Final version of the published paper. Cite as: *"Asimul Haque and Muhammad Abulaish, An Emotion-Enriched and Psycholinguistics Features-Based Approach for Rumor Detection on Online Social Media, In Proceedings of the 11th International Workshop on Natural Language Processing for Social Media (SocialNLP), Co-located with IJCNLP-AACL, Bali, Indonesia, 1-4 November 2023, pp. 28-37."*

An Emotion-Enriched and Psycholinguistic Features-Based Approach for Rumor Detection on Online Social Media

Asimul Haque
Department of Computer Science
South Asian University
New Delhi-68, India
asimulhaque@gmail.com

Muhammad Abulaish, SMIEEE
Department of Computer Science
South Asian University
New Delhi-68, India
abulaish@sau.int

Abstract

Social media platforms and online communication tools, once lauded as sources of information, unity, and global connectivity, are now breeding grounds for the unprecedented spread of false, misleading, and manipulative information. Rumor is one type of such information. It consists of untrue information and deceptive messages and is constructed to manipulate emotions. Consequently, emotions are crucial in determining the veracity of rumors. In this paper, we introduce an emotion-enhanced and psycholinguistic features-based approach for rumors detection on social social media. It entails detecting rumors utilizing lexicons and various linguistic features-based learning approaches, primarily by extracting the psychological association of words with their emotions. Emotional and psycholinguistic features are extracted from both posts and comments to enhance the approach and make rumor detection more effective. Using word-level GloVe embedding, the semantic relationships between a post and its comments and their underlying emotions are preserved. The proposed method is evaluated on the popular PHEME dataset and compared to various baselines and SOTA methods, demonstrating substantially superior performance for rumor detection on social media platforms.

1 Introduction

The advent and rapid growth of social media enables users to construct, facilitate, and create connections through the exchange of ongoing ideas, emotions, thoughts, and experiences (Zubiaga et al., 2018; Zhou and Zafarani, 2020). Some individuals acquire information, while others use it to publish and disseminate rumors or false narratives. This rumor continues to spread unchecked across social media platforms, posing numerous explicit and implicit threats to social stability and public trust (Zubiaga et al., 2018). Different manifestations of rumors make their detection a growing concern and necessitate a refined approach to addressing them. During the COVID-19 pandemic, it was observed that thousands of rumors proliferated on social media, ranging from the origin of the corona virus to its cures. For example, the nature of rumor in India differs from that in the West. In India, claims ranged from immunity to natural remedies, whereas in the West, anti-vaccine arguments garnered popularity[1]. Recently, misinformation and propaganda regarding the Ukraine war and France unrest have circulated on social media[2]. Social media posts are intentionally created to delude and elicit strong emotions in users in order to spread them across the network. The information published online has real-world consequences. False information spreads more quickly if it is disseminated, as demonstrated by the authors in (Vosoughi et al., 2018).

Numerous fact-checking websites, such as Snopes, PolitiFact, and FactCheck, manually evaluate the accuracy of claims. For instance, PolitiFact evaluates claims using a Truth-O-Meter that indicates the relative accuracy of a statement. To combat misinformation, more than simply manual fact-checking is required. It requires adaptability on multiple fronts, including social media platform regulation, media literacy programs, and the adoption of advanced emerging tools in a responsible manner, in order to address the issue on a large scale. Various governments, international and regional organizations have adopted numerous strategies to combat rumors and fake news on a global scale. For example, (i) the United Nations exhorted users to exercise caution prior to sharing. During the COVID-19 infodemic, social media users were encouraged to consider the 5W's; *Who made it, what is the source of informa-*

[1] http://www.globaltimes.cn/content/1178157.shtml

[2] http://www.bbc.com/news/world-europe-66081671

Proceedings of the 11th International Workshop on Natural Language Processing for Social Media, pages 28–37
November 1, 2023. ©2023 Association for Computational Linguistics

tion, where did it come from, why are you sharing this, and *when was it published*[3]? (ii) the European Union has devised a self-regulatory code of conduct and imposed content moderation requirements on social media platforms such as `Facebook` and `YouTube`, (iii) Singapore has enacted stringent criminal laws to combat online misinformation, (iv) India is combating the proliferation of online misinformation through the Digital India Bill[4]. The evolving hazards are further complicated by technological advances such as `ChatGPT` and `Generative AI`. Automatic detection of rumors is a formidable challenge for the research community and an essential requirement for contemporary society. The varying degrees of falsified information used to deceive users make it more difficult to find a method to prevent the spread of rumors that undermine trust in the social media ecosystem.

