture has been a key enabler for PLMs, including BERT [20], OpenAP GPT [21], RoBERTa [22], and XLNet [23]. Transformers are composed of multiple encoder and decoder components, as well as multi-head attention mechanisms, which contribute to their effectiveness in various NLP tasks. The attention mechanism allows for selectively focusing on pertinent information while disregarding irrelevant information. The fundamental principle underlying PLMs is using pre-trained models and then applying fine-tuning for tasks rather than training models from scratch. These models have been proposed to cope with the issue of polysemy in static word embeddings (i.e., word2vec and GloVe) by considering the context in which the word occurs (referred to as contextualized word embeddings).

BERT was released in 2018 by Google researchers Jacob Devlin and colleagues. It is now widely accepted as the preliminary step for research on NLP. The model was trained with two tasks. First, masked language modeling allows the representation to integrate the right and left contexts, thus enabling it to learn the context in contrast to previous language representation models that only captured context in one direction. Second, it was trained on next-sentence prediction, which simultaneously pre-trains text-pair representation. BERT performs remarkably well across a range of natural language processing tasks, including text categorization, thanks to this mix of features. It includes an encoder with 12 transformer blocks and a hidden size of 768 with 12 self-attention heads. BERT provides two predominant methods: feature-based and fine-tuning [20]. The feature-based approach extracts pre-training embedding vectors at various levels (word, phrase, or paragraph), which are subsequently supplied as extra features to a given model [24]. The second method, fine-tuning, entails training pre-trained parameters on the desired task. The fine-tuning method is simple in BERT since the self-attention feature enables it to represent several classification problems by switching appropriate inputs and outputs [20]. The most recent research demonstrates that fine-tuning outperforms the feature-based method on text classification tasks [25]. Models are initialized for diverse downstream tasks using the same pre-trained model parameters, and all parameters are adjusted during fine-tuning. Every single instance has the special symbol [CLS] before it and [SEP] as a separate separator token.

Fortunately, a multilingual version of BERT was released for the top 104 languages, including Arabic [20]. In addition, the more recently developed XLM-RoBERTa model provides a framework containing languages, including Arabic, trained on a large corpus of text at 2.5 TB [26]. However, according to [27], the multilingual versions are not as effective compared to the English version due to the lack of language resources. Several pre-training monolingual models in Arabic have been introduced to address this issue of limited performance, including AraBERT and MARBERT. The outstanding performance of monolingual models may be determined by the amount of pre-training corpora that were utilized to train these models. AraBERT, for instance, was pre-trained with a corpus that comprises 2.7 billion tokens, in contrast to Multilingual BERT and XLM-RoBERTa, which had pre-training corpora with 110,000 tokens and 250,000 tokens, respectively.

AraBERT and AraBERT-Twitter: The AraBERT model was trained specifically for the Arabic language on a sizable corpus in Modern Standard Arabic (MSA) news, which consisted of 70 million sentences and around 3 billion words. The aim was to achieve, in the Arabic language, what the BERT model had accomplished in the English language. Three NLP downstream tasks—sentiment analysis, question resolution, and Named Entity Recognition (NER)—were used to assess the model. The trials on these Arabic NLP tasks

demonstrated that the AraBERT model outperformed other baselines, including earlier single-language and multilingual techniques, on the majority of assessed tasks. AraBERT's new version has approximately 60 times the text of the multilingual BERT and includes 110 million trainable parameters. AraBERT-Twitter is an improved version of the AraBERT model for the Arabic language. It was trained on 60 million Arabic tweets in addition to the data used to train the original AraBERT model. This enhancement allows AraBERT-Twitter to better handle the nuances and complexities of Arabic text found on social media platforms like Twitter [28].

MARBERT and ARBERT: MARBERT and ARBERT are two Arabic versions of BERT that were created in 2021 by [29]. ARBERT was trained on 61 GB of MSA text from a variety of Arabic datasets, while MARBERT, unlike AraBERT, was pre-trained on enormous datasets (6 billion Arabic tweets) with different Arabic dialects. Due to the character limit on tweets, MARBERT uses the same network architecture as the BERT model but does not include the next-sentence prediction objective. Six NLP tasks, including sentiment analysis, topic categorization, dialect identification, question-answering, NER, and social meaning, were used to evaluate MARBERT. According to [29], these experiments revealed that MARBERT is noticeably superior to AraBERT.

2.3. Emotion Classification

Emotions are an important part of language and are regarded as complicated. Emotion classification is a multi-label classification problem that detects the emotion for textual input from pre-defined emotion labels. Several models have been developed to characterize emotions, including the six-category model of [30] of basic emotions: joy, sadness, anger, fear, disgust, and surprise. The model proposed by [31] extended these six categories to include trust and anticipation. Arabic NLP is not as mature as English NLP, and thus fewer research articles have been published on emotion detection in the Arabic language, which is mainly due to resource limitations. Prior studies on emotion recognition for Arabic text varied from traditional Machine Learning (ML) to Deep Learning (DL) advancement. The three major approaches are the lexicon-based approach, the ML approach, and the hybrid approach. The lexicon-based solutions are based on computing the overall intensity for each emotion using a lexicon [32]. On the other hand, various algorithms have been used for emotion recognition, including Support Vector Machine (SVM) [33], Complement Naïve Bayes (CNB) [34], Recurrent Neural Networks (RNN) [35], and Convolutional Neural Networks (CNN). The authors in [35] proposed three models for emotion detection in Arabic text: a feature-engineered model, a DL model, and a hybrid model consisting of elements of both. They used various text features in their feature-engineered model, including stylistic, lexical, syntactic, and semantic features. Various deep neural networks were used with word embeddings. The approach was tested on a range of datasets and outperformed state-of-the-art models in a variety of measurements. Recently, a shift toward the advanced pre-trained language model in the study of [36] involved proposing that multilingual BERT build a toolkit for social media processing in Arabic. It focused on several tasks, including delicate, gender, and emotion prediction. Emotions have recently been used as indicators for false information detection. In [10], the authors investigated false and true Twitter rumors and discovered that false rumors tend to elicit emotions such as anxiety, disgust, and surprise, while true rumors evoke emotions like pleasure and anticipation. This finding has inspired researchers to explore the role of emotions in the task of rumor detection.

3. Related Works

In recent years, rumor detection has garnered significant attention from researchers, leading to the exploration of various methods and approaches to address this challenge. Two predominant approaches now in use are content-based and context-based models. The textual information included in news articles, as well as image and video information, are used in content-based techniques. However, these approaches have mainly concentrated

on hand-engineered features or shallow representations, which are unable to adequately handle the tremendous complexity of the rumor detection task [37]. On the other hand, context-based models take into account details such as how news is distributed across social media platforms, author profiles, and user interactions with news [38–40]. However, since rumors take time to propagate among users and are not easily detectable in the early stages, these approaches struggle to identify them rapidly. This section reviews the work conducted on rumor detection in two areas: studies in Arabic and studies that have explored emotions in other languages.

3.1. Emotions-Based Rumor Detection Approaches

A selection of studies that address rumor detection from an emotional perspective is reviewed in this section. The study [41] involved the first exploration into the role of emotions in false news detection, taking into account a set of false information types (propaganda, hoax, clickbait, and satire) derived from social media and online news article resources. The research revealed that emotions have a significant influence on false information detection. In false news detection, two sources of emotion must be considered: source emotion (i.e., the source news content) and audience emotion (i.e., those who receive the information disseminated on social media) [42]. The authors in [42] implemented a novel framework that exploited both publisher and social emotions to detect false news. The framework consisted of three parts. First, the content component extracted semantic content combined with additional emotional features. Second, the public responses component extracted the main emotional features from people's reactions. Third, the prediction component used the fused information mentioned above for model production. Another study explored whether a relationship exists between these dual emotions [43]. Statistical analysis was performed on these emotions to test their relationship and to differentiate between fake and real news. Based on the results, there are two forms of appearance: emotion resonances, where the dual emotions are similar, and emotion dissonances, where the dual emotions are not similar. The authors reported that the results showed considerable improvement when considering the gap between the two emotions. The authors in [44] used emotion lexicons to detect fake news in healthcare. The experimental findings demonstrated the effectiveness of the representations in both supervised and unsupervised settings. In [45], the authors explored various approaches to extract rich emotion representations and to determine their intensity, drawing on lexicon-based and neural network approaches for false news detection. The experimental findings on real-world datasets attested to the importance of considering emotions in a credibility detection framework. The study [46] presents a promising approach for detecting fake news that comprises two main components: the multi-modal fusion module and the adaptive interaction module. The multi-modal fusion module combines textual features with social interaction features, such as user engagement and sentiment analysis. This module aims to capture both the content and social context of news articles. The adaptive interaction module uses attention mechanisms to learn the importance of different social interactions and dynamically adjust the attention weights, allowing the model to focus on the most relevant pieces of information. The proposed method improves the accuracy of fake news detection and aids in the reduction of misinformation spread on social media. The study [47] proposed a fake news detection model for social media that leverages sentiment analysis of news content and emotion analysis of users' comments. The model uses ML techniques to classify news articles as either fake or genuine based on the sentiment expressed in the article and the emotions expressed in the comments. The proposed model was evaluated using a dataset of news articles and comments from Twitter and achieved high accuracy in detecting fake news. The study concludes that combining sentiment analysis and emotion analysis can improve the accuracy of fake news detection on social media. The study [48] proposes a new approach for detecting fake news called "FakeFlow", which focuses on the flow of affective information through social media. This approach involves modeling the propagation of emotions within a network and using features such as sentiment,

emotion, and interpersonal relationships to identify fake news. The authors evaluated the effectiveness of their approach on two datasets and compared it with existing methods, showing that FakeFlow outperforms them by a significant margin. They also conducted a sensitivity analysis to demonstrate the robustness of their approach and identify potential areas for future research. Overall, the paper presents a promising approach to tackling the problem of fake news detection by leveraging the social and emotional dynamics of online communities. In Table 1, we present a comprehensive comparison of the papers reviewed in the field of emotion-based rumor detection, highlighting their domain, dataset source, and language.

Table 1. Comparison of reviewed papers in emotion-based rumor detection.

