Final version of the accepted paper. Cite as: *"Asimul Haque and Muhammad Abulaish, Unmasking Rumors on Social Media by Combining Affective Patterns Flow with User Behaviors, In Proceedings of the 17th International Conference on COMmunication Systems and NETworks (COMSNETS), Bengaluru, India, January 6-10, 2025, IEEE, pp. 126-131. DOI: https://doi.org/10.1109/COMSNETS63942.2025.10885610 "*

Unmasking Rumors on Social Media by Combining Affective Patterns Flow with User Behaviors

Asimul Haque
Department of Computer Science
South Asian University, New Delhi, India
asimulhaque@students.sau.ac.in

Muhammad Abulaish, *SMIEEE*
Department of Computer Science
South Asian University, New Delhi, India
abulaish@ieee.org

***Abstract*—Social media, a platform for rapid information dissemination, has undergone a significant shift. Its users are no longer passive recipients but active creators of content. This paradigm shift has led to a surge in the quantity and diversity of information, reducing their credibility and paving the way for the production and spread of rumors. The prevalence of rumors poses a severe threat to the trustworthiness of the entire social media ecosystem. The adverse impact of rumors underscores the urgent need for their detection, making it a pressing concern and a crucial requirement. This paper introduces a rumor detection approach that combines affective and psycholinguistic information flow on social media, user attributes, propagation structures, and a word embedding model. It leverages external lexicon-based knowledge to extract the affective and psycholinguistic information flow on social media. It entails detecting rumors utilizing users' profiles and propagation structure information, and uses word-level `GloVe` embedding to preserve the semantic relations of the posts and their underlying affective information. The proposed approach is evaluated over the publicly available `Twitter`/𝕏 dataset and compared to eight baselines and state-of-the-art approaches, demonstrating substantially remarkable performance for rumor detection on social media.**

***Index Terms*—Rumor Detection, Social Network Analysis, Affective Information, User Behaviour, Machine Learning.**

I. Introduction

Social media platforms have drastically changed the landscape of news consumption and the pattern of information flow in the past decade. They have revolutionized the means of socializing among the general populace. The sheer volume of information encountered on social media stems from its user-friendly interface and the seamless ease of content modification and sharing [1]. These attractive features increased their popularity so much that they gained 4.89 billion users (about 60% of the world's population) by 2023, including `Facebook` and `Twitter`, hosting 3 billion and 66 million users,[1] respectively. Moreover, they continue to grow, with 150 million new users added in the last year, representing an annual growth rate of 3.2%. Despite their benefits and popularity, spreading rumors through them poses a consequential concern to civil society, impacting the explicit and implicit threats to social stability and public trust, especially during emergencies, hazards, or crises when people eagerly seek updates [1].

Rumors disseminate faster, farther, and broader than true news [2]. This faster dissemination has many consequential threats. Extensive studies have revealed that rumors adversely affect almost every realm of human activity, including financial markets, democracy, political affairs, and the health system [3]. Moreover, the World Health Organization termed misinformation as an *"infodemic,"*[2] and the World Economic Forum has ranked misinformation as the biggest global risk in their Global Risks Report 2021 and 2024[3]. The severe implications of misinformation have motivated many governments, regional bodies, and international organizations to adopt multiple combating strategies. For example, the United Nations urged social media users to *"pause take care before share."* In addition, numerous fact-checking sites like `Snopes`, `PolitiFact`, and `NewsGuard` are available to evaluate the accuracy of claims and news websites. At a time when the availability of advanced technologies, such as LLMs, is both enabling and complicating evolving hazards and threats of the rapid creation of false information and the ease of access to the web across all demographics aids. These combating approaches and human-dependable manual fact-checking resorts become myopic. Thus, there is a need for adaptability on multiple fronts, including social media platform regulations, media literacy programs, the responsible adoption of advanced emerging tools, and the devising of robust rumor detection models for addressing the issue on a larger scale.

