PL-NCC: A Novel Approach for Fake News Detection through Data Augmentation

Keshopan Arunthavachelvan
Toronto Metropolitan University
Toronto, Canada
karunthavachelvan@torontomu.ca

Shaina Raza
Vector Institute of Artificial Intelligence
Toronto, Canada
shaina.raza@vectorinstitute.ai

Chen Ding
Toronto Metropolitan University
Toronto, Canada
cding@torontomu.ca

***Abstract*—Fake news has become rampant with the prominence of social media, jeopardizing information integrity and the reliability of news. It becomes imperative for media platforms to develop robust and efficient strategies to mitigate the spread of fake news. We make two contributions to fake news research in this work. Firstly, we propose the Psycho-Linguistic News Content and Comments (PL-NCC) dataset, a consolidated dataset of two leading fake news datasets, NELA-GT and Fakeddit, which leverages linguistic and psychological characteristics from the news article and user comments to improve the classification accuracy of benchmark models. Secondly, we propose the News Content and Comments (NCC) classification model to leverage the psychological features extracted from our PL-NCC dataset, which introduces a feed-forward layer to a deep learning model, enhancing the efficacy of the extracted psychological features to better identify fake news. Our approach achieves a classification accuracy of over ninety percent, exceeding baseline results.**

***Index Terms*—fake news detection, linguistic features, psychological features, user engagement, user comments, fake news dataset**

I. Introduction

In an era where social media has popularized the digital dissemination of information, the spread of false information presents a substantial threat to maintaining the trust and integrity of information in modern-day society [1], [2]. As fake news continues to become increasingly prominent, it becomes crucial for social media platforms to develop preventative measures to effectively recognize and mitigate fake news on their platforms.

Earlier fake news detection models [3], [4] primarily relied on the news article itself for new classification by analyzing the news's content, writing style, and source credibility. However, researchers have started including users' engagements and propagation data (user profiles for news distribution and tracking) as part of the classification task. Previous studies [5], [6] show that user engagement and news propagation provide valuable information for the classification of fake news. However, other works have identified through analysis that identifying the propagation graph or news cascade can be time-consuming [7], and this type of information may not be available if we want to detect fake news in the early stages [8]. Developing a classification model to rely mostly on news content would be a better option for early fake news detection.

Additionally, recent work in fake news detection explores the use of Cognitive Natural Language Processing (NLP) [9]–[11], which studies elements of cognitive science, linguistics, psychology and artificial intelligence to better understand human language. Cognitive NLP provides valuable details about the writing style of an article, including grammar and language analysis, as well as emotions portrayed in the writing. This data greatly enhances the ability for classification models to determine the truthfulness of news.

Several studies [12] have uncovered the role of psychological characteristics in differentiating between authentic and fake news. Other related studies [12], [13] have highlighted the potential for extracting various sentiment-based characteristics from news data, including the article's text and user comments. These studies consider the factors, such as emotions, swear words, and social behaviour for detecting the veracity of the news.

An in-depth analysis of these feature groups indicates that social behavioural traits, such as politeness, interpersonal conflict, moralization, pro-social behaviour and communication have elevated effects in fake and real news. Studies [12], [13] have shown that fake news often exhibits emotional bias [14] and profanity, while truthful articles are more neutral. Positive and negative emotions, as well as tones of anxiety, sadness, and anger, are common traits found in fake news and are beneficial when incorporated into fake news classification [15].

The goal of this work is to enhance the accuracy of existing fake news classification models by leveraging the linguistic and psychological features extracted from the textual components of news articles and associated user engagement. To achieve this, we first develop a consolidated Psycho-Linguistic News Content and Comments (PL-NCC) dataset built on state-of-the-art NELA [16] and Fakeddit [17] fake news datasets and expand on them by including various linguistic and psychological features unique to fake and real news. In addition, when the propagation data is available, we develop a simple way to include the propagation data in our dataset by including the related user comments for each news article, making it a desirable approach for early detection models. Secondly, we come up with an early fake news detection model using these linguistic and psychological features of both news content and user comments. The model functions well when only relying on the news content, and it also has the flexibility to take into consideration the user comments as they become available.