Paper	Dataset Source	Domain	Languages
Juan Cao, Chuan Guo, Xueyao Zhang. Sheng, Kai Shu, and Miao Yu (2019) [42]	Weibo	Various domains	Chinese
Juan Cao, Xueyao Zhang, Xirong Li, Qiang Sheng, Kai Shu, and Lei Zhong (2021) [43]	Twitter, Weibo	Various domains	English, Chinese
Anoop k, Deepak P and Lajish V (2020) [44]	News	Health	English
Lianwei Wu and Yuan Rao (2020) [46]	Twitter	Various domains	English
Bilal Ghanem, Paolo Rosso, Francisco Rangel (2020) [41]	Twitter, News	Various domains	English
Anastasia Giachanou, Paolo Rosso, Fabio Crestani (2019) [45]	News	Political	English
Bilal Ghanem, Simone Paolo Ponzetto, Paolo Rosso, Francisco Rangel (202l) [48]	News	Various domains	English
Suhaib Kh Hamed, Mohd Juzaiddin Ab Aziz, Mohd Ridzwan Yaakub (2023) [47]	Reddit	Various domains	English

3.2. Arabic Language Rumor Detection Approaches

The Arabic language is one of the most widely spoken languages around the world. Indeed, it is one of the top six most commonly used languages, according to the United Nations [49]. Its speakers are spread across Arab countries located in Africa and Asia and in some countries that border the Arab world. However, in comparison to other languages, rumor detection in Arabic is still limited, with relatively few studies addressing the issue. This section highlights existing Arabic language rumor detection approaches, which can be classified into ML and DL approaches. A feature-based approach utilizing classic ML methods has been reported in the literature [50–53]. The authors in [52] used content-related and user-related features to detect Arabic rumors. They tested their approach with the Random Forest (RF), Decision Tree (DT), AdaBoost, and Linear Regression (LR) algorithms, and the findings showed that it had an accuracy of 76%. Another study [51] used hybrid non-textual features to evaluate the trustworthiness of Arabic news on Twitter. The authors used DT, SVM, and Naïve Bayes (NB) classifiers on a dataset of 800 Arabic news tweets that had been manually annotated. The findings showed that DT outperformed SVM and NB by about 2% and 7% in accuracy, respectively. Another study [50] proposed an approach for identifying rumors in Arabic tweets utilizing non-textual information, such as user profiles, using a semi-supervised expectation maximization algorithm. The findings demonstrated that the proposed model beat Gaussian Naive Bayes (GNB) with an f1-score of 78.6%. In a study [53], user- and content-based features were extracted. The eXtreme Gradient Boosting (XGBoost) algorithm was employed, achieving an accuracy rate of 97%. A further study conducted by Al-Khalifa et al. [54] proposed an approach for assessing the reliability of news items posted on Twitter by assigning each tweet one of three levels of credibility—low, medium, or high—using two methods. The first method makes use of legitimate news sources to identify a link between a Twitter tweet and these sources. The second method employs the outcomes of the first technique, as well as a range of other features. The results showed that the first approach outperforms the second approach.

A few studies have examined this issue from a textual content perspective [55–59]. Traditional ML classifiers such as SVM, LR, and NB have been investigated in [56,57,59]. DL approaches were investigated in [55,58,59] along with several transformer-based approaches to analyzing textual content (i.e., news) [55,60–62]. However, these approaches face the limitation of relying on domain-dependent features, which restricts their ability to detect new rumors effectively. The study [62] employed a CNN model on a balanced Arabic corpus uniting stance identification with fact-checking. The model performed better than state-of-the-art methods, with the highest accuracy of 91%. Research [59] was carried out to identify COVID-19 rumors using textual features in both traditional ML and DL approaches. The performance of algorithms using different feature representations, optimizers, and ensemble learning was evaluated in this research. The results showed that ensemble learning improved machine learning performance. Contextualized embedding models were used in [60] to detect false news. Although the bulk of these algorithms has never been applied to detect Arabic false news, experimental findings show that these cutting-edge models are resilient, with accuracy surpassing 98%. Research in [63] proposed an Arabic rumor detection approach that employs both textual and visual features. Several experiments were carried out in order to select the best pre-trained model to extract features for developing a multi-modal model. Ultimately, the MARBERTv2 model was used to extract textual features, while a combination of VGG-19 and ResNet50 was used to extract visual features. Lastly, the experimental findings showed that textual features alone could outperform multi-modal models in rumor detection tasks.

4. Research Methodology

The proposed methodology investigates cutting-edge deep learning approaches, such as PLMs, that can capture both semantic and contextual aspects of textual data. Figure 2 illustrates the components of our emotion-based rumor detection model. Data collection serves as the initial stage in building a model for rumor detection. Our model comprises three branches: the first branch extracts topic features from the textual content of news articles; the second branch extracts reaction features from the textual content of comments on news articles; and the third branch extracts emotions from both news content and comments. Following this, emotions are concatenated with the output from the language model and fed into a dense layer to extract emotion-related representation. A summary of the methodology is outlined below.

4.1. Dataset Collection

There was no ready-made dataset in the Arabic language suitable for our experiments. While available datasets include real rumor incidents, they suffer from several limitations, such as the absence of reactions (i.e., replies) towards rumors. Additionally, these datasets contain duplicates and noise. To address these limitations, we created a new dataset following the methodology employed in the construction of the English rumor detection dataset [64]. Our dataset collection process spanned approximately two months and involved searching for rumors and non-rumors that occurred within the last ten years. We gathered false claims from Norumors (http://norumors.net/, accessed on 12 January 2023), a non-profit organization founded in 2012 to combat rumors manually in Saudi society, and true claims from news websites. The claims ranged from 2012 to 2022 and encompassed a broad array of topics, including politics, society, and health. To investigate the influence of crowd response, we collected related tweets that post claims on Twitter and then gathered the related responses posted for these claims using the Twitter API. After filtering and removing unrelated responses, such as advertisements or blank comments, we labeled false news as rumor and true news as non-rumor. The dataset we created consists of 403 events, including 202 false rumors and 201 true rumors. The total number of news posts and replies within these events, along with their veracity, are displayed in Table 2.

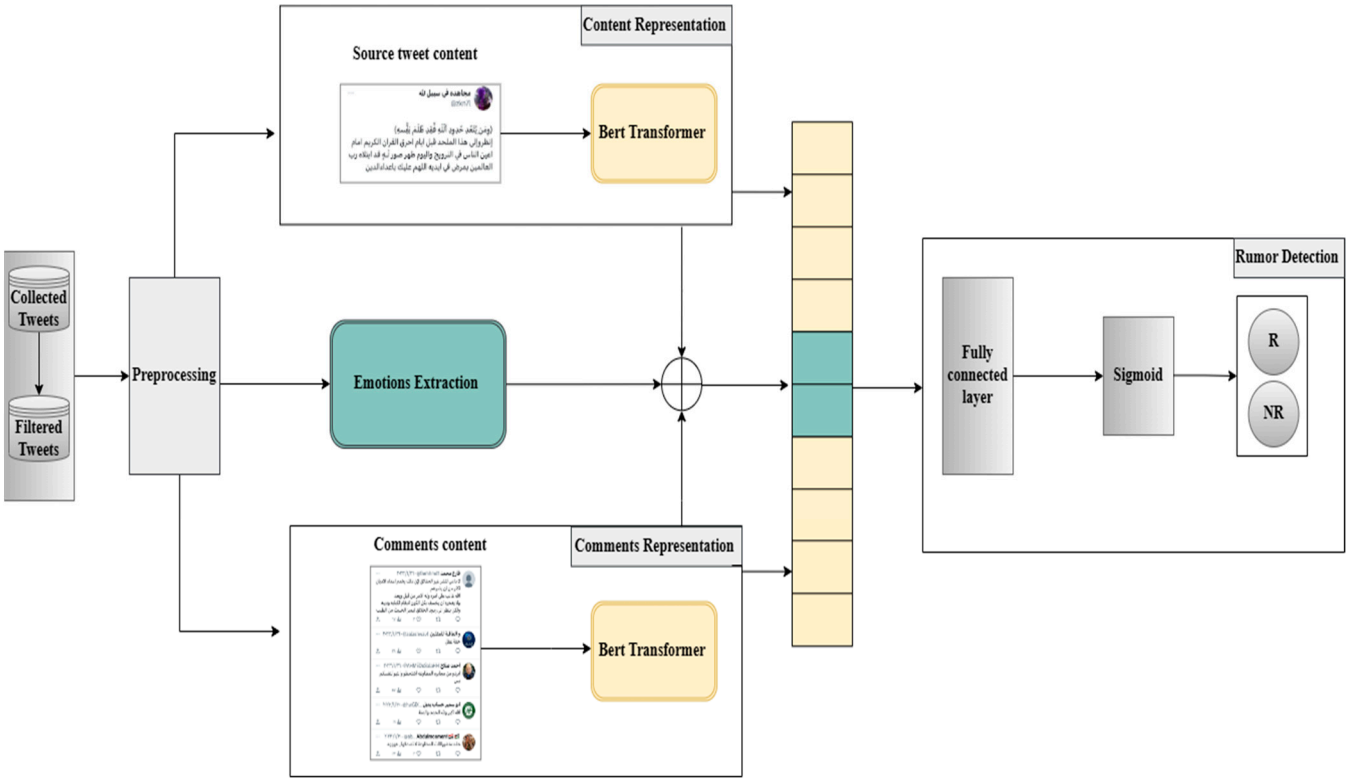

Figure 2. The proposed emotion-based rumor detection model.

Table 2. Dataset Statistics.

Statistics	Count
Events#	403
Rumor#	202
Non-rumor#	201
News Posts#	20,493
Replies#	40,759

4.2. Feature Extraction

This section provides a detailed explanation of the user, content, and emotion features that were used in our framework to detect non-credible tweets.