In this paper, we relate the psychological theories indicating that users with malicious intent exhibit uncertainty, vagueness, emotion, and indirect forms of expression, whereas trusted users cite primary sources more frequently (Buller and Burgoon, 1996). It has also been observed in the literature that some people use simple words to express their emotions on social media platforms, while others prefer exaggerated and provocative language (Rashkin et al., 2017). In addition, researchers are investigating the representation of emotion and sentiment as structural properties for disseminating false information (Pröllochs et al., 2021; Martel et al., 2020). We augment these works in order to identify the combinations of feature sets that yield the best predictive capabilities for rumors classification. We intend to demonstrate the effectiveness of emotion-related features and combine them with psycholinguistic features in order to facilitate classification tasks. The extraction of emotion-related features is facilitated by lexicons developed by researchers to aid in emotion analysis (Mohammad and Turney, 2013; Mohammad, 2018). The word-level `GloVe` embedding (Pennington et al., 2014) is also used to determine the semantic affinities of posts and comments with their underlying emotion words. Despite the fact that emotion is a key factor in rumor propagation, the scholarly community has paid little attention to comprehending the prevalence of emotions in online posts and comments and their usefulness in detecting rumors.

This paper employs emotions for rumor detection by defining two emotion categories for social media posts: *post emotion* and *comment emotion*. We provide the features for the rumor detection framework by capturing and analyzing the user's emotions when publishing social media posts and the user's reactive emotions when the posts reach them. The numerous emotion-related expressions of social media posts and comments are analyzed, and a straightforward and persuasive approach is proposed, taking into account emotion and sentiment polarities that investigate their aspects in rumor detection. The following is a summary of the main contributions of this paper:

- Presenting techniques for extracting emotion-related features from social media posts and comments and integrating them with their psycholinguistic and syntactic features.
- Introducing an emotion-enhanced and psycholinguistic features-based approach for detecting rumors on social media that combines emotion-related aspects, psycholinguistic features, and a word embedding model.
- An empirical study using the popular `PHEME` dataset to assess the efficacy of emotion-enhanced and psycholinguistic features-based approaches for rumor detection.

The remaining parts of the paper are organized as follows. Section 2 presents a brief review of the related literature. Section 3 presents the architectural and functional details of the proposed approach. Section 4 presents the dataset, experimental results, and comparative analysis of the proposed approach. Finally, Section 5 concludes the paper and presents future research directions.

2 Related Works

There are two widely accepted theories on basic emotion models based on psychological science (Plutchik, 1982; Ekman, 1992). The study of emotions and their implications gets the interest of several fields, especially affective computing (Picard, 2000), where the emotions and sentiments of words and phrases are analyzed. Emotion analysis frequently resembles sentiment analysis and opinion mining, as well as the study of affective lexicons in psycholinguistics, which assesses the

[3] https://news.un.org/en/story/2020/06/1067422

[4] https://indianexpress.com/article/opinion/columns/india-show-way-combatting-fake-news-global-south-8646961/

connection between psychological processes and linguistic behaviors (Pang et al., 2008). In contrast to opinion mining and sentiment analysis, emotion analysis is not limited to analyzing the polarity but also associating text with a predefined set of psychological terms determined by dimensions such as valence, arousal, and dominance (Russell, 2003).

There are recent works on determining the veracity of rumors; however, very few were focused on detecting rumors based on emotion-related features. Vosoughi et al. (2018) explored emotions in rumors on social media and found that true rumors contain *joy, sadness, trust, and anticipation*, whereas false rumors trigger *fear, disgust*, and *surprise*. Ajao et al. (2019) determined the relationship between fake news and the sentiments of the social media posts. Giachanou et al. (2019) proposed an approach based on an `LSTM` neural network, which incorporates emotional signals to differentiate between credible and non-credible claims. Abulaish et al. (2019) incorporated sentimental aspects through a graph-based approach using POS tags to identify anxious and doubtful terms for rumor detection. Wasi and Abulaish (2020) proposed a logistic regression-based sentiment classification approach that uses prior domain knowledge extracted from a lexicon and unlabeled domain data.

Dong et al. (2022) proposed a method for fake news detection based on hypergraph attention networks, which employed two hypergraphs to model news contents and user comments to capture high-order relations between words in a news document and comments with the same sentimental polarity. Haque and Abulaish (2022) proposed a graph-based contextual and semantic learning approach using posts and comments to understand the underlying linguistic patterns that exploited the textual and latent information. Xu et al. (2022) studied the role of comments in rumor detection and proposed a method that extracted the features from the original post and associated comments. Choudhry et al. (2022) annotated fake news and rumor datasets with their emotion labels using transfer learning. They proposed a multitasking framework for fake news and rumor detection, predicting the text's emotion and legitimacy. In (Kumari et al., 2021) and (Gupta et al., 2022), the authors investigated the performance improvement of fake news detection with joint learning of novelty, emotion, and sentiments.