To build the PL-NCC dataset, we take the news content of

NELA combined with the user engagement from Fakeddit to create a more diverse and larger dataset for the classification task. Additionally, we expand on the consolidated dataset by introducing additional pre-processed linguistic and psychological features from the article's text and user comments to be ready to use for classification models. Linguistic features could reveal the detailed information about the writing style and formality of the news. Real news articles are typically written in a more formal manner, while fake news often exhibits informal language and numerous abbreviations [18]. Disinformation-related traits and clickbait terms from the headline could also be used to identify fake news [19]. Additionally, incorporating psychological features provides insights into the emotional aspects and biases present in the writing, which are common traits in fake news. We provide the embedding values of these extracted features in the PL-NCC dataset.

After preparing the dataset, we analyse the patterns of these linguistic and psychological features existing in fake and real news, and find out that leveraging these characteristics can enhance the fake news classification. We further analyze the patterns in the user comments and observe that these features sometimes exhibit different patterns in fake and real news. By including the user comments, we could further enhance the detection accuracy. Based on these findings, we propose a content-based early fake news detection model - News Content and Comments (NCC) based model, which combines deep neural networks with the lexical, semantic, syntactical, and psychological features extracted in the proposed dataset. To enhance the efficacy of the proposed psychological features in our PL-NCC dataset, we build upon a standard MLP classification model by introducing a feed-forward linear layer, allowing the classifier to better train on unique characteristics of fake news.

Finally, we conduct an extensive set of experiments to compare the performance of the proposed NCC classification model against the latest state-of-the-art fake news detection models. Using fake news datasets from NELA-GT and Fakeddit as benchmarks, we evaluate the performance of each model with various feature sets and embedding systems. Our model outperformed existing classifiers with fake news detection.

II. Related Work

Fake news detection is a specific type of text classification [45], where the goal is to categorize news articles as either true or fake. "Fake news" is identified as misinformation that is presented as genuine, aiming to deceive or mislead the audience [14]. Recent research has focused on leveraging artificial intelligence (AI) and machine learning (ML) algorithms to identify and mitigate the spread of false information [44].

Fake news manifests in several formats, including but not limited to clickbait (misleading titles), disinformation (deliberately misleading information), misinformation (incorrect information without intent to deceive), hoaxes, parodies, satires, rumors, and deceptive news [46]. The spread of fake news is further compounded by the psychological and social dynamics of misinformation sharing [47].

We categorize fake news datasets into three types: text-based, image-based, and propagation-based datasets.

A. Text-Based Datasets

Most common fake news detection models [14], [20] focus on a text-based approach for classification. To train such models, several text-based datasets have emerged to streamline the training process. Additionally, these datasets contain the largest sample size compared to other forms of datasets. Datasets such as NELA-GT [16], Credbank [21], FakeNewsNet [22], and Liar [23] are considered benchmark datasets for text-based classification as they contain a large sample of articles collected from several sources of news media. Datasets such as FakeNewsNet provide quality labelling methods [24], through manual annotations on the data curated from the fact-checking platforms such as PolitiFact and GossipCop. Many datasets such as NELA and Fakeddit perform their labelling at the news source level instead of at the article level, which may be lesser accurate compared to manually fact-checked datasets such as FakeNewsNet. However, manually fact-checking news articles is a tedious task, with much smaller sample size as in FakeNewsNet.

State-of-the-art fake news detection models [3], [4], [25], [26] focus on extracting different linguistic features using NLP tools from the article's text to perform the classification task. Previous research, such as Verma et al. [27], also incorporated pre-processed feature sets in their WELfake dataset to enhance classification accuracy and runtime. The WELfake dataset combines news articles from existing datasets to create a diverse and unbiased dataset for fake news classification. Their approach involves analyzing linguistic patterns like sentence length, readability, and subjectivity to assign a binary label to each article. Additionally, they provide a comprehensive list of linguistic features, such as readability and writing patterns, as well as sentiment analysis, to improve the effectiveness of existing fake news classification models.

B. Image-Based Datasets

Some studies in fake news detection [26], [28] have experimented with multimodal classification and have found promising results in the models' performance. Research has shown that images in news articles provide important details about the veracity of news articles and can be used to generate misinformation through techniques such as clickbait. Thus, research in multimodal datasets has become more prevalent in recent years. Datasets such as the Image Manipulation dataset [29] and PS-Battles dataset [30] provide an expansive sample of manually fetched articles which contain manipulative images used for fake news. Classification models can leverage these datasets to train models on image-based classification for effective fake news mitigation. Some datasets that are multimodal are Fakeddit [17], MedEval [1] and FakeNewsNet [22].