Content and User-Based Features: Content and user-based features are crucial for detecting a non-credible tweet. A total of 47 features were extracted from each tweet using the Twitter API and divided into two categories: content-based and user-based features. Table 3 lists the user features in our dataset [39,65,66]. Twitter API documents contain a detailed explanation of each feature. The content-based features include the number of retweets, replies, favorites, presence of hashtags and URLs, the existence of photos, tweet length, and use of punctuation marks. These features are important indicators in social media of rumor verification, as non-credible tweets tend to have lower engagement rates and lack supporting evidence. For example, the presence of hashtags (e.g., #fakenews) was based on the observation of real posts that tended to contain more hashtags [50]. Furthermore, the presence of URLs in a post has been investigated in prior studies [67,68]. According to their findings, supporting posts are more likely to contain links, indicating that they refer to information that confirms their claim. Moreover, the existence of photos is crucial for determining news veracity since real news tends to have photos or videos that support its content. Furthermore, replies can deny false news by posting images from authorities as evidence [69]. The word count was also used since false rumors are typically more detailed and lengthier, whereas true news is often concise [70,71]. On the other hand, user-based

features include features such as follower count, number of status updates, number of friends, number of likes, verification status, length of time a user has held an account, and whether the user has a URL or textual description on their profile. Some features were directly obtained from the Twitter API, while others were derived indirectly. For example, the time span was computed by subtracting the time when a tweet was posted from the time when the related user was registered. User profile features have been used in previous studies to determine rumor veracity, with the findings indicating that false rumor conversations were often initiated by users with fewer friends and lower status counts, and that non-verified accounts were more likely to share false news [72–74].

Table 3. User and content feature explanation.

Type	Feature	Type	Description
User-based features	List count	Float	Number of lists that author participates in
	Description status	Boolean	Whether user provides a personal description
	Char-description length	Integer	Personal description length (in words)
	Word-description length	Integer	Personal description length (in char)
	User-favorite count	Integer	Number of posts that author favors
	Username length	Integer	Number of characters in username
	Screenname length	Integer	Number of characters in screenname
	User-followers count	Integer	Number of accounts that follow author
	User-friends count	Integer	Number of accounts that author follows
	Geo-enabled	Boolean	Whether account has enabled geographic location
	Media count	Integer	Number of media posted by author
	Custom timelines	Boolean	Whether user has a custom timeline
	Status count	Integer	Number of posts written by author account
	Verified	Boolean	Whether user has verified accounts
	URL statues	Integer	Whether author account has a URL in homepage
	Protected	Boolean	Whether user has protected their tweets
	Consent status	Boolean	Whether account requires consent
	Can dm	Boolean	Whether account allows users to send direct messages privately
	Profile background tile	Boolean	Whether author account has background profiles tile
	Profile background image	Boolean	Whether author account has background profiles images
	Default profile	Boolean	Whether author account has not changed theme or background of profiles
	Default profile images	Boolean	Whether author account has changed the default profiles images
	User engagement	Float	User engagement (# posts/(account age + 1))
	Following rate	Float	Following rate (i.e., followings/(account age + 1))
	Favorite rate	Float	Favorite rate (i.e., user favorites/(account age + 1))
	User effects	Float	Whether the author is a producer or recipient determined by the formula # followers/# following
	Reputation score	Float	Reputation score of accounts, calculated by the formula # followers/(# followers + # following + 1)
	Account age	Integer	Number of years since account creation
	Following	Boolean	Whether author account is followed by authenticated user
	Follow request sent	Boolean	Whether author account is requested to follow by authenticated user
	Notifications	Boolean	Whether author account has turned on notifications
	Contributor-enabled	Boolean	Whether account has enabled contributors
	Translation-enabled	Boolean	Whether account has enabled translations

Table 3. *Cont.*

Type	Feature	Type	Description
	Is translator	Boolean	Whether account has a translator
	Timespan	Float	Difference in years between account creation and tweet posted
	Promotable	Boolean	Whether account is promotable
	Normal followers	Boolean	Number of normal followers
Content-based features	Retweet count	Integer	Total number of post retweets
	Reply count	Integer	Total number of post replies
	Favorite count	Integer	Total number of post favorites
	URL presence	Boolean	Whether tweets have a URL
	Question mark presence	Boolean	Whether tweets have a question mark
	Exclamation mark presence	Boolean	Whether tweets have an exclamation mark
	Media presence	Boolean	Whether tweets have media videos or images
	Hashtag presence	Boolean	Whether tweets have hashtags
	Word length	Integer	Tweet length in char

Sentiments and Emotions Features: Sentiments and emotional cues can be crucial in distinguishing between rumor and non-rumor. Estimating the emotional signals was the first step in our process. To circumvent the time-consuming annotation process for each response, we considered various public approaches. Two methods are employed for calculating emotional signals: the lexicon-based method and pre-trained models. The Arabic language is considered a low-resource language, and the available emotion-based lexicons suffer from a limited vocabulary size. To address this challenge, we used the Google API translator to translate text from Arabic into English, allowing us to utilize English lexicons to extract sentiments and emotions. There are two popular lexicons for emotion recognition:

1. SenticNet (https://sentic.net/, accessed on 12 January 2023) [75]: SenticNet is a concept-level lexicon that utilizes denotative and connotative information-associated concepts from the WordNet (https://wordnet.princeton.edu/, accessed on 12 January 2023) lexical database to perform emotion recognition. It is employed to extract mood tags at the word level. It covers eight emotions: anger, calmness, eagerness, disgust, fear, joy, pleasantness, and sadness. Additionally, we can extract negative and positive sentiment tags for each word in a sentence.
2. NRC Emotion Lexicon (http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm, accessed on 12 January 2023): The NRC emotion lexicon is used to assign availability statuses to the eight emotions based on the Plutchik model, namely, anger, trust, disgust, fear, joy, sadness, surprise, and anticipation, as well as sentiment. Two Python packages are utilized to extract emotions: the LeXmo and NRCLex packages.

The method we used is straightforward. First, we concatenated all the replies or news posts about a specific event; then, if certain words existed in a sentence, the sentence was considered to reflect a particular sentiment or emotion. For example, a statement containing the word "happy" expresses happiness and joy, while a sentence containing the word "scare" expresses the emotion of fear. More formally, given textual content with the length L, $W = [w_1, w_2, \ldots, w_i, \ldots, w_L]$, where w_i is the ith word in the text. We aimed to extract emotions related to emotional words $E = [e_1, e_2, \ldots, e_i]$ that convey a certain emotion e. Then, we calculated the emotion frequency and normalized it by sentence length to obtain a feature value representing the intensity of that emotion in the given text. This process was applied to all emotions to create a feature vector for the entire text.

We also used two publicly available pre-trained models that were specifically trained for the Arabic language to derive emotion categories and sentiments:

1. AraNet tools [36]: AraNet tools support a wide variety of Arabic NLP tasks, including dialect, emotion, and irony prediction. Built on BERT architecture, AraNet provides state-of-the-art performance for these tasks. It covers eight emotions: sadness, anticipation, surprise, anger, fear, happiness, disgust, and trust.
2. CAMeL [76]: CAMeL tools are used for extracting sentiments, such as positive, negative, and neutral.

By using these pre-trained models and tools, we can obtain emotion categories and sentiment scores for Arabic text without the need to translate it into English or rely on limited lexicons. These models can capture more complex emotional patterns in the text and offer improved accuracy and generalization capabilities compared to lexicon-based approaches. Using these approaches, we extracted emotions from each post in both news and comments. Then, we aggregated the emotion features to obtain a mean representation, reflecting the average emotions across events. Subsequently, the extracted features were concatenated for news content and replies separately. The final feature representation for news content and replies were obtained as follows:

$$\text{Basic Features} = \text{SenticNet-emotions} \oplus \text{NRC-emotions} \oplus \text{AraNet-emotions} \oplus \text{LeXmo-emotions} \oplus \text{CAMeL-sentiments}$$

The study [42] stated that using gaps in emotions to model the resonances and dissonances of emotions showed a considerable improvement. We followed their work, using a subtracting operator as in the following equation:

$$\text{Features-Gap} = \text{News features} - \text{Replies features}$$

The final representation of all emotions and sentiments contains both original features with gaps as in the following equation:

$$\text{All Features} = \text{News-features} \oplus \text{Replies-features} \oplus \text{Features-Gap}$$

4.3. Pre-Processing Unit

The majority of the words in our dataset are in colloquial Arabic, which negatively affects spelling and grammar. Also, there are a lot of lost spaces between words due to either written mistakes to the existence of words in the hashtag format, in which Twitter policy requires a tick "_" between words. To prepare the dataset, we performed the following pre-processing steps using the NLTK library:

- Diacritical Mark Elimination: We systematically removed diacritical marks from the Arabic text, resulting in a more consistent and standardized representation of the language.
- Exclusion of Non-Arabic Text: Our pre-processing strategy involved the removal of all non-Arabic content, such as hyperlinks, symbols, mentions, usernames, English characters, and numerals. Simultaneously, hashtags and keywords associated with news agencies or anti-rumor organizations were eliminated to prevent the model from favoring accurate tweet recognition.
- Character Normalization: To ensure uniformity in the Arabic text, we normalized specific characters by converting أ, آ, and إ into ا.
- Stop Word Elimination: This phase involved filtering and removing common articles, pronouns, and prepositions from the Arabic text, such as ("في", "in") or ("على", "on") which typically offer minimal analytical value. However, some essential stop words for our task, such as لا and غير, were retained as they can enhance performance.

4.4. Textual Representation of PLM Unit

This unit is responsible for extracting text-based features from posts using the BERT model. It processes pre-processed news posts and comments obtained from the previously described unit and feeds them into BERT models for representation extraction. We opted for PLMs due to their exceptional performance and suitability for small data sizes. As these models are trained on vast amounts of data, they help mitigate the overfitting problem. Furthermore, these models generally perform well in inferring context and implicitly recognizing emotions. This presents an additional challenge when trying to leverage emotional features for rumor identification. As a result, we decided to utilize these models to demonstrate the effectiveness of incorporating emotion features to enhance rumor detection capabilities. We explored two recently developed Arabic PLMs, AraBERT-Twitter, and MARBERT, for extracting textual representations in the context of rumor detection. These models were selected based on their impressive performance across various classification tasks and their training on Twitter data, which aligns with our dataset comprising Twitter comments in colloquial languages.