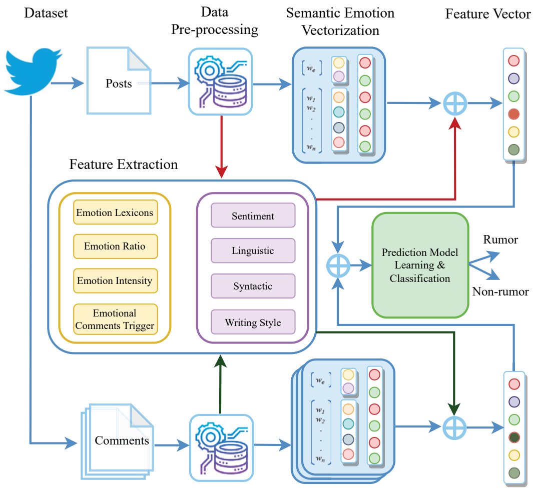

Figure 1: Work-flow of the proposed approach

Researchers have also explored the misinformation and emotions related to COVID-19 and studied their impacts on the pandemic (Caceres et al., 2022; Gupta et al., 2022; Sosea et al., 2022). Bhardwaj and Abulaish (2023) collected and annotated a large-scale, multi-labeled emotion and sentiment classification dataset that contains COVID-19-related textual data. They analyzed user sentiments and emotions expressed in tweets during the third wave of the omicron sub-variant pandemic. Furthermore, some researchers have explored different emotional patterns for analyzing various forms of false information, such as hoaxes, propaganda, clickbait, and satire (Rashkin et al., 2017; Ghanem et al., 2020; Mackey et al., 2021).

3 Proposed Approach

In this section, we explain the proposed rumor detection approach and its functional module. The work-flow of the proposed approach for rumor detection is represented in Figure 1.

3.1 Rumor Detection Formalization

Following the state-of-the-art approaches, we consider rumor detection as a binary classification problem. Given any i^{th} social media post p_i and its set of comments $C_{p_i} = \{c_1^{p_i}, c_2^{p_i}, ..., c_m^{p_i}\}$, where $p_i, C_{p_i} \in P$ of a social media dataset P, a rumor function f_x predict whether a post p_i is a rumor or non-rumor. Further, it is defined as f_x: (p, C_p) $\rightarrow \{0, 1\}$ such that,

$$f_x(p, C_p) = \begin{cases} 1, & p \text{ is a rumor} \\ 0, & \text{non-rumor.} \end{cases} \tag{1}$$

3.2 Data Preprocessing

In this step, data preprocessing tasks, such as tokenization, cleaning, and normalization, are conducted on the `Twitter` dataset. All the posts and comments are tokenized with white space. In normalization, we replace the Twitter-specific tokens, i.e., *hashtags, urls, retweets*, and *mentions*, with the tags *≺hashtags≻, ≺urls≻, ≺retweets≻*, and *≺mentions≻*, respectively. Emojis or emoticons are essential to convey emotion and are frequently used in informal text; therefore, we included them in our dataset by converting them into text using the `Python demoji` module.

3.3 Emotion Recognition

Trusted or genuine posts retain their natural way of writing without affecting the opinion of users. In contrast, rumorous posts are usually presented in a manner that manipulates emotions using the advantage of arousal and sensitive behaviors of users. Social media posts with emotional words such as astonishment, anger, and anxiety have a broader scope to get viral (Vosoughi et al., 2018). The nature of emotions in posts and comments varies based on social media users' intentions, content, and writing styles. Generally, emotions of posts and comments resonate in cases of outrage and anxiety, whereas, sometimes, posts having neutral feelings raise doubt and fear in their comments (Pröllochs et al., 2021; Vosoughi et al., 2018). Given any i^{th} post p_i and its comments C_{p_i}, the expression of emotions are categorized as post emotion and comments emotion, for which the emotion-related features are extracted to generate features vector $\mathcal{F}_e$.

Post emotion contains emotion-related information about the source post. It focuses on the users' emotions while posting any post on social media. Malign users intentionally feed intense emotions in social media posts to exploit the sensitive behavior of users. Purposefully, they depict themselves as genuine in publishing false but convincing news on controversial topics. We define the post emotion as $post_{emo} \in \mathcal{R}^{\mathcal{F}_e}$, where $\mathcal{R}^{\mathcal{F}_e}$ is the feature space of emotion-related feature $\mathcal{F}_e$.

Comments emotions are reactive emotions that capture emotion-related details from comments. Users commonly get deceived by the provocative emotion fed in the content of source posts and start commenting based on preconceived ideas and aroused feelings. Similarly, as a post, we define comments emotion as $comment_{emo} \in \mathcal{R}^{\mathcal{F}_e}$, where $\mathcal{R}^{\mathcal{F}_e}$ is the feature space of emotion-related feature $\mathcal{F}_e$.

3.4 Emotion-Related Feature Extraction

This section assesses several emotion resources to analyze and extract the range of emotion-related features from the text. Posts and comments are in textual form, represented as textual input $T = \{w_1, w_2, ..., w_n\}$. The following processes are considered to extract the emotion-related features for the textual input T.