On social media, crowd intelligence is a valuable source for debunking rumors, with social entities serving as crucial sources of semantic information [1], [4]. These entities foster close connections that reinforce shared narratives and influenced behaviors. If a post appears unreliable, users express opinions on its veracity and provide insightful clues, producing many retweets and replies within very few minutes, considerably higher than those generated by reliable ones [5]. The circulation of original posts enables users to provide semantic information and reveals emotion, sentiment, and skepticism in their reactions [2], [6]. Specifically, rumorous posts have a different distribution of affective information across the social content than authentic ones [2], [6], [7]. We focus on leveraging the underlying semantic and affective information flow on social media and integrating them with psycholinguistic, user, and propagation information. The motivation behind

[1]https://tinyurl.com/2vmfv9d4

[2]https://www.who.int/health-topics/infodemic

[3]https://www.weforum.org/press/2024/01/global-risks-report-2024-press-release/

our work is to capture knowledge from social media content and user interactions involving engagements and propagation structure, which can significantly enhance understanding of how information spreads and influences behavior.

In this paper, we present a rumor detection approach that combines affective and psycholinguistic information flow on social media, user behaviors, and propagation structure. It explores features from social media content, context, and temporal diffusion information. We delve into the social context, including information about users and the propagation structure containing the variation of responsive engagement and auxiliary information for facilitating rumor classification tasks [5], [8]. Understanding user behaviors is crucial in the context of rumor diffusion; for instance, users with a lower ratio of following-to-follower are more likely to spark rumor diffusion [9]. We consider user demographic features, profile information, and user-influenced factors, such as user-following networks. These features are essential in quantifying opinion leadership, influence, and social circles, thereby providing valuable insights into the social dynamics of rumor diffusion [10]. We also leverage propagation information to capture the degree of connection and interaction with temporal diffusion information. Moreover, the semantic relation of posts and reactions and their underlying affective information are also considered using the word-level `GloVe` embedding model. We augment this information in order to identify the combinations of feature sets that yield the best predictive capabilities for rumor classification. In summary, the main contributions of this paper can be briefly described as follows:

- Considering the usefulness of affective information flow on social media and the dynamics of social interaction, such as users' social relations and propagation structures for effectively detecting rumors on social media.
- Extracting the features for rumor detection from the social media content, users, and propagation structure. It leverages affective information flow and psycholinguistic features from social media posts and reactions using external lexicon-based knowledge and integrates them with user characteristics, diffusion information, and embedding model.
- Performing an empirical study on the publicly available `Pheme` dataset to assess the efficacy of the proposed approach. It also highlights the effectiveness of fusing affective and psycholinguistic patterns, user attributes, and propagation features for rumor detection.

II. Related Works

In recent years, the interdisciplinary field of rumor detection on social media has attracted the attention of many researchers. The textual and visual content of social media, with its linguistic, stylistic, and visual clues, provides a rich source for extracting discriminative characteristics for rumor detection [1], [4]. For instance, linguistic-based features can capture specific language styles, such as opinionated and inflammatory language in text, which are crucial for identifying rumors [4], [6]. Similarly, some researchers have explored emotional, sentimental, and skeptical clues for detecting rumors and fake news [7], [11], [12]. In [2], the authors observed that on social media, true rumors contain *joy, sadness, trust*, and *anticipation*, whereas false rumors trigger *fear, disgust*, and *surprise*. Moreover, the role of external lexicon-based knowledge has been instrumental in extracting text clues for rumors and other research areas [4], [6], [13].