Contextual embeddings of textual inputs were acquired using two approaches: fine-tuning the BERT model and employing a feature-based method without fine-tuning any BERT parameters. Comparing the fine-tuning and feature-based approaches for each model can shed light on the best strategy for extracting and utilizing textual and emotional features in the task of rumor detection. Furthermore, we argue that better representation of sentences can affect utilizing emotions. This motivated us to investigate various textual representations through three methods: first, by extracting the hidden state of the last layer; second, by concatenating the hidden states of the last four layers; and lastly, by applying mean pooling on the last four layers to compute the average of all tokens. Padding tokens were immediately removed to avoid generating distinct sentence embeddings based on the number of padding tokens. To extract representations, we tokenized each instance using the HuggingFace library, incorporating [SEP] and [CLS] tags and subsequently encoded them to produce token IDs. For sentences exceeding 512 tokens in length, we truncated them to accommodate language models limited to a maximum of 512 tokens. Conversely, shorter sentences were padded to achieve a uniform size. The resulting output representation consisted of 768 tokens, which collectively represented each sentence.

4.5. Concatenation Layer

In this layer, the output representations of BERT for the news and comment branches are combined with features derived from the feature-extraction unit. By combining these elements, the model can effectively identify correlations between the BERT-generated representations and the extracted features, thereby enhancing its ability to accurately classify rumors and non-rumors.

4.6. Dense Layer

In the final stage of the model architecture, the output of the concatenation layer is fed into a dense layer. Then, the resulting output is passed through a Sigmoid layer, which predicts the output class as either rumor or non-rumor by generating values between 0 and 1.

4.7. Experimental Setup

In this research, a quantitative experimental approach was utilized to evaluate emotion features for rumor detection in the Arabic language. The experiments used the Python programming PyTorch library on Google Colab (Python development environment). This section outlines the experiments that were carried out in classifying false news. Hyper-parameter tuning is a crucial aspect of the experimental process, and several factors are considered, including the optimizer, learning rate, number of neurons, and epoch numbers throughout the tuning process on the development dataset. The parameters we utilized are listed below:

- Optimizer: we tested various optimizers: SGD, ADAM, and ADAMW optimizers
- Learning rate: we explored a variety of values between 1×10^{-2} and 1×10^{-6} before deciding that 1×10^{-6} was the best value for fine-tuning and 1×10^{-4} was the best value for feature-based learning.
- Neuron numbers: several numbers were evaluated for the dense layer, including 256, 512, 1024, and 2048 neurons. After multiple tests, we settled on 1024 neurons since it provided the highest accuracy.
- Epoch numbers: we investigated with a range of epoch numbers, from 1 to 50. We employed an early stopping technique. This involves halting the training phase before the validation error arises, and the model's performance cannot be enhanced to prevent overfitting and underfitting problems.

To prevent overfitting and ensure the generalization of the model, it was crucial to monitor the accuracy and loss values on both the training and validation sets during the training phase. To achieve this, we examined the accuracy and loss values at each training step to identify when the model was starting to overfit.

4.8. Evaluation Measurements

In assessing the performance of a proposed model, evaluation metrics such as accuracy, recall, precision, and F1-score are employed. Equations for these measurements are provided below:

$$\text{Accuracy} = \frac{\text{Instances predicted correctly}}{\text{Total number of instances}} \tag{1}$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{2}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{3}$$

$$\text{F1-score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

5. Results and Discussion

In this section, we present the results and a discussion of our experiments designed to evaluate the proposed approach for investigating the role of sentiments and emotional features in enhancing the accuracy of false news detection. We summarize our observations and discuss our major findings, primarily addressing the research questions in Section 1.

In order to address Research Question 1 (RQ1), we evaluated performance by comparing the approaches through three well-established ML models, with the aim of assessing their efficacy in differentiating between rumors and non-rumors. The results, as presented in Table 4, illustrate the impact of emotions extracted from news articles and comments, as well as the combination of both emotion types. Although the primary focus of research in this area is on news-based emotions, the findings indicate that response-based emotional features contribute to the accurate detection of rumors, often surpassing the performance of news-based emotional features. This observation can be attributed to the fact that ordinary individuals are more expressive in conveying their emotions when reacting to news articles, whereas news reports typically strive for neutrality and objectivity.

Generally, the results reveal that the AraNet model performed well in terms of news emotions, which can be attributed to its training on state-of-the-art language models, such as BERT. In contrast, SenticNet did not perform well. Moreover, the AraNet pre-trained model exhibited superior accuracy. Among the three classifiers, the RF classifier consistently produced favorable results, particularly as the number of features increased in most cases. Additionally, the combination of both news-based and comment-based emotions yielded improved results compared to the individual performance of either emotion type. For instance, when all features were utilized, comment-based performance improved from

65.4% to 67.9% in terms of accuracy when using the SVM classifier. Additionally, the inclusion of gap features between news and comments generally led to a positive impact on performance enhancement. These findings align with the study by [43], which highlighted the significance of gap features in emotion-based rumor detection.

Table 4. News and comments emotions performance in rumor detection.

Approach	Features	SVM		LR		RF	
		Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score
News-based emotions	LeXmo	58%	55.3%	56.8%	58.8%	58%	58.8%
	NRCLex	49.4%	64.3%	60.5%	64.4%	63%	62.5%
	SenticNet	45.7%	59.3%	51.9%	48%	44.4%	45.8%
	AraNet + CAMeL	63%	68.1%	54.3%	54.3%	67.9%	71.7%
	ALL Emotions	64.2%	64.2%	64.2%	62.8%	60.5%	62.8%
comments-based emotions	LeXmo	59.3%	60.2%	54.3%	56.5%	58%	58.5%
	NRCLex	55.6%	62.5%	59.2%	62.9%	64.2%	62.3%
	SenticNet	48.1%	51.2%	53.1%	52.5%	58%	56.4%
	AraNet + CAMeL	64.2%	63.3%	59.3%	60.2%	65.4%	64.1%
	ALL Emotions	65.4%	65%	56.8%	58.8%	67.9%	69%
Combining-based emotions	LeXmo	64.2%	67.4%	55.6%	57.1%	63%	64.3%
	NRCLex	53.1%	64.8%	60.5%	63.6%	71.6%	72.3%
	SenticNet	53.1%	52.5%	56.8%	57.8%	53.1%	53.7%
	AraNet + CAMeL	63%	65.9%	67.9%	69.8%	67.6%	69%
	ALL Emotions	67.9%	69.8%	66.7%	69.7%	67.9%	69.7%
	ALL + Gap emotions	67.9%	70.5%	69.1%	71.3%	69.1%	70.6%

To address Research Question 2 (RQ2), we investigated the impact of combining textual features with emotional features. To this end, we extracted sentence representations using several contextual embeddings. To ascertain the importance of user comments in news-based rumor detection, we examined the effect of including textual features of comments. Table 5 presents the performance of the emotion-based rumor detection approaches, including the fine-tuning and feature-based methods, compared with two baselines. Baseline 1 considers only news textual features without comment-based features and emotional features, while Baseline 2 incorporates both news and comment textual features without the emotional branch.

The results indicate that incorporating the textual features of comments significantly enhances rumor detection performance. Overall, the inclusion of emotions positively impacted all models, regardless of whether they were fine-tuned or feature-based. Regarding emotions, AraBERT-Twitter demonstrated better performance with basic emotions when utilizing the last layers and mean pooling approaches. The gap emotions did not notably improve the performance of either model, except when using concatenation. In contrast, the feature-based approach benefited more from the gap emotions, as its performance surpassed that of the basic emotions in most cases. The findings revealed that the utilization of concatenation and mean pooling of the last four layers did not yield a significant improvement in performance. In light of these results, a subsequent experiment was carried out, concentrating solely on the last layer.

Table 5. PLMs performance with the incorporation of emotion features in rumor detection.

Approach	Model	Features	Last Layer (cls)		Concatenation		Mean Pooling	
			Accuracy	F-Score	Accuracy	F-Score	Accuracy	F-Score
Fine-tuning approach	AraBERT-Twitter	News Features	66.66%	64.01%	62.96%	59.71%	65.43%	61.79%
		News + Comments Features	77.77%	78.81%	82.71%	82.59%	79.01%	78.54%
		(+) Basic Features	85.18%	85.13%	82.71%	82.59%	81.48%	80.91%
		(+) All Features	74.04%	72.73%	83.95%	83.79%	70.37%	67.77%
	MARBERT	News Features	59.25%	51.72%	56.79%	46.15%	69.13%	66.17%
		News + Comments Features	87.65%	87.64%	87.65%	87.64%	85.18%	85.18%
		(+) Basic Features	87.65%	87.64%	88.88%	88.86%	88.88%	88.86%
		(+) All Features	87.65%	87.64%	88.88%	88.86%	86.41%	86.41%
Features-based approach	AraBERT-Twitter	News Features	66.66%	66.15%	69.13%	68.44%	69.13%	65.59%
		News + Comments Features	77.77%	77.77%	77.77%	77.5%	81.48%	81.38%
		(+) Basic Features	81.48%	81.48%	80.24%	80.24%	82.71%	82.65%
		(+) All Features	80.24%	80.24%	81.48%	81.47%	82.71%	82.65%
	MARBERT	News Features	65.43%	61.79%	65.43%	61.79%	65.43%	61.79%
		News + Comments Features	79.01%	78.81%	69.13%	66.68%	83.95%	83.79%
		(+) Basic Features	83.95%	83.86%	76.54%	75.59%	70.37%	68.65%
		(+) All Features	83.95%	83.91%	83.95%	83.91%	82.71%	82.26%

The study conducted a *t*-test to compare the results presented in Table 5 with Baseline 2, following an NLP-specific guide in [77]. A significance level of 0.05 was employed, which means that differences in results between adding emotional features and Baseline 2 were statistically significant if the p-value was less than 0.05. The analysis indicated that there was no statistical significance in the utilization of gap features for either type. However, for AraBERT-Twitter, statistical significance was observed with a p-value of 0.00062 when using basic features. This finding may be attributed to the capacity of MARBERT to comprehend emotions and sentiments proficiently, rendering the inclusion of explicit features redundant in order to predict them. When compared to the performance of emotion-based rumor detection in English-language studies [41–45], the performance of emotions in the Arabic language was found to be lower than expected. This could be due to the use of PLMs, which present a challenge as they possess the ability to discern context and emotions. Another possible explanation could be the suboptimal performance of available Arabic-language methods for emotion extraction and the lack of a good emotions-based lexicon that forces us to translate text into English and may lead to translation inaccuracies and cultural differences in emotional expression.