3.4.1 Emotion Lexicon

Rumors evoke certain emotions in users; consequently, we consider eight types of emotions based on Plutchik's wheel (Plutchik, 2001) to investigate rumors, i.e., *anger, fear, disgust, trust, sadness, joy, anticipation*, and *surprise*. We use `NRC Word-Emotion Association Lexicon` (Mohammad and Turney, 2013) for extraction of the emotion categories that persist in the textual input. The emotion categories in the dictionary are represented as $E = \{e_1, e_2, ..., e_r\}$, where r is the total number of emotion categories. The frequency of emotion words belonging to each emotion category for a textual input T is computed using equation 2, where $f_{(T,e_i)}$ is the occurrence of the emotion words of T in a category e_i and $emo_{lex}(w_j, e_i)$ counts if a word $w_j, \forall w_j \in T$ is present in a emotion category e_i.

$$f_{(T,e_i)} = freq\{emo_{lex}(w_j, e_i)\}, \forall e_i \in E \quad (2)$$

Finally, the emotion lexicon feature vector $lex_{(T)}$ for textual input T is expressed by equation 3.

$$lex_{(T)} = [f_{(T,e_1)}, f_{(T,e_2)}, ..., f_{(T,e_r)}] \quad (3)$$

3.4.2 Emotion Ratio

We evaluate the emotional alignment of the posts and comments toward positive and negative emotion cues. Malign users make complicated stories and allude to negative words more frequently (Vosoughi et al., 2018). As determined in (Ajao et al., 2019), the fraction measure of emotion words at the sentence level is calculated to assess the influence toward positive or negative emotional states. For this purpose, unigrams of input text are matched with the existing lexicon dictionary (Mohammad and Turney, 2013) that counts the number of positive and negative emotions. The consideration of the emotion ratio feature $\mathcal{ER}$ is defined by equation 4, where $freq(e^+, T)$ and

$freq(e^-, T)$ counts the frequency of positive and negative emotions in textual input T, respectively.

$$\mathcal{ER}_{(T)} = \frac{freq(e^-, T)}{freq(e^+, T) + 1}, \forall e^-, e^+ \in E \quad (4)$$

3.4.3 Emotion Intensity

People used to express false narratives with higher intensities to emphasize and arouse intense emotions in the users (Pröllochs et al., 2021). Each word in posts and comments signifies diverse emotion-related signals based on their intensity levels. The `NRC Emotion Intensity Lexicon` (Mohammad, 2018) is used for calculating the emotion intensity score, which contains the same emotion categories considered in Section 3.4.1, with a real-valued score representing that particular word's intensity. We compute the intensity score corresponding to each emotion category by summing the intensity values of emotion words belonging to that category. The emotion intensity feature is defined by equation 5, where $emo_{int}(w_j, e_i)$ is the intensity of a word w_j in an emotion category e_i.

$$int_{(T,e_i)} = \sum_{w_j \in T}^{|T|} emo_{int}(w_j, e_i), \forall e_i \in E \quad (5)$$

The final emotion intensity feature vector $\mathcal{EI}_{(T)}$ for textual input T is obtained by equation 6.

$$\mathcal{EI}_{(T)} = \left[int_{(T,e_1)}, int_{(T,e_2)}, ..., int_{(T,e_r)}\right] \quad (6)$$

3.4.4 Triggered Emotional Comments

Social media users' sensitive and responsive behavior drives them to trigger their emotions after seeing provocating or controversial messages (Giachanou et al., 2018; Martel et al., 2020). They tend to reply or react in emotionalized ways. We incorporated the triggered emotions in comments using counts of their emotional expressions. Accordingly, the calculation is conducted to count the number of comments having positive and negative emotional words. We flag if positive e^+ and negative e^- emotions are present in the textual input T. In the end, for a post p_i the frequency of its flagged comments of C_{p_i} is counted using equation 7.

$$\mathcal{CT}_{(T)} = \left[freq(c, e^+), freq(c, e^-)\right], \forall e^+, e^- \in E \quad (7)$$

Eventually, the emotion-related feature vector $\mathcal{F}_e$ for an input text T is obtained by concatenating the feature vectors generated above, expressed by equation 8.

$$\mathcal{F}_e = lex_{(T)} \oplus \mathcal{ER}_{(T)} \oplus \mathcal{EI}_{(T)} \oplus \mathcal{CT}_{(T)} \quad (8)$$

3.5 Seed Feature Extraction

People cultivate rumors using exaggerated words such as superlatives, subjective, assertive, hedge, and manner adverbs (Wilson et al., 2005; Ott et al., 2011); subjective words are used to dramatize or sensationalize, whereas hedge words indicate vagueness, mystifying, and obscuring language (Islam et al., 2020). Also, mistrust toward online information evokes skepticism and doubt in users (Martel et al., 2020). Beyond emotions, these words need to be focused to enhance the framework's efficiency. We use lexical resources from existing works in communication theory and stylistic analysis of computational linguistics to extract these words from social media posts to characterize between rumor and non-rumor. The following features are comprised in that direction to extract the seed feature vector for the input text T.

Linguistic features: Analyze the psycholinguistic patterns in the text that are incorporated through the score of subjective, aggressive, hateful, and hedging words. Lexicons from (Islam et al., 2020) are used for hedge words to calculate the hedge score for an input text. We also used `LIWC` (Pennebaker et al., 2015), a linguistic dictionary widely used in social science studies composed of $6,400$ (approx.) words with different categories to extract psycholinguistic patterns. We consider categories, i.e., *total pronouns, common adverbs, common adjectives, comparisons, affective processes, negations, anger, sadness, positive emotion, negative emotion, anxiety*, and *swear* words. The subjectivity of an input text is obtained using the total number of words present in an input text by limiting the word count to 15 as an empirically calculated threshold. A transformer-based `Python` library for SocialNLP tasks described in (Pérez et al., 2021) is used for measuring hateful and aggressive scores. In the end, these measures are concatenated to construct a linguistic feature vector $\mathcal{L}_f(T)$.