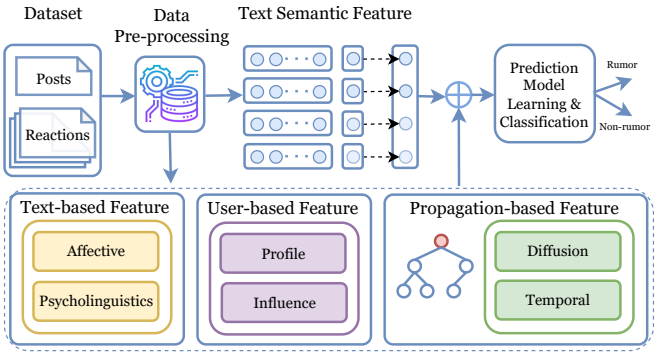

Fig. 1: Work-flow of the proposed approach

Researchers have focused on the role of social context and explored it alongside content-based approaches for rumor detection. These social context-based features mainly include user-based, propagation-based, and network-based information containing user profiles and social relationships [1]. Multiple studies have shown that social context against rumors encompasses distinct characteristics and factors, such as timing, speed, quantity, and nature of their responses [5], [14]. In [5], the authors created a kernel-based method that captures high-order patterns for differentiating rumors. In [14], the authors proposed a CNN-based model using text and propagation structure. The social context-based approaches used the crowd's wisdom as an auxiliary judgment and learned the contextual and semantic features of posts and reactions for rumor detection [6], [8], [12], [15]. In [8], the authors leveraged the homophily nature of social media users. Moreover, some studies have focused on correlating social media user profiles and rumors [9], [10]. Likewise, some have analyzed the characteristics and motivational factors of rumor and fake news spreaders on social media [16]. It is observed from the above discussions that many studies have focused on various aspects of content and social context for rumor detection. However, there is significant scope for exploring affective information inherent in posts and reactions for rumor detection, with user attributes and propagation structure.

III. Proposed Approach

In this section, we present the work-flow and functional details of the proposed rumor detection approach, as visualized in Figure 1. Let $T = \{t_1, t_2, ..., t_{|T|}\}$ be the set of source posts, each $t_i \in T$ is a text composed of a sequence of words $< w_1^{t_i}, w_2^{t_i}, ..., w_k^{t_i} >$. Each source post t_i is associated with its set of reactions $R_{t_i} = \{r_1^{t_i}, r_2^{t_i}, ..., r_{|R_{t_i}|}^{t_i}\}$, users U_i, and propagation network G_{t_i}. Similarly, as t_i, each $r_j^{t_i} \in R_{t_i}$ is

also composed of a word sequence of reply texts. For the social media dataset, wherein each t_i is associated with binary label {rumor, non-rumor}, a function f predicts whether a post t_i is a rumor or non-rumor. Mathematically, it is defined as f: $(t_i, R_{t_i}, U_{t_i}, G_{t_i}) \rightarrow \{0, 1\}$ such that,

$$f(t_i, R_{t_i}, U_{t_i}, G_{t_i}) = \begin{cases} 1, & t_i \text{ is a rumor} \\ 0, & \text{non-rumor} \end{cases} \quad (1)$$

A. Text-Based Feature Generation

The linguistic, psychological, and semantic cues underlying the textual content of social media are always powerful indicators for assessing their credibility. Social media posts that carry the flow of psychological and affective information, such as *uncertainty, astonishment, anger*, and *anxiety*, have a broader scope to go viral [2]. The nature of these flows in posts and reactions varies based on social media users' intentions, motivations, and writing styles. In this section, we explore the relevance of affective and psychological information flow in social media posts with their underlying semantic knowledge for detecting rumors. Given an input text, the following features are extracted to generate text-based features $\mathcal{T}_f$.

1) Affective and Psycholiguistic Information-Based Features: This section captures the affective information flow on social media with the assumption that rumorous posts have different distributions of these affective and psychological words. We use lexical resources from existing works in communication theory and psychological analysis of computational linguistics to extract this information from social media posts and reactions. The following scores are computed to comprise affective and psycholinguistic information-based features $\mathcal{A}_f$.

Emotion Scores: In today's information landscape, malign social media users deliberately inject intense emotions into their posts that exploit the sensitive behavior of others. This emotional coercion drives them to trigger their emotions in response to provocating or controversial messages. In our investigation of rumors, we consider eight types of emotions based on Plutchik's wheel [17], i.e., *anger, fear, disgust, trust, sadness, joy, anticipation*, and *surprise*. We use the `NRC Word-Emotion Association Lexicon` [18] to extract these emotion categories from the input text to measure their emotion scores E_s.