In this study, we also explored the effect of incorporating user features in an emotion-based rumor detection system. The results presented in Table 6 demonstrate that the inclusion of user-centric features significantly improves the model's performance in identifying and mitigating the dissemination of misinformation. This finding underscores the importance of developing more accurate and robust rumor detection algorithms that consider not only emotional features but also the characteristics of their authors.

Table 6. Emotion-based language model results, augmented with user and content features, for rumor detection.

Approach	Model	Features	Accuracy	F-Score
Fine-tuning approach	AraBERT-Twitter	Emotion-based model	74.04%	72.73%
		(+) user and contents feature	86.41%	86.34%
	MARBERT	Emotion-based model	87.65%	87.64%
		(+) user and contents feature	85.18%	85.13%
Features-based approach	AraBERT-Twitter	Emotion-based model	80.24%	80.24%
		(+) user and contents feature	81.48%	81.44%
	MARBERT	Emotion-based model	83.95%	83.91%
		(+) user and contents feature	88.88%	88.83%

Emotions that are elicited by news authors and the public are typically sensational and provocative [44,45]. Some forms of false information elicit emotions that are similar across languages. For instance, ironic news tends to evoke positive emotions in various languages. However, other forms of false information differ from language to language due to cultural differences. For example, rumors in non-Muslim countries about the Islamic religion often carry negative sentiments and evoke feelings of intimidation, while religious rumors in Islamic countries tend to be positive and inspire feelings of admiration and reflection. This inspired us to investigate the role of emotions in Arabic fake news. Several studies have recently been undertaken to analyze the language of false information [41,42]. In this study, we will examine it from different perspectives, expanding the scope to include comments and news perspectives in Arabic languages. Based on previous experiments, we noted that the AraNet model produced good results across all types; thus, it was chosen to finalize the analysis and distribution across types. Moreover, user and content features were examined to demonstrate their contribution to improved performance, allowing researchers to investigate potential avenues for enhancing and incorporating these features into current and new rumor detection methods.

A comparative analysis was conducted to comprehend the role of emotions, user, and content features in Arabic social media to distinguish between true and false news and facilitate answering question **QR3**, including the following:

Top features in distinguishing false and true rumors: In this study, we employed the RF Classifier to show feature contribution, where a higher score indicates greater relevance to the target variable in the dataset. Figure 3 depicts the most informative user and content features for news and comments, respectively. The most important features for detecting rumors from the news publisher's perspective were the number of followers, followed by the number of lists the account subscribed to, and the presence of a link in the account's description. In contrast, the features contributing to the commenter's perspective were the number of characters in the name, favorite count, and the number of words in the description.

Regarding emotional features, we observed that the most important emotions for detecting rumors in the news were neutral, negative sentiments, and anticipation. This aligns with social studies, which propose that rumors propagate negativity, while true news is generally unbiased towards specific moods. Conversely, we found that the most helpful emotions for detecting rumors from the comments perspective were positive and neutral, followed by fear and anger emotions, which had almost equal importance in emotions-based rumor detection. This indicates that replies to accurate news tend to express positive and neutral sentiments as they discuss the news, while responses to false information often employ words of fear or anger to convey rejection. Examining both perspectives broadly, we find that the effectiveness and performance of each feature vary significantly, depending on whether it is presented by the publishers or by commenters. For instance,

account verification status can be beneficial for publishers, but it was the least helpful feature in terms of comments. This highlights the importance of separating the two types and extracting features from each perspective individually rather than considering the overall total regardless of its type or disregarding commenter features entirely.

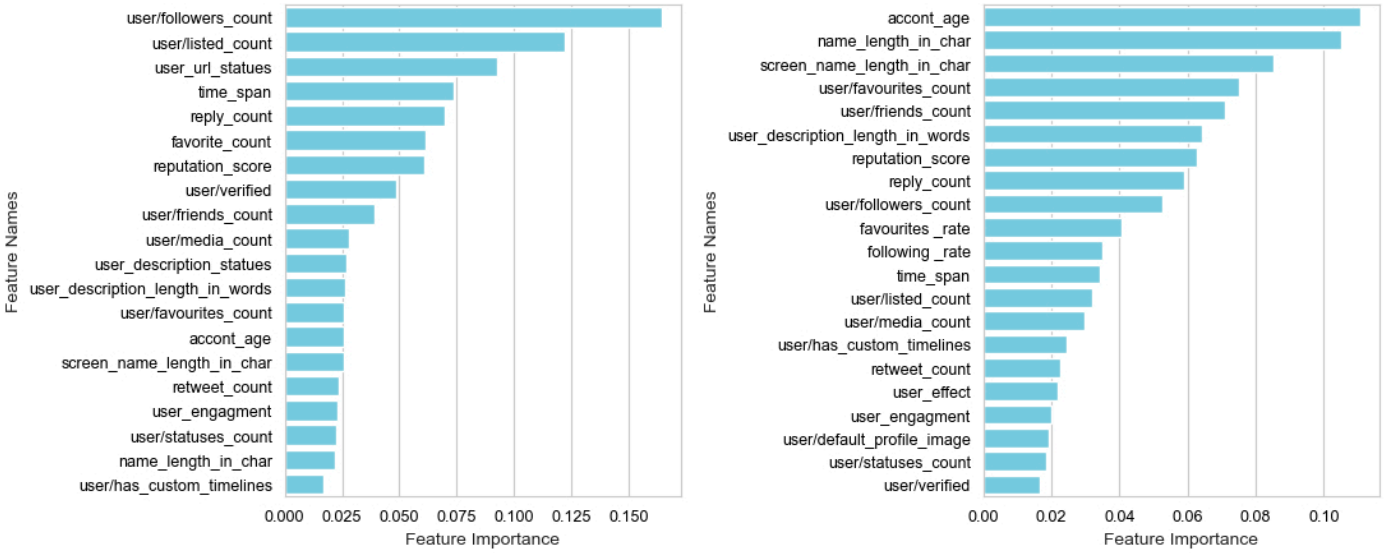

Figure 3. RF feature importance scores for user-centric and content features in rumor detection. (**a**) Informative features on the news disseminators' side. (**b**) Informative features on the commenters' side.

Feature importance distribution similarity in news content and comments perspective: Upon analyzing the data, we concluded that the ability of a feature to detect rumors varied considerably depending on whether it was made in comments or by news publishers. To test our hypothesis, we used the Chi-Square score, which demonstrates the strength of the relationship between two variables. Figure 4 shows feature correlation using the Chi-Square test, indicating a clear divergence in the prominence of each feature per type. For emotional features, our findings reveal similarities in the distribution of trust, neutrality, happiness, and sadness across both news content and comments perspectives. However, there are noteworthy distinctions. For example, anger and disgust hold minimal importance in news content, while surprise and negative emotions play a more significant role. This can be attributed to false news writers employing negative and surprising words to capture public attention. On the other hand, concerning user and content features, we observed a significant number of features in news posts that were correlated with class, such as follower count, list count, account verification status, URL status, and reply count. These results are consistent with the more informative features identified using RF in previous analyses. On the other hand, from the comments' perspective, features such as favorites count, account age, follower count, and screen name length were correlated with class. These findings revealed a marked difference in the prominence of features for each perspective, with certain features proving to be more effective in the news publisher context than in the replies context.

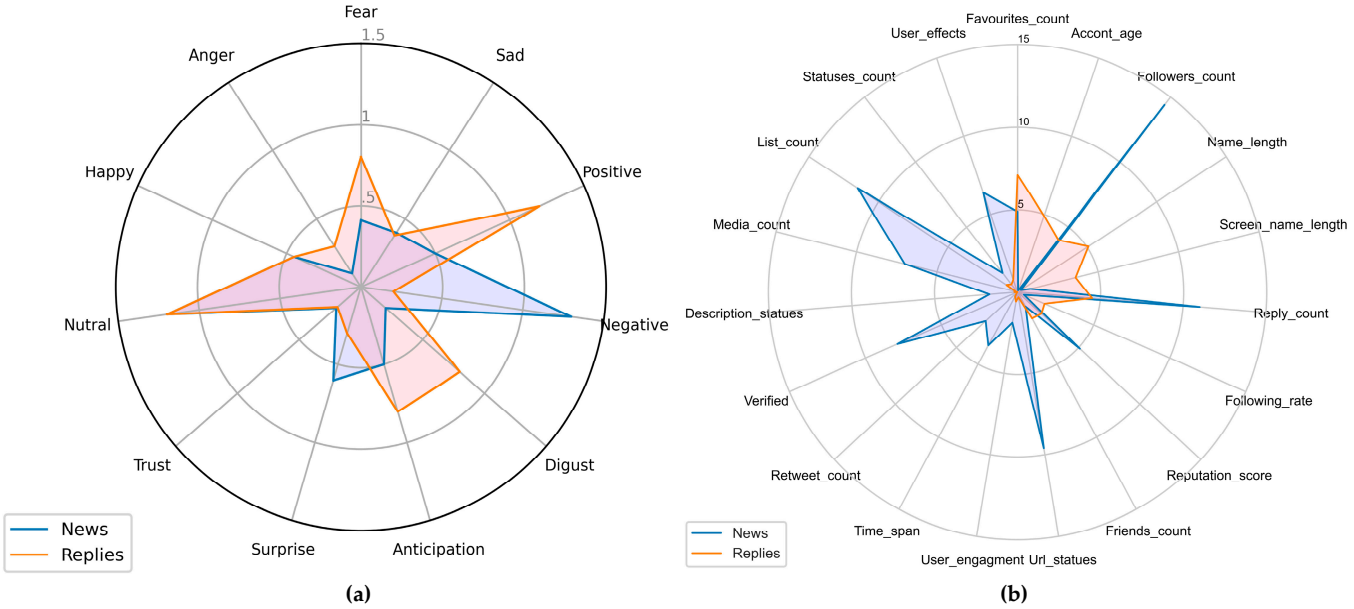

Figure 4. Features chi-square score (**a**) Emotional features chi-square scores for news content and comments. (**b**) User-centric and content features chi-square scores for news content and comments.