Sentiment scores: Capture underlying sentimental signals of messages conveyed on social media. Rumorous posts manifest a higher negative sentiment than genuine posts (Vosoughi et al., 2018). To determine the sentiment expressed in an input text, we use the VADER (Hutto and Gilbert, 2014), a lexicon and rule-based sentiment analysis tool devised explicitly for sentiments expressed in microblogs with four dimensions, i.e., *positive, negative, neutral*, and *compound*. We select *positive, negative* and *neutral* sentiment polarity and construct the sentiment score $\mathcal{S}_s(T)$.

Syntactic features: Role are vital from the inception of the literature on rumor identification. The POS tagging is applied using spaCy to extract syntactic features based on the presence of *noun (NN), verb (VB), adverb (RB)*, and *adjective (JJ)* tags along with their fine-grained tags, i.e., *comparative adjective (JJR), superlative adjective (JJS), comparative adverb (RBR), superlative adverb (RBS)*, and *personal pronoun (PRP)*. The tokenized input text is assigned with respective POS tags using Penn Treebank for designing syntactic feature $\mathcal{S}_f(T)$.

Writing style: Include features of users' writing manners that contribute significantly to propagating false and ambiguous information, i.e., the use of punctuations $\{?, !, ...\}$, capital letters, and uppercase words. We consider the count of these measures when they appear in the input text T to construct the feature vector $\mathcal{W}_f(T)$. Finally, the seed feature vector $\mathcal{F}_s$ for the input text T is obtained by concatenating all extracted features, expressed by equation 9.

$$\mathcal{F}_s = \mathcal{L}_f(T) \oplus \mathcal{S}_s(T) \oplus \mathcal{S}_f(T) \oplus \mathcal{W}_f(T) \quad (9)$$

3.6 Semantic Emotion Vectorization

Text embedding, which provides learned word representation as low-dimension dense vectors in continuous embedding space, is applied to preserve the semantic relation of the posts and comments. We employed two embedding tasks, one for document representation and the second for underlying emotion representation in an input text. The emotion representation analyzes the intrinsic emotion in the input text T. It is attained by embedding the most frequent emotion word of the input text using the same embedding model. The textual input T of length n is represented as vector $[v_1, v_2, ..., v_n]$ where $v_i \in \mathcal{R}^d$; $\mathcal{R}^d$ is a d-dimensional word embedding vector for the i^{th} word in the textual input T. For this purpose, we use publicly available 100-dimensional word-level pre-trained GloVe embedding vectors trained over a Twitter dataset with 27 billion tokens[5]. Similarly, we construct the emotion embedding vector $[v_e]$ for the most frequent emotion word with the same embedding model. When extracting the most frequent emotion word, in the case where more than one occurrence appears, we consider the most intensified emotional word based on intensity score. Finally, the emotionalized embedding feature vector $\mathcal{F}_{sv}$ is obtained as the mean of both the embedding vectors defined by equation 10.

$$\mathcal{F}_{sv} = \frac{\sum_1^n [v_i] + [v_e]}{n+1} \forall v_i, v_e \in \mathcal{R}^d \quad (10)$$

Table 1: Statistics of the dataset

Events Name	Posts	Comments	Rumors	Non-Rumors	Total
Charlie Hebdo	2,079	36,189	458	1,621	38,268
Sydney Siege	1,221	22,775	522	699	23,996
Ferguson	1,143	23,032	284	859	24,175
Ottawa Shooting	890	11,394	470	420	12,284
Germanwings Crash	469	4,020	238	231	4,489
Putin Missing	238	597	126	112	835
Prince Toronto	233	669	229	4	902
Gurlitt	138	41	61	77	179
Ebola Essien	14	212	14	0	226
Total	6,425	98,929	2,402	4,023	105,354

3.7 Feature Vector Generation

In this step, the feature vectors for posts and comments are generated using the above-described extraction techniques. The posts and comments are treated concurrently since comments have different emotion-related features and linguistic information depending on the crowd's moods and interest in the ongoing events. Each statistical feature is calculated as a normalized sum of their score. All the above features are extracted for a post p_i and its comments $C_{p_i} = \{c_1^{p_i}, c_2^{p_i}, ..., c_m^{p_i}\}$. Feature vector associated with C_{p_i} is calculated as an average of all $\{C_{p_i}\}_1^m$. The final feature vector $\mathcal{FV}$ is obtained by concatenating the features of a post, i.e., $post_f = \mathcal{F}_e^p \oplus \mathcal{F}_s^p \oplus \mathcal{F}_{sv}^p$ and its comments, i.e., $comments_f = \mathcal{F}_e^c \oplus \mathcal{F}_s^c \oplus \mathcal{F}_{sv}^c$ using equation 11.

$$\mathcal{FV} = post_f \oplus comments_f \quad (11)$$

4 Experimental Setup and Results

In this section, we describe the experimental setup, results, and comparative analysis of the proposed approach with the baseline methods and state-of-the-art approaches. The effectiveness of the proposed approach is evaluated using the evaluation matrix described below.