Emotion Ratio: Rumor stories are complicated and allude to a higher frequency of negative emotion words [2]. As determined in [11] and [6], we also calculate the fraction measure of emotion words that are inclined toward positive or negative emotional states using the `NRC Emotion Lexicon`. It is calculated as emotion ratio E_r using equation 2, where $freq(e^+)$ and $freq(e^-)$ count the frequency of positive and negative emotion words, respectively, in the input text.

$$E_r = \frac{freq(e^-)}{freq(e^+) + 1} \quad (2)$$

Sentiment Scores: Rumorous posts also manifest a higher negative sentiment, such as *fear, anxiety*, and *disgust* in their reactions [2]. We extract *positive* and *negative* sentiments from the textual input using the `NRC Emotion Lexicon` [18] and `SenticNet` [19]. `SenticNet` is based on commonsense knowledge, consisting of word sentiment scores.

Morality Scores: Beyond emotion and sentiment, morality is one of the significant factors in the virality of online content. We measure the morality score to gain a deeper understanding of the affective information in the input text, considering the word categories from the `Moral Foundations Dictionary` [20] as determined in [7]. These categories are *care, harm, fairness, unfairness (cheating), loyalty, betrayal, authority, subversion, sanctity*, and *degradation*.

Stress Scores: The manipulative, inflammatory, and sensitive terms of social media have an effect on users that increases their stress levels. These stress elements, such as *worries, frustration*, and *irritation*, are important to capture from the language used on social media. We compute the stress score using the lexical dictionary created by Wang et al. [21].

Uncertainty Scores: Uncertain situations often lead to increased rumor dissemination. It affects the information-seeking behavior of users during crises. If credible sources are unable to provide information, individuals resort to making predictions based on their judgment to fill the information gap. To measure this uncertainty, we utilize the *discrepancy (discrep), tentativeness (tentat)*, and *certainty (certain)* categories from `LIWC` [22], a linguistic dictionary widely used in social science studies composed of $\sim 6,400$ words.

Readability Scores: The complexity of text in NLP depends on how easily it can be read and understood. We consider this feature by measuring the `Simple Measure of Gobbledygook Index (SMOG)` readability score using the `Textstat`[4] `Python` library. Higher scores of readability indicate that the text is easier to read.

2) Text Semantic Feature: In this section, text embedding, which provides learned word representation as low-dimension dense vectors in continuous embedding space, is applied for semantic feature generation. The input text is a sequence of n-words represented as vector $[w'_1, w'_2, ..., w'_n]$ where $w'_i \in \mathcal{R}^d$; $\mathcal{R}^d$ is a d-dimensional word embedding vector for the i^{th} word. We use publicly available 100-dimensional word-level pre-trained `GloVe`[5] embedding vectors. As described in [6], we construct the emotion embedding vector $[w'_e]$ for the most frequent emotion word with the same embedding model. Thus, the emotionalized embedding feature vector $\mathcal{S}_f$ is obtained by the mean pooling of both the embedding vectors.

Finally, the text-based feature vector $\mathcal{T}_f$ for each input text is obtained using equation 3 by concatenating the above-generated features.

$$\mathcal{T}_f = \mathcal{A}_f \oplus \mathcal{S}_f \quad (3)$$

B. User-Based Feature Generation

In this section, we leverage the intricate information of social media users to understand their characteristics, which are valuable in detecting their rumorous posts. A user's decision to share a social media post depends on the influence

[4]https://pypi.org/project/textstat/

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

of their friends along with their specific characteristics, such as demographics, profile properties, influence, activities, etc. Exploring all these characteristics aligns with the diffusion of innovations theory [23], which explains how breaking news diffuses on a social network. We explore the following user information to generate user-based features $\mathcal{U}_f$.

1) Profile-Based Features: This section considers user profile information, which is a strong indicator for rumor detection. It is assessed from different characteristics, i.e., implicit (e.g., gender and age) and explicit (e.g., username and geolocations), for extracting profile-based features $\mathcal{P}_f$.