Emotions' difference according to the news topic: The intensity of emotions varies with respect to the topic being discussed. Figures 5 and 6 illustrate this; for example, we see in real news content that the intensity of positive sentiment in religious rumors is quite high, in contrast to true religious news. When it comes to health news, real news surpasses rumors in terms of positive and neutral sentiments and anticipation. Sad emotions are also strong in false news in all genres except religious rumors, where the author attempts to instill sadness in false information to attract attention. Additionally, trust and neutrality are similar in all types, with trust being higher in all kinds of rumors to deceive the reader into believing the news is real, while neutral sentiment is higher in all types of real news. On the other hand, the variation in intensity in replies to rumors is more noticeable than in the news. For example, we see a rise in trust emotions in religious and health rumors, whereas it is lower in political and social real news.

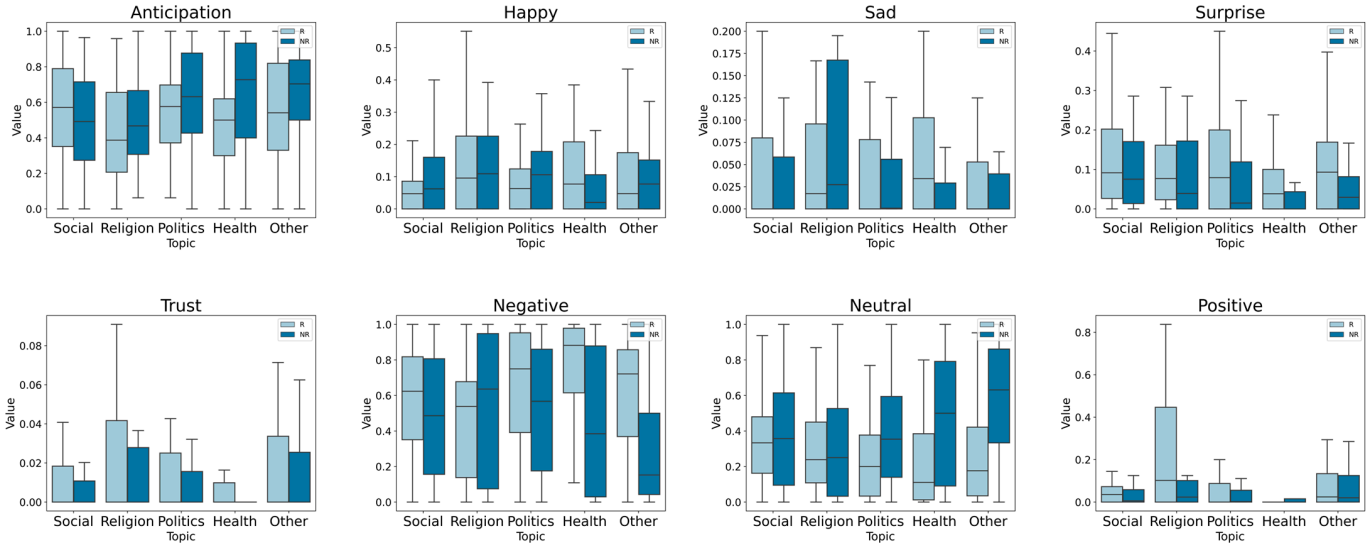

Figure 5. Features density in news content on various topics.

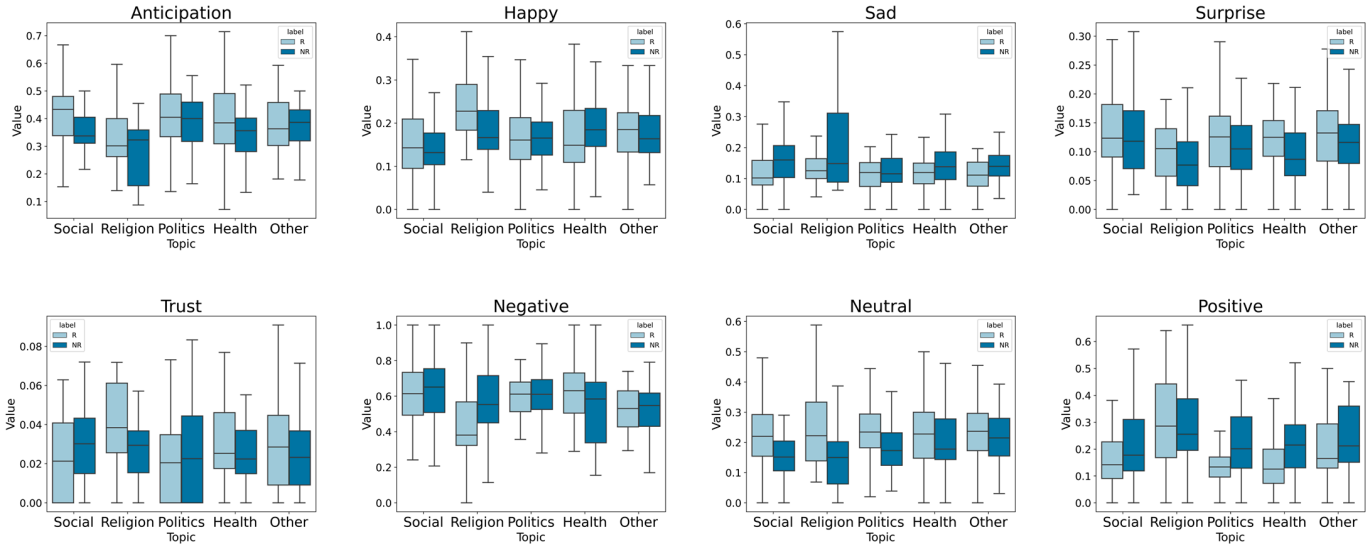

Figure 6. Features density in replies on various topics.

In conclusion, the findings revealed a significant disparity in the density of emotions according to whether they occurred in news or comments, highlighting the necessity of incorporating textual contents via topic modeling or language models that can infer the news topic for accurate rumor detection. In summary, this study demonstrates the varying importance of emotions, user, and content features in detecting rumors from both the news publisher and commenter perspectives. Consequently, it is crucial to consider both aspects when identifying rumors to optimize the use of available information.

6. Conclusions and Future Work

In this study, we investigated the influence of various user and content attributes, as well as affective dimensions, on rumor detection from two distinct vantage points: news sources and user comments. To accomplish this, we assembled a novel dataset derived from the Twitter platform. Our experimental analysis examined the distribution of these features and identified the most impactful attributes for each perspective. The findings revealed that sentiments, emotions, and user features exhibit notable differences between rumor and non-rumor instances. Moreover, the significance of these features varies depending on whether they originate from news content or user comments. Nonetheless, one of the primary limitations of this study pertains to the inadequacy of the available techniques for extracting emotions in the Arabic language.

These observations underscore the potential utility of emotional and user attributes in the identification of false news, culminating in our proposed rumor detection algorithms. One potential avenue for further exploration is the incorporation of affective resources into deep learning models, a technique that has been previously applied to the English language [78]. Additionally, it may be worthwhile to investigate the potential benefits of including other psycholinguistic traits, such as personality, which have demonstrated utility in certain prior studies [79]. Furthermore, recent literature [80–86] suggests that the stance of user comments can serve as a primary indicator for verifying deceptive rumors. However, prior research endeavors in crafting rumor detection systems have predominantly focused on stance, neglecting the nuanced interplay of emotions and their potential contributions. Additionally, emotions have been primarily employed as supplementary features for stance identification. The weak association between emotions and stance, as reported in [87], implies that this relationship is limited and may contribute to the propagation of errors. Consequently, there is considerable scope for enhancing the efficacy of existing false rumor verification techniques. As a result, current false rumor verification methods obviously have room for improvement. Future research should delve further into the emotions, and specific user features that contribute to this enhanced performance and explore potential

avenues for refining and optimizing the integration of these features into existing and novel rumor detection methodologies.

Author Contributions: Conceptualization, H.F.A.-S. and H.Z.A.-D.; data curation, H.F.A.-S.; formal analysis, H.F.A.-S.; funding acquisition, H.Z.A.-D.; methodology, H.F.A.-S.; supervision, H.Z.A.-D.; validation, H.Z.A.-D.; writing—original draft, H.F.A.-S.; writing—review and editing, H.Z.A.-D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Deanship of Scientific Research, King Saud University.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Tweet IDs and user IDs are available upon request via correspondence email for academic use, in accordance with Twitter's Developer Policy.