4.1 Dataset

We conduct our experiment on a publicly available PHEME dataset used in (Kochkina et al., 2018). The dataset contains a collection of posts with their

[5] https://nlp.stanford.edu/projects/glove/

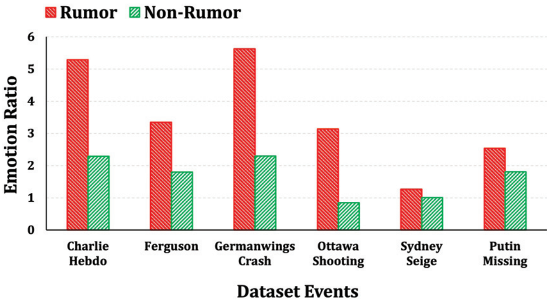

Figure 2: The distribution of negative and positive emotion words in the dataset represented by their ratio

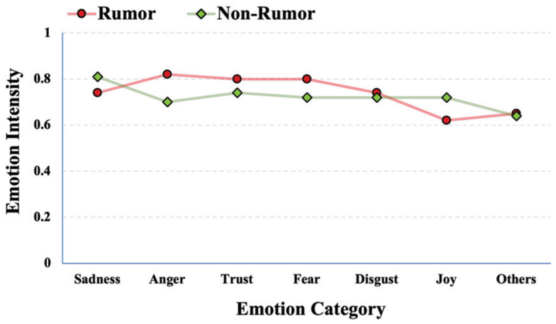

Figure 3: The distribution of emotion words based on their intensity scores in the dataset under the various emotion categories

comments related to nine events posted on `Twitter` during breaking news. The detailed dataset statistics are presented in Table 1.

4.2 Dataset Analysis

This section discusses the analysis of the dataset. The emotion expressed in the dataset varies since people have different expression styles according to their language, culture, and interest in ongoing topics. There are nine events in the dataset; the events containing at least 100 posts of rumor and non-rumor are considered for analysis. Figure 2 shows the presence of negative and positive emotions in the six events of the dataset using their emotion ratio. Figure 3 shows the distribution of emotion words based on their intensity scores under various emotion categories. The analysis also found that the proportion of negative emotions is higher in the rumor compared to non-rumor. Moreover, Emotion words present in the comments of rumor and non-rumor are analyzed, and it was found that the severity of anger, doubt, and fear is much greater in rumors. Extreme forms of words are deliberately chosen to unleash anger, such as deadly, brutal, and hatred.

4.3 Evaluation Metrics

This section discusses performance evaluation metrics for classification. The standard evaluation metrics- *Precision (P), Recall (R)*, and *F1-score* are defined in equations 12, 13, and 14, respectively. The evaluation metrics are defined using the concepts of *true-positive (tp), false-positive (fp)*, and *false-negative (fn)*. *True-positive* is the total number of rumors specified as rumors class correctly; *fp* is the total number of non-rumors specified as rumors; *fn* is the number of rumors identified as non-rumors. *Precision* assesses the correctness, whereas *Recall* evaluates the completeness of coverage of the classifier. Meanwhile, the *F1-score* uses harmonic means to provide a way to combine the contribution of both *Precision* and *Recall* evenly.

$$P = \frac{tp}{tp + fp} \tag{12}$$

$$R = \frac{tp}{tp + fn} \tag{13}$$

$$F1 - score = \frac{2 \times P \times R}{P + R} \tag{14}$$

4.4 Evaluation Results and Comparative Analysis

We performed experiments for the evaluation of the proposed approach using machine learning classification algorithms – *Support Vector Machine (SVM), Gradient Boosting (GB), Logistic Regression (LR)*, and *Multi-layer Perceptron (MLP)*, implemented using the `scikit-learn Python` library with the default parameter settings.

The proposed approach is compared with the following baseline methods and state-of-the-art approaches for evaluating the significance of extracted features:

Baseline 1: In this method, only posts are considered to evaluate the significance of emotion-related and seed features.

Baseline 2: In this method, posts and comments are considered to evaluate the seed features only.

Baseline 3: In this method, posts and comments are considered to evaluate the emotion-related features only.