Demographic Features: User demographics, such as age, gender, and political orientation, are not explicitly available on social media platforms. Therefore, we utilized m3inference [24], a deep-learning-based system trained on `Twitter` data, to infer user demographic characteristics. This tool predicts the gender as male or female and the user's age into four categories ($\leq$ 20, 20-30, 30-40, and $\geq$ 40). It also predicts whether the given account is handled by an organization. In this paper, we only use age and gender as discriminative features.

Explicit Features: We consider the following user profile information as explicit features: (1) the length of the user name, i.e., the screen name that identifies a user; (2) the length of the description, i.e., the user's self-description; (3) Is geo-enabled, i.e., the geographical location of the user is enabled or not; (4) is verified, i.e., indicating whether the platform has verified the account; (5) the friends count; and (6) the favor count, i.e., the number of posts a user has liked.

2) Influenced-Based Features: This section focuses on users' attributes that describe their influences on social media, i.e., user activities and social relations. The following measures are considered to generate influenced-based features $\mathcal{I}_f$.

Popularity Ratio: This metric measures the degree of interconnection of users within social media circles. It is quantified in terms of popularity ratio (PR) using equation 4. The higher the PR value, the more popular a user is on social media.

$$PR = \frac{Number\ of\ followers + 1}{Number\ of\ followings + 1} \tag{4}$$

Profile Lifetime Index: It quantifies the span of user accounts on social media, calculated as the $Current\ Date - User\ Signup\ Date$.

Status Count: It counts the number of posts a user issues in a lifetime.

Finally, the user-based feature vector $\mathcal{U}_f$ of the users involved in the social media post is obtained by concatenating the above-generated features as defined in equation 5.

$$\mathcal{U}_f = \mathcal{P}_f \oplus \mathcal{I}_f \tag{5}$$

C. Propagation-Based Feature Extraction

In this section, we analyze propagation structures of social media posts for capturing network-based features and degree of connections, which also have a critical role in detecting falsehood. We define a propagation graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_1, v_2, ..., v_n\}$ represent a set of nodes, and $\mathcal{E} = \{e_{tr} | t, r = 1, 2, ..., n\}$ refers to set of edges from post v_t to response v_c. Suppose $\alpha = \{a_1, a_2, ..., a_n\}$ be a set of timestamps for the corresponding posts publication. More precisely, v_1 is the source node, $v_i (2 \leq i \leq n)$ represents the i^{th} response node, and α is the timestamps, for which the following propagation-based features $\mathcal{G}_f$ is generated.

Temporal Features: The sequential response patterns and users' engagement time contain indicative clues that can assess social media posts' credibility. The drift of posts through social networks also establishes linear interactions. We extract the sequential user engagement information on posts within a specific time period. Also, we integrate the chronological order of retweets time with their user-based features.

Propagation Cascades: The propagation cascades are defined as a group of relationships of continuous retweets between users. We consider the number of cascades as a feature by calculating the depth of the propagation graph.

Interaction Features: The dynamics of social interactions and users' responses to posts are useful and crucial in detecting rumors. We calculate the statistical features based on the users' interactions. These features are (1) the number of retweets / total responses, (2) the number of likes / total responses, (3) the number of support responses / total responses, and (4) the number of refute responses / total responses.

D. Fusion of Features

This section fuses the above-described features into a unique vector representation. These features, each playing a distinct role in unmasking rumors, are based on the diverse aspects of social media information. The final feature vector representation $\mathcal{F}$ is obtained by concatenating text-based, user-based, and propagation-based feature vectors using equation 6.

$$\mathcal{F} = \mathcal{T}_f \oplus \mathcal{U}_f \oplus \mathcal{G}_f \tag{6}$$

IV. Experimental Setup and Results

This section presents a detailed description of the dataset, experimental setup, results, and comparative analysis of the proposed approach. We analyze the experimental results over the `Pheme` dataset and compare the performance of the proposed approach with the eight baselines and state-of-the-art approaches. Finally, an ablation study is presented to provide a deeper analysis of the proposed approach.