Acknowledgments: The authors would like to thank the Deanship of scientific research King Saud University for funding and supporting this research through the initiative of the DSR Graduate Students Research Support (GSR).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mason, W.; Katrena, E.M. News Consumption across Social Media in 2021. Available online: http://www.pewresearch.org (accessed on 21 September 2021).
2. Antonakaki, D.; Fragopoulou, P.; Ioannidis, S. A survey of Twitter research: Data model, graph structure, sentiment analysis and attacks. *Expert. Syst. Appl.* **2021**, *164*, 114006. [CrossRef]
3. Shu, K.; Sliva, A.; Wang, S.; Tang, J.; Liu, H. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explor. Newsl.* **2017**, *19*, 22–36. [CrossRef]
4. Gupta, A.; Kumaraguru, P.; Castillo, C.; Meier, P. Tweetcred: Real-time credibility assessment of content on twitter. *Lect. Notes Comput. Sci.* **2014**, *8851*, 228–243. [CrossRef]
5. Allcott, H.; Gentzkow, M. Nber Working Paper Series Social Media and Fake News in the 2016 Election. *J. Econ. Perspect.* **2017**, *31*, 211–236. [CrossRef]
6. Tandoc, E.C.; Lim, Z.W.; Ling, R. Defining 'Fake News': A typology of scholarly definitions. *Digit. J.* **2018**, *6*, 137–153. [CrossRef]
7. Samy-Tayie, S.; Tejedor, S.; Pulido, C. News literacy and online news between Egyptian and Spanish youth: Fake news, hate speech and trust in the media. *Comunicar* **2023**, *31*, 73–87. [CrossRef]
8. Bovet, A.; Makse, H.A. Influence of fake news in Twitter during the 2016 US presidential election. *Nat. Commun.* **2019**, *10*, 1657. [CrossRef]
9. NY Times. As Fake News Spreads Lies, More Readers Shrug at the Truth. 2016. Available online: https://www.nytimes.com/2016/12/06/us/fake-news-partisan-republican-democrat.html (accessed on 12 January 2023).
10. Zubiaga, A.; Kochkina, E.; Liakata, M.; Procter, R.; Lukasik, M. Stance classification in rumours as a sequential task exploiting the tree structure of social media conversations. In Proceedings of the COLING 2016—26th International Conference on Computational Linguistics, Proceedings of COLING 2016: Technical Papers, Osaka, Japan, 11–16 December 2016; pp. 2438–2448.
11. Vosoughi, S.; Roy, D.; Aral, S. The spread of true and false news online. *Science* **2018**, *1151*, 1146–1151. [CrossRef]
12. Kumar, S.; Shah, N. False Information on Web and Social. Media: A Survey. *arXiv* **2018**, arXiv:1804.08559.
13. Ruffo, G.; Semeraro, A.; Giachanou, A.; Rosso, P. Studying fake news spreading, polarization dynamics, and manipulation by bots: A tale of networks and language. *Comput. Sci. Rev.* **2023**, *47*, 100531. [CrossRef]
14. Meel, P.; Vishwakarma, D.K. Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. *Expert. Syst. Appl.* **2020**, *153*, 112986. [CrossRef]
15. Zannettou, S.; Sirivianos, M.; Blackburn, J.; Kourtellis, N. The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans. *J. Data Inf. Qual.* **2019**, *11*, 1–37. [CrossRef]
16. Liang, G.; He, W.; Xu, C.; Chen, L.; Zeng, J. Rumor Identification in Microblogging Systems Based on Users' Behavior. *IEEE Trans. Comput. Soc. Syst.* **2015**, *2*, 99–108. [CrossRef]
17. Zubiaga, A.; Aker, A.; Bontcheva, K.; Liakata, M.; Procter, R. Detection and resolution of rumours in social media: A survey. *ACM Comput. Surv.* **2017**, *51*, 1–36. [CrossRef]
18. Derczynski, L.; Bontcheva, K.; Liakata, M. SemEval-2017 Task 8: RumourEval: Determining rumour veracity and support for rumours. In Proceedings of the 11th International Workshop on Semantic Evaluations, Vancouver, BC, Canada, 3–4 August 2017; Association for Computational Linguistics: Stroudsburg, PA, USA, 2017. [CrossRef]
19. Jaeger, M.E.; Anthony, S.; Rosnow, R.L. Who Hears What from Whom and with What Effect: A Study of Rumor. *Pers. Soc. Psychol. Bull.* **1980**, *6*, 473–478. [CrossRef]

20. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the NAACL HLT 2019—2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; Volume 1, pp. 4171–4186.
21. Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. Improving Language Understanding by Generative Pre-Training. 2018. *In preprint*. Available online: https://scholar.google.com/scholar?q=Improving+Language+Understanding+by+Generative+Pre-Training&hl=ar&as_sdt=0&as_vis=1&oi=scholart (accessed on 16 May 2023).
22. Ott, Y.L.M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv* **2019**, arXiv:1907.11692.
23. Yang, Z.; Dai, Z.; Yang, Y.; Carbonell, J.; Salakhutdinov, R.R.; Le, Q.V. XLNet: Generalized Autoregressive Pretraining for Language Understanding. In *Advances in Neural Information Processing Systems*; Wallach, H., Larochelle, H., Beygelzimer, A., Alché-Buc, F., Fox, E., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2019.
24. Giurgiu, N.H.A.; Jastrzebski, S.; Morrone, B.; Laroussilhe, Q. Parameter-Efficient Transfer Learning for NLP. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019.
25. Howard, J.; Ruder, S. Universal Language Model Fine-tuning for Text Classification. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 15–20 July 2018; pp. 328–339.
26. Khandelwal, A.C.K.; Goyal, N.; Chaudhary, V.; Wenzek, G.; Guzmán, F.; Grave, E.; Ott, M.; Zettlemoyer, L.; Stoyanov, V. Unsupervised cross-lingual representation learning at scale. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 8440–8451. [CrossRef]
27. Agerri, R.; Campos, J.; Barrena, A.; Saralegi, X.; Soroa, A.; Agirre, E. Give your text representation models some Love: The case for Basque. In Proceedings of the LREC 2020—12th International Conference on Language Resources and Evaluation, Conference Proceedings, Marseille, France, 11–16 May 2020; pp. 4781–4788.
28. Antoun, W.; Baly, F.; Hajj, H. AraBERT: Transformer-based Model for Arabic Language Understanding. In Proceedings of the LREC 2020 Workshop Language Resources and Evaluation Conference, Marseille, France, 11–16 May 2020.
29. Abdul-Mageed, M.; Elmadany, A.; Nagoudi, E.M.B. ARBERT & MARBERT: Deep Bidirectional Transformers for Arabic. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Online, 1 August 2021; pp. 7088–7105.
30. Ekman, P. An argument for basic emotions. *Cogn. Emot.* **1992**, *6*, 169–200. [CrossRef]
31. Plutchic, R. General Psychoevolutionary Theory of Emotion. In *Theories of Emotion*; Plutchik, R., Kellerman, H., Eds.; Academic Press: Cambridge, MA, USA, 1980; pp. 3–33. [CrossRef]
32. Al-A'abed, M.; Al-Ayyoub, M. A Lexicon—Based Approach for Emotion Analysis of Arabic Social Media Content. In Proceedings of the International Computer Sciences and Informatics Conference (ICSIC 2016), Amman, Jordan, 12–13 January 2016.
33. El Jundi, G.B.O.; Khaddaj, A.; Maarouf, A.; Kain, R.; Hajj, H.; El-Hajj, W. EMA at SemEval-2018 Task 1: Emotion Mining for Arabic. In Proceedings of the SemEval@NAACL-HLT, New Orleans, LA, USA, 5–6 June 2018.
34. Al-Khatib, A.; El-Beltagy, S.R. Emotional tone detection in Arabic tweets. *Lect. Notes Comput. Sci.* **2018**, *10762*, 105–114. [CrossRef]
35. Alswaidan, N.; Menai, M.E.B. Hybrid Feature Model for Emotion Recognition in Arabic Text. *IEEE Access* **2020**, *8*, 37843–37854. [CrossRef]
36. Abdul-Mageed, M.; Zhang, C.; Hashemi, A.; Nagoudi, E.M.B. AraNet: A deep learning toolkit for arabic social media. In Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection, Marseille, France, 12 May 2019; European Language Resource Association: Paris, France, 2019; pp. 16–23.
37. Tian, L.; Zhang, X.; Wang, Y.; Liu, H. Early Detection of Rumours on Twitter via Stance Transfer Learning. *Lect. Notes Comput. Sci.* **2020**, *12035*, 575–588. [CrossRef]
38. Kwon, S.; Cha, M.; Jung, K.; Chen, W.; Wang, Y. Prominent features of rumor propagation in online social media. In Proceedings of the IEEE International Conference on Data Mining, ICDM, Dallas, TX, USA, 7–10 December 2013; pp. 1103–1108. [CrossRef]
39. Wu, K.; Yang, S.; Zhu, K.Q. False rumors detection on Sina Weibo by propagation structures. In Proceedings of the 2015 IEEE 31st International Conference on Data Engineering, Seoul, Republic of Korea, 13–17 April 2015; pp. 651–662. [CrossRef]
40. Jin, Z.; Cao, J.; Jiang, Y.-G.; Zhang, Y. News Credibility Evaluation on Microblog with a Hierarchical Propagation Model. In Proceedings of the 2014 IEEE International Conference on Data Mining, in ICDM'14, Shenzhen, China, 14–17 December 2014; IEEE Computer Society: Washington, DC, USA, 2014; pp. 230–239. [CrossRef]
41. Ghanem, B.; Rosso, P.; Rangel, F. An Emotional Analysis of False Information in Social Media and News Articles. *ACM Trans. Internet Technol.* **2020**, *20*, 1–17. [CrossRef]
42. Guo, C.; Cao, J.; Zhang, X.; Shu, K.; Yu, M. Exploiting emotions for fake news detection on social media. *arXiv* **2019**, arXiv:1903.01728.
43. Zhang, X.; Cao, J.; Li, X.; Sheng, Q.; Zhong, L.; Shu, K. Mining Dual Emotion for Fake News Detection. In *WWW'21: Proceedings of the Web Conference 2021*; Association for Computing Machinery: New York, NY, USA, 2021; pp. 3465–3476. [CrossRef]
44. Anoop, K.; Deepak, P.; Lajish, V. Emotion Cognizance Improves Health Fake News Identification. In Proceedings of the 24th Symposium on International Database Engineering & Applications, Seoul, Republic of Korea, 12–14 August 2020.