Table 2: Comparative performance evaluation results of our proposed approach with the state-of-the-art approaches and baseline methods

Approach	SVM			LR			GB			MLP		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Baseline 1	58.48	50.74	54.33	59.10	55.21	57.08	64.61	66.86	65.71	69.09	65.35	67.16
Baseline 2	70.19	66.35	68.21	69.20	67.32	68.24	72.32	73.34	72.82	75.00	77.71	76.33
Baseline 3	87.42	90.60	88.98	82.49	78.40	80.39	86.34	89.56	87.92	84.38	91.50	87.79
Baseline 4	70.62	66.66	68.58	66.49	69.92	68.16	71.43	72.69	72.05	71.31	70.25	70.77
Abulaish et al. (2019)	41.30	45.56	43.32	40.80	41.93	41.11	56.26	55.40	55.82	64.62	60.10	62.28
Ajao et al. (2019)	87.18	86.00	86.58	84.63	84.88	84.75	85.28	86.82	86.04	88.83	87.21	88.01
Proposed Approach	88.56	92.42	90.44	90.35	86.52	88.39	87.47	93.26	90.27	91.54	94.96	93.21

Baseline 4: In this method, posts and comments are considered to evaluate the significance of emotion-related and seed features without considering the embedding model.

Abulaish et al. (2019): This approach incorporated sentimental aspects, such as anxiety and doubtful terms from the social media posts, and the embedding model for detecting rumors.

Ajao et al. (2019): This approach considered the relationship of rumors with the sentiments of the social media posts. It used the ratio of negative and positive emotions for the detection of sentiment-aware misinformation.

The experiments of the baselines and state-of-the-art approaches are performed to compare them with the proposed approach. The reasons for choosing these four classifiers are to follow the state-of-the-art approaches and exhibit the robustness of the extracted features. The performance of the classical machine learning algorithms is comparable to the *MLP* with one hidden layer, which signifies the robustness of the extracted features. The summarization of comparative results in terms of evaluation metrics, i.e., *Precision (P), Recall (R)*, and *F1-score (F1)*, are presented in Table 2. The proposed approach results show that the performance is significantly better for all four classification algorithms.

The best result of the proposed approach is obtained for the *MLP* classifier for all evaluation metrics. The outperformance of the proposed approach with one of the state-of-the-art methods (Ajao et al., 2019) ranges from $4.5 - 6.0\%$. The proposed approach remarkably outperformed (Abulaish et al., 2019) over all the classification algorithms. It shows that the features in the proposed approach are more versatile and broadened than (Abulaish et al., 2019), leading to a significant improvement in the performance. Moreover, baseline 3, considered only the emotion-related features for posts and comments, outperformed the state-of-the-art approach (Ajao et al., 2019) range from $2.1 - 2.7\%$ for the *SVM* and *GB* classification algorithms. It can also be observed from Table 2 that when the feature size is large, *MLP* performs better, while in baseline 4, where the feature size is small, *GB* performs better.

The experiments demonstrate the significant improvement of the detection task when combining emotion-related features with a embedding model. The performance analysis of the baselines also demonstrates that consideration of the emotions in comments improved the weight of the feature vectors and enhanced the performance significantly. The outperformance of the proposed approach and remarkable improvement against the baselines and state-of-the-art methods signifies that incorporating emotion-related features with psycholinguistic features enhances the rumor detection task.

5 Conclusion and Future Work

In this paper, we have presented an emotion-enhanced and psycholinguistic features-based approach for rumor detection that makes use of various emotion-related and linguistic features extracted from posts and comments. The embedding model is used to learn emotion-related semantic information from the posts and comments. The experimental results reveal a significant correlation between rumors and emotions, and demonstrate that emotion-related features substantially enhance rumor detection tasks, and are effective when combined with psycholinguistic features and an embedding model. It can also be concluded that utilizing the underlying features of comments improves the performance of the rumor detection task. The proposed approach can be extended to improve the extraction of emotion-related features that are not explicitly mentioned in the posts and comments. It can also be expanded to take into account the emotions that are embedded in image captions and multimedia.

References

Muhammad Abulaish, Nikita Kumari, Mohd Fazil, and Basanta Singh. 2019. A graph-theoretic embedding-based approach for rumor detection in twitter. In *Proceedings of the 18th IEEE/WIC/ACM International Conference on Web Intelligence, Thessaloniki, Greece*, pages 466–470. Association for Computing Machinery.

Oluwaseun Ajao, Deepayan Bhowmik, and Shahrzad Zargari. 2019. Sentiment aware fake news detection on online social networks. In *Proceedings of the 44th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK*, pages 2507–2511. IEEE.

Anjali Bhardwaj and Muhammad Abulaish. 2023. MINDS: A multi-label emotion and sentiment classification dataset related to covid-19. In *Proceedings of The First Workshop on Context-aware NLP in eHealth (WNLPe-Health 2022) co-located with The nineteenth International Conference on Natural Language Processing (ICON-2022), Delhi, India, December 15-18, 2022*.

David B Buller and Judee K Burgoon. 1996. Interpersonal deception theory. *Communication theory*, 6(3):203–242.

Maria Mercedes Ferreira Caceres, Juan Pablo Sosa, Jannel A Lawrence, Cristina Sestacovschi, Atiyah Tidd-Johnson, Muhammad Haseeb Ul Rasool, Vinay Kumar Gadamidi, Saleha Ozair, Krunal Pandav, Claudia Cuevas-Lou, et al. 2022. The impact of misinformation on the covid-19 pandemic. *AIMS public health*, 9(2):262.