A. Dataset

The experiment is conducted over the publicly available `Pheme` dataset [15]. This dataset is a collection of source posts, their reactions, and metadata related to five breaking

TABLE I: Statistics of the Pheme dataset

Events Name	Rumors	Non-Rumors	Posts	Reactions	Total
Charlie Hebdo	458	1,621	2,079	36,189	38,268
Sydney Siege	522	699	1,221	22,775	23,996
Ferguson	284	859	1,143	23,032	24,175
Ottawa Shooting	470	420	890	11,394	12,284
Germanwings Crash	238	231	469	4,020	4,489
Total	1,972	3,830	5,802	97,410	103,212

TABLE II: Comparative performance evaluation results of the proposed approach with baselines and state-of-the-art approaches

Approach	GB			SVM			LR			MLP		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Baseline 1	89.52	88.36	88.93	89.32	87.68	88.49	87.60	85.94	86.76	92.08	90.43	91.24
Baseline 2	68.48	66.75	67.59	69.15	64.88	66.94	62.33	63.47	62.89	71.26	70.35	70.80
Baseline 3	60.33	61.24	60.78	60.00	59.21	59.60	56.54	59.40	57.93	63.83	61.26	62.52
PTK	74.09	67.59	70.69	73.94	66.38	69.95	69.00	65.05	66.96	76.25	70.06	73.02
Ajao et al. [11]	85.28	86.82	86.04	87.18	86.00	86.58	84.63	84.88	84.75	88.83	87.21	88.01
Shu et al. [9]	92.16	88.78	90.43	91.33	89.26	90.28	90.71	87.00	88.81	92.52	91.14	91.84
Lathiya et al. [8]	65.23	61.35	63.23	60.87	52.65	56.46	48.97	51.27	50.09	69.43	66.17	67.76
Haque & Abulaish [6]	87.47	93.26	90.27	88.56	92.42	90.44	90.35	86.52	88.39	92.96	92.44	92.69
Proposed Approach	95.76	93.00	94.36	95.20	92.88	94.02	93.65	93.47	93.55	97.62	94.81	96.19

news events posted on Twitter/𝕏. The source posts are annotated as rumors or non-rumors. The detailed dataset statistics are presented in Table I.

B. Experimental Setup

All textual contents of the dataset undergo preprocessing phase, including noise filtering, cleaning, contraction, tokenization, and normalization. We employ pycontractions,[6] a Python library for expanding contractions, improving text clarity, and maintaining the original meaning of the text. Non-alphabetic words, including numbers and symbols, and one-lettered words are removed from the text. We keep the period in a sentence to extract readability features. Stop-words and capital letters are essential to understanding the context of posts. Likewise, emojis are also frequently used to convey emotion in informal texts; therefore, we include them in our work. We convert emojis into text using the Python demoji module.

The evaluation is underpinned by an efficient hold-out validation strategy. The datasets are split into training and testing sets in a proportion of $80 : 20$. We used Python programming language to implement the proposed approach, baselines, and state-of-the-art methods for empirical study. We choose four classification algorithms for evaluation, each representing a different learning paradigm. These include support vector machine (SVM) as a kernel machine, gradient boosting (GB) as an ensemble technique, logistic regression (LR) as a non-sequential linear model, and multi-layer perceptron (MLP) as a neural network-based model. These classifiers are implemented at their default hyper-parameter settings in Scikit-learn,[7] a widely used machine learning library in Python. The hyper-parameter settings of state-of-the-art methods are followed as recommended in their respective papers. We have used three standard evaluation metrics-*precision, recall*, and *f1-score* to assess the performance of the proposed approach. We have employed publicly available Glove embedding model, which is a 100-dimensional word-level pre-trained embedding over the Twitter dataset with 27 billion tokens. Those words that do not exist in the pre-trained word vectors are ignored.

[6]https://pypi.org/project/pycontractions/

[7]https://scikit-learn.org/stable/

C. Evaluation Results and Comparative Analysis

This section presents the comparative performance evaluation results of the proposed approach with the following baselines and state-of-the-art approaches for evaluating the significance of extracted features for rumor detection.