45. Giachanou, A.; Rosso, P.; Crestani, F. Leveraging Emotional Signals for Credibility Detection. In *SIGIR'19: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*; Association for Computing Machinery: New York, NY, USA, 2019; pp. 877–880.
46. Wu, L.; Rao, Y. Adaptive interaction fusion networks for fake news detection. *Front. Artif. Intell. Appl.* **2020**, *325*, 2220–2227. [CrossRef]
47. Hamed, S.K.; Aziz, M.J.A.; Yaakub, M.R. Fake News Detection Model on Social Media by Leveraging Sentiment Analysis of News Content and Emotion Analysis of Users' Comments. *Sensors* **2023**, *23*, 1748. [CrossRef] [PubMed]
48. Ghanem, B.; Ponzetto, S.P.; Rosso, P.; Rangel, F. FakeFlow: Fake News Detection by Modeling the Flow of Affective Information. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics, Kyiv, Ukraine, 19–23 April 2021; pp. 679–689.
49. United Nations Top Languges. 2021. Available online: https://www.un.org/en/our-work/official-languages (accessed on 12 January 2023).
50. Alzanin, S.M.; Azmi, A.M. Rumor detection in Arabic tweets using semi-supervised and unsupervised expectation–maximization. *Knowl. Based Syst.* **2019**, *185*, 104945. [CrossRef]
51. Sabbeh, S.F.; Baatwah, S.Y. Arabic news credibility on twitter: An enhanced model using hybrid features. *J. Theor. Appl. Inf. Technol.* **2018**, *96*, 2327–2338.
52. Jardaneh, G.; Abdelhaq, H.; Buzz, M.; Johnson, D. Classifying Arabic tweets based on credibility using content and user features. In Proceedings of the 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology, JEEIT, Amman, Jordan, 9–11 April 2019; pp. 596–601. [CrossRef]
53. Gumaei, A.; Al-Rakhami, M.S.; Hassan, M.M.; De Albuquerque, V.H.C.; Camacho, D. An Effective Approach for Rumor Detection of Arabic Tweets Using eXtreme Gradient Boosting Method. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **2022**, *21*, 1–16. [CrossRef]
54. Al-Khalifa, H.S.; Al-Eidan, R.M. An experimental system for measuring the credibility of news content in Twitter. *Int. J. Web Inf. Syst.* **2011**, *7*, 130–151. [CrossRef]
55. Al-yahya, M.; Al-khalifa, H.; Al-baity, H.; Alsaeed, D.; Essam, A. Arabic Fake News Detection: Comparative Study of Neural Networks and Transformer-Based Approaches. *Complexity* **2021**, *2021*, 5516945. [CrossRef]
56. Alsudias, L.; Rayson, P. COVID-19 and Arabic Twitter: How can Arab World Governments and Public Health Organizations Learn from Social Media? In Proceedings of the 1st Workshop on NLP for COVID-19 at ACL, Online, 9–10 July 2020; pp. 1–9.
57. Mahlous, A.R.; Al-Laith, A. Fake News Detection in Arabic Tweets during the COVID-19 Pandemic. *Int. J. Adv. Comput. Sci. Appl.* **2021**, *12*, 778–788. [CrossRef]
58. Fouad, K.M.; Sabbeh, S.F.; Medhat, W. Arabic fake news detection using deep learning. *Comput. Mater. Contin.* **2022**, *71*, 3647–3665. [CrossRef]
59. Amoudi, G.; Albalawi, R.; Baothman, F.; Jamal, A.; Alghamdi, H.; Alhothali, A. Arabic rumor detection: A comparative study. *Alex. Eng. J.* **2022**, *61*, 12511–12523. [CrossRef]
60. Nassif, A.B.; Elnagar, A.; Elgendy, O.; Afadar, Y. Arabic fake news detection based on deep contextualized embedding models. *Neural Comput. Appl.* **2022**, *34*, 16019–16032. [CrossRef]
61. Albalawi, R.M.; Jamal, A.T.; Khadidos, A.O.; Alhothali, A.M. Multimodal Arabic Rumors Detection. *IEEE Access* **2023**, *11*, 9716–9730. [CrossRef]
62. Harrag, F.; Djahli, M.K. Arabic Fake News Detection: A Fact Checking Based Deep Learning Approach. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **2022**, *21*, 1–34. [CrossRef]
63. Giachanou, A.; Zhang, G.; Rosso, P. Multimodal Multi-image Fake News Detection. In Proceedings of the 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), Sydney, Australia, 6–9 October 2020; pp. 647–654. [CrossRef]
64. Gorrell, G.; Bontcheva, K.; Derczynski, L.; Kochkina, E.; Liakata, M.; Zubiaga, A. RumourEval 2019: Determining rumour veracity and support for rumours. In Proceedings of the 13th International Workshop on Semantic Evaluation, Minneapolis, MN, USA, 6–7 June 2019; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 845–854.
65. Castillo, C.; Mendoza, M.; Poblete, B. Information Credibility on Twitter. In *WWW '11: Proceedings of the 20th International Conference on World Wide Web*; Association for Computing Machinery: New York, NY, USA, 2011; pp. 675–684. [CrossRef]
66. Yang, F.; Liu, Y.; Yu, X.; Yang, M. Automatic Detection of Rumor on Sina Weibo. In Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics, Beijing, China, 12–16 August 2012; Volume 2, pp. 13:1–13:7.
67. Ghanem, B.; Cignarella, A.T.; Bosco, C.; Rosso, P.; Pardo, F.M.R. UPV-28-UNITO at SemEval-2019 Task 7: Exploiting Post's Nesting and Syntax Information for Rumor Stance Classification. In Proceedings of the 13th International Workshop on Semantic Evaluation, Minneapolis, MN, USA, 6–7 June 2019; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 1125–1131. [CrossRef]
68. Kochkina, A.Z.E.; Liakata, M.; Procter, R.; Lukasik, M.; Bontcheva, K.; Cohn, T.; Augenstein, I. Discourse-aware rumour stance classification in social media using sequential classifiers. *Inf. Process. Manag.* **2018**, *54*, 273–290. [CrossRef]
69. Alsaif, H.; Aldossari, H. Review of stance detection for rumor verification in social media. *Eng. Appl. Artif. Intell.* **2023**, *119*, 105801. [CrossRef]

70. Singh, V.; Narayan, S.; Akhtar, M.S.; Ekbal, A.; Bhattacharyya, P. IITP at SemEval-2017 Task 8: A Supervised Approach for Rumour Evaluation. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), Vancouver, BC, Canada, 3–4 August 2017; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 497–501. [CrossRef]
71. Bahuleyan, H.; Vechtomova, O. Detecting Stance towards Rumours with with Topic Independent Features. In Proceedings of the 11th International Workshop on Semantic Evaluations, Vancouver, BC, Canada, 3–4 August 2017; pp. 145–183. [CrossRef]
72. Baris, I.; Schmelzeisen, L.; Staab, S. CLEARumor at SemEval-2019 Task 7: ConvoLving ELMo against rumors. In Proceedings of the 13th International Workshop on Semantic Evaluation, Minneapolis, MN, USA, 6–7 June 2019; pp. 1105–1109. [CrossRef]
73. Islam, M.R.; Muthiah, S.; Ramakrishnan, N. Rumorsleuth: Joint detection of rumor veracity and user stance. In Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social. Networks Analysis and Mining, ASONAM 2019, Vancouver, BC, Canada, 27–30 August 2019; pp. 131–136. [CrossRef]
74. Janchevski, A.; Gievska, S. AndrejJan at SemEval-2019 Task 7: A Fusion Approach for Exploring the Key Factors pertaining to Rumour Analysis. In Proceedings of the 13th International Workshop on Semantic Evaluation, Minneapolis, MN, USA, 6–7 June 2019; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 1083–1089. [CrossRef]
75. Cambria, E.; Poria, S.; Bajpai, R.; Schuller, B. SenticNet 4: A Semantic Resource for Sentiment Analysis Based on Conceptual Primitives. In Proceedings of the COLING 2016, 26th International Conference on Computational Linguistics: Technical Papers, Osaka, Japan, 11–16 December 2016; The COLING 2016 Organizing Committee: Osaka, Japan, 2016; pp. 2666–2677.
76. Zalmout, O.O.N.; Khalifa, S.; Taji, D.; Oudah, M.; Alhafni, B.; Inoue, G.; Eryani, F.; Erdmann, A.; Habash, N. CAMeL Tools: An Open Source Python Toolkit for Arabic Natural Language Processing. In Proceedings of the Twelfth Language Resources and Evaluation Conference, Marseille, France, 11–16 May 2020; pp. 7022–7032. Available online: http://qatsdemo.cloudapp.net/farasa/ (accessed on 12 January 2023).
77. Dror, R.; Baumer, G.; Shlomov, S.; Reichart, R. The Hitchhiker's Guide to Testing Statistical Significance in Natural Language Processing. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia, 15–20 July 2018; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 1383–1392. [CrossRef]
78. Giachanou, A.; Rosso, P.; Crestani, F. The impact of emotional signals on credibility assessment. *J. Assoc. Inf. Sci. Technol.* **2021**, *72*, 1117–1132. [CrossRef]
79. Giachanou, A.; Ghanem, B.; Ríssola, E.A.; Rosso, P.; Crestani, F.; Oberski, D. The impact of psycholinguistic patterns in discriminating between fake news spreaders and fact checkers. *Data Knowl. Eng.* **2022**, *138*, 101960. [CrossRef]
80. Lillie, A.E.; Middelboe, E.R.; Derczynski, L. Joint Rumour Stance and Veracity. In Proceedings of the 22nd Nordic Conference on Computational Linguistics, Turku, Finland, 30 September–2 October 2019; Linköping University Electronic Press: Linköping, Sweden, 2019; pp. 208–221. [CrossRef]
81. Li, Y.; Scarton, C. Revisiting Rumour Stance Classification: Dealing with Imbalanced Data. In Proceedings of the 3rd International Workshop on Rumours and Deception in Social Media (RDSM), Barcelona, Spain, 13 December 2020; pp. 38–44. Available online: https://www.aclweb.org/anthology/2020.rdsm-1.4 (accessed on 12 January 2023).
82. Kumar, S.; Carley, K.M. Tree LSTMs with convolution units to predict stance and rumor veracity in social media conversations. In Proceedings of the ACL 2019—57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2020; pp. 5047–5058. [CrossRef]
83. Yang, R.; Ma, J.; Lin, H.; Gao, W. A Weakly Supervised Propagation Model for Rumor Verification and Stance Detection with Multiple Instance Learning. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, 11–15 July 2022; ACM: New York, NY, USA, 2022; pp. 1761–1772. [CrossRef]
84. Wei, P.; Xu, N.; Mao, W. Modeling conversation structure and temporal dynamics for jointly predicting rumor stance and veracity. In Proceedings of the EMNLP-IJCNLP 2019—2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Hong Kong, China, 3–7 November 2019; pp. 4787–4798. [CrossRef]
85. Ma, J.; Gao, W.; Wong, K. Detect rumor and stance jointly by neural multi-task learning. In Proceedings of the Companion Proceedings of the Web Conference 2018, Lyon, France, 23–27 April 2018; pp. 585–593. [CrossRef]
86. Xuan, K.; Xia, R. Rumor stance classification via machine learning with text, user and propagation features. In Proceedings of the IEEE International Conference on Data Mining Workshops, ICDMW, Beijing, China, 8–11 November 2019; pp. 560–566. [CrossRef]
87. Schuff, H.; Barnes, J.; Mohme, J.; Padó, S.; Klinger, R. Annotation, Modelling and Analysis of Fine-Grained Emotions on a Stance and Sentiment Detection Corpus. In Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social. Media Analysis, Copenhagen, Denmark, 8 September 2018; pp. 13–23. [CrossRef]