Arjun Choudhry, Inder Khatri, Minni Jain, and Dinesh Kumar Vishwakarma. 2022. An emotion-aware multitask approach to fake news and rumor detection using transfer learning. *IEEE Transactions on Computational Social Systems*, pages 1–12.

Diwen Dong, Fuqiang Lin, Guowei Li, and Bo Liu. 2022. Sentiment-aware fake news detection on social media with hypergraph attention networks. In *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2174–2180. IEEE.

Paul Ekman. 1992. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200.

Bilal Ghanem, Paolo Rosso, and Francisco Rangel. 2020. An emotional analysis of false information in social media and news articles. *ACM Trans. Internet Technol.*, 20(2).

Anastasia Giachanou, Paolo Rosso, and Fabio Crestani. 2019. Leveraging emotional signals for credibility detection. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR'19, page 877–880, New York, NY, USA. Association for Computing Machinery.

Anastasia Giachanou, Paolo Rosso, Ida Mele, and Fabio Crestani. 2018. Emotional influence prediction of news posts. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 12.

Vipin Gupta, Rina Kumari, Nischal Ashok, Tirthankar Ghosal, and Asif Ekbal. 2022. MMM: An emotion and novelty-aware approach for multilingual multimodal misinformation detection. In *Findings of the Association for Computational Linguistics: AACL-IJCNLP 2022*, pages 464–477, Online only. Association for Computational Linguistics.

Asimul Haque and Muhammad Abulaish. 2022. A graph-based approach leveraging posts and reactions for detecting rumors on online social media. In *Proceedings of the 36th Pacific Asia Conference on Language, Information and Computation*, pages 533–544, Manila, Philippines. De La Salle University.

Clayton Hutto and Eric Gilbert. 2014. VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225.

Jumayel Islam, Lu Xiao, and Robert E. Mercer. 2020. A lexicon-based approach for detecting hedges in informal text. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 3109–3113, Marseille, France. European Language Resources Association.

Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. 2018. All-in-one: Multi-task learning for rumour verification. In *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Rina Kumari, Nischal Ashok, Tirthankar Ghosal, and Asif Ekbal. 2021. A multitask learning approach for fake news detection: Novelty, emotion, and sentiment lend a helping hand. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.

A Mackey, Susan Gauch, and Kevin Labille. 2021. Detecting fake news through emotion analysis. In *Proceedings of the 13th International Conference on Information, Process, and Knowledge Management*, pages 65–71.

Cameron Martel, Gordon Pennycook, and David G Rand. 2020. Reliance on emotion promotes belief in fake news. *Cognitive research: principles and implications*, 5:1–20.

Saif Mohammad. 2018. Word affect intensities. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Saif M. Mohammad and Peter D. Turney. 2013. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3):436–465.

Myle Ott, Yejin Choi, Claire Cardie, and Jeffrey T. Hancock. 2011. Finding deceptive opinion spam by any stretch of the imagination. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 309–319, Portland, Oregon, USA. Association for Computational Linguistics.

Bo Pang, Lillian Lee, et al. 2008. Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval*, 2(1–2):1–135.

James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of liwc2015. Technical report.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Rosalind W Picard. 2000. *Affective computing*. MIT press.

Robert Plutchik. 1982. A psychoevolutionary theory of emotions.

Robert Plutchik. 2001. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist*, 89(4):344–350.

Juan Manuel Pérez, Juan Carlos Giudici, and Franco Luque. 2021. pysentimiento: A python toolkit for sentiment analysis and socialnlp tasks.

Nicolas Pröllochs, Dominik Bär, and Stefan Feuerriegel. 2021. Emotions explain differences in the diffusion of true vs. false social media rumors. *Scientific Reports*, 11(1):22721.

Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2931–2937, Copenhagen, Denmark. Association for Computational Linguistics.

James A Russell. 2003. Core affect and the psychological construction of emotion. *Psychological review*, 110(1):145.

Tiberiu Sosea, Chau Pham, Alexander Tekle, Cornelia Caragea, and Junyi Jessy Li. 2022. Emotion analysis and detection during COVID-19. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 6938–6947, Marseille, France. European Language Resources Association.

Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science*, 359(6380):1146–1151.

Nesar Ahmad Wasi and Muhammad Abulaish. 2020. Document-level sentiment analysis through incorporating prior domain knowledge into logistic regression. In *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pages 969–974. IEEE.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 347–354, Vancouver, British Columbia, Canada. Association for Computational Linguistics.

Yang Xu, Jie Guo, Weidong Qiu, Zheng Huang, Enes Altuncu, and Shujun Li. 2022. " comments matter and the more the better!": Improving rumor detection with user comments. In *2022 IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, pages 383–390. IEEE.

Xinyi Zhou and Reza Zafarani. 2020. A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, 53(5):1–40.

Arkaitz Zubiaga, Ahmet Aker, Kalina Bontcheva, Maria Liakata, and Rob Procter. 2018. Detection and resolution of rumours in social media: A survey. *ACM Computing Surveys (CSUR)*, 51(2):1–36.