- Baseline 1: In this method, only text-based features are considered for detecting rumors.
- Baseline 2: In this method, only user-based features are considered for detecting rumors.
- Baseline 3:In this method, only propagation-based features are considered for detecting rumors.
- PTK [5]: This approach proposed a kernel-based propagation tree for rumor detection. It used high-order patterns of rumors by evaluating the similarities between their propagation tree structures. It captured the salient substructures in the propagation trees, considering different propagation paths from source posts to the roots of subtrees, representing the context of transmission.
- Ajao et al. [11]: This approach proposed sentiment-aware misinformation detection, using emotion ratio to represent the relation of rumors with the sentiments of the social media posts.
- Shu et al. [9]: This study presented the correlation between social media user profiles and rumors, utilizing various user profile features for rumor detection.
- Lathiya et al. [8]: This approach considered the content and user features of social media posts, as well as the users' homophilic connections for rumor detection.
- Haque & Abulaish [6]: This approach presented emotion-related aspects, psycholinguistic features, and a word embedding model for rumor detection on social media.

A summary of comparative results in terms of *precision* (P), *recall* (R), and *f1-score* (F1) is presented in Table II. It can be observed from the results that the proposed approach achieves the best performance compared to all baselines and state-of-the-art approaches, validating the effectiveness of the extracted features for rumor detection. The proposed approach significantly outperforms all the classifiers, with the MLP yielding the best result. The proposed approach achieved comparatively significant performance in terms of *f1-score* by the margin of $3 - 5\%$, with two good performing state-of-the-art

TABLE III: Ablation study results in terms of f1-score

Approach	GB	SVM	LR	MLP
w/o text-based features	71.23	70.86	69.92	72.50
w/o user-based features	90.24	90.36	88.74	92.80
w/o propagation-based features	91.48	92.18	90.25	93.46

approaches [6], [9]. The experimental results also demonstrate that the features extracted in the proposed approach are more versatile and diverse, enhancing the rumor detection task.

D. Ablation Study

In this section, we conduct the ablation study by excluding each component from the model to check its associated effectiveness and systematically validate its relative importance in the proposed approach.

- *w/o text-based features:* This method considers only user-based and propagation-based features for the model.
- *w/o user-based features:* This method considers only text-based and propagation-based features for the model.
- *w/o propagation-based features:* This method considers only text-based and user-based features for the model.

The experimental results of the ablation study are presented in Table III. The results demonstrate that removing text-based features has a pronounced impact on performance, with the *f1-score* dropping by 23.69% in the best case when compared with the proposed approach. However, this performance is significantly enhanced when text-based features are combined with user-based features, as they effectively capture the cues and dynamics of social interactions through user responses.

V. Conclusion and Future Work

In this paper, we have presented a rumor detection approach that utilizes affective and psycholinguistic aspects of social media text, users' behaviors, and propagation information. It considers text-based features from affective and psycholinguistic information flow in posts and reactions. The user-based features utilize explicit and implicit profile information and user-influenced aspects. The propagation-based features leverage the degree of connections, interactions, and temporal information. The embedding model is used to learn the semantic information underlying posts and reactions. The experimental results have demonstrated the significance of these features in rumor detection. It also has shown the effectiveness of text-based features in the detection task and enhanced performance when combined with user and propagation features. The proposed approach can be extended by incorporating more nuanced features from users' interactions and social circles. It can also be expanded to detect fine-grained rumors, such as true rumors, false rumors, hoaxes, and propaganda.

References

[1] X. Zhou and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," *ACM Computing Surveys (CSUR)*, vol. 53, no. 5, pp. 1–40, 2020.

[2] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.

[3] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *Journal of economic perspectives*, vol. 31, pp. 211–236, 2017.

[4] Z. Liu, T. Zhang, K. Yang, P. Thompson, Z. Yu, and S. Ananiadou, "Emotion detection for misinformation: A review," *Inf. Fusion*, vol. 107, p. 102300, 2024.

[5] J. Ma, W. Gao, and K. Wong, "Detect rumors in microblog posts using propagation structure via kernel learning," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, (Vancouver, Canada), pp. 708–717, ACL, 2017.

[6] A. Haque and M. 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*, (Bali, Indonesia), pp. 28–37, ACL, 2023.

[7] B. Ghanem, S. P. Ponzetto, P. Rosso, and F. Rangel, "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*, pp. 679–689, ACL, 2021.

[8] S. Lathiya, J. S. Dhobi, A. Zubiaga, M. Liakata, and R. Procter, "Birds of a feather check together: Leveraging homophily for sequential rumour detection," *Online Soc. Networks Media*, vol. 19, p. 100097, 2020.

[9] K. Shu, X. Zhou, S. Wang, R. Zafarani, and H. Liu, "The role of user profiles for fake news detection," in *ASONAM '19: International Conference on Advances in Social Networks Analysis and Mining*, (Vancouver, British Columbia, Canada), pp. 436–439, ACM, 2019.

[10] X. Chen, F. Zhou, G. Trajcevski, and M. M. Bonsangue, "Multi-view learning with distinguishable feature fusion for rumor detection," *Knowl. Based Syst.*, vol. 240, p. 108085, 2022.

[11] O. Ajao, D. Bhowmik, and S. Zargari, "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)*, pp. 2507–2511, IEEE, 2019.

[12] A. Haque and M. Abulaish, "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, PACLIC*, (Manila, Philippines), pp. 533–544, ACL, 2022.

[13] B. Anjali, W. N. Ahmad, and A. Muhammad, "Unlocking emotions in text: A fusion of word embeddings and lexical knowledge for emotion classification," in *Proceedings of the 20th International Conference on Natural Language Processing, ICON, (Goa, India)*, pp. 766–772, 2023.

[14] K. Tu, C. Chen, C. Hou, J. Yuan, J. Li, and X. Yuan, "Rumor2vec: A rumor detection framework with joint text and propagation structure representation learning," *Inf. Science*, vol. 560, pp. 137–151, 2021.

[15] A. Zubiaga, M. Liakata, and R. Procter, "Exploiting context for rumour detection in social media," in *Proceedings of International Conference on Social Informatics*, (Oxford, UK), pp. 109–123, Springer, 2017.

[16] M. Karami, T. H. Nazer, and H. Liu, "Profiling fake news spreaders on social media through psychological and motivational factors," in *HT'21: 32nd ACM Conference on Hypertext and Social Media, Virtual Event*, (Ireland), pp. 225–230, ACM, 2021.

[17] R. Plutchik, "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*, vol. 89, no. 4, pp. 344–350, 2001.

[18] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word-emotion association lexicon," *Computational Intelligence*, vol. 29, no. 3, pp. 436–465, 2013.

[19] E. Cambria, S. Poria, R. Bajpai, and B. W. Schuller, "Senticnet 4: A semantic resource for sentiment analysis based on conceptual primitives," in *Proceedings of the 26th International Conference on Computational Linguistics*, (Osaka, Japan), pp. 2666–2677, ACL, 2016.

[20] J. Graham, J. Haidt, and B. A. Nosek, "Liberals and conservatives rely on different sets of moral foundations.," *Journal of personality and social psychology*, vol. 96, no. 5, p. 1029, 2009.

[21] W. Wang, I. Hernandez, D. A. Newman, J. He, and J. Bian, "Twitter analysis: Studying us weekly trends in work stress and emotion," *Applied Psychology*, vol. 65, no. 2, pp. 355–378, 2016.

[22] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The development and psychometric properties of liwc2015," tech. rep., 2015.

[23] L. Ma, C. S. Lee, and D. H. Goh, "Understanding news sharing in social media: An explanation from the diffusion of innovations theory," *Online Inf. Rev.*, vol. 38, no. 5, pp. 598–615, 2014.

[24] Z. Wang, S. A. Hale, D. I. Adelani, P. A. Grabowicz, T. Hartmann, F. Flöck, and D. Jurgens, "Demographic inference and representative population estimates from multilingual social media data," in *The World Wide Web Conference*, (San Francisco, CA, USA), pp. 2056–2067, ACM, 2019.