[1] https://multimediaeval.github.io/editions/2022/tasks/fakenews/

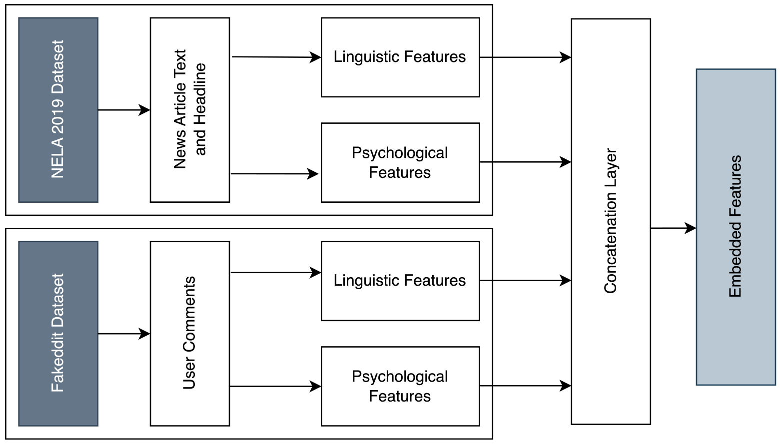


Fig. 1. Overview of proposed dataset

C. Propagation-Based Datasets

More recent work [5], [7], [17] in fake news has explored methods of fake news detection outside of the traditional text and image-based classification methods. While the previous datasets collect data from news sources, datasets such as Truth Seeker [31], PHEME [32], and RumourEval [33] are created for rumour detection by collecting user engagements from tweets on Twitter. User propagation data provides valuable insights into the likelihood of different users spreading fake news. This information is vital in social media-based classification, as studies [34] have shown that individuals with a high risk of spreading or interacting with fake news tend to propagate fake news more frequently. By limiting the spread of news by high-risk users early, the spread of fake news can be greatly reduced.

D. Multi-Category Datasets

Finally, datasets such as Fakeddit, Credbank, and FakeNewsNet incorporate multiple categories of data to create a robust dataset. They are becoming increasingly popular as they provide a comprehensive list of propagation-related features, as well as the article's original headline and text content for fake news detection. Among them, Fakeddit provides details about user engagement through user comments and user IDs, which can then be expanded to incorporate user profiles. It also includes images from all collected Reddit posts for effective multimodal classification. Furthermore, user comment data is extracted from Reddit posts to aid in text-based classification.

E. Multi-Class Labeling

Modern research in fake news [3] has expanded from the binary classification problem of fake news detection to explain the type of fake news each article imposes. Datasets such as Liar, BuzzFeed, FEVER [35], Fakeddit [17], and Credbank have developed their datasets to incorporate three-, four-, or six-way labelling methods. Accurate categorization of fake news into types such as satire, manipulation, and false image connection offers significant advantages in psychological-based fake news classification. This form of categorization greatly assists detection models to better understand psychological cues within fake news to improve the classification accuracy.

In this work, we incorporate the user comment data from the Fakeddit dataset and utilize news articles and labelling inherited from the NELA-GT-2019 dataset as components of the PLNCC dataset. Compared to the Fakeddit dataset, the PLNCC dataset includes the news article with the headline, as well as labels based on the article instead of user comments. Unlike the mentioned propagation-based datasets, PLNCC includes user comments instead of the propagation graph, which allows researchers to use simpler and more efficient models for earlier fake news detection. In addition, we expand beyond sentiment analysis and emotions used in other work and explore a wide range of psychological features through cognitive NLP.

There are multimodal fake news detection methods [**?**] that can detect the intramodality and cross-modal correlation relationships between image regions and text fragments. Combining information from various modalities, such as textual content [22], user metadata [14], and network structures [8], and using language models [49] can make fake news detection systems more efficient and robust. Our work draws inspiration from prior research, with a particular emphasis on data and feature representations.

III. Methodology

A. Problem Definition

Our text-based Psycho-Linguistic News Content and Comments (PL-NCC) fake news dataset is designed to aid in identifying fake news by analyzing the text within news articles and user comments. We further leverage this dataset to develop a News Content and Comments (NCC) based classification model to detect fake news. The goal of this work is to determine the veracity of news articles in the input news corpus using a binary classification model.

B. Data Description

In this work, we merge the text-based content from NELA-GT-2019 with the user comments from Fakeddit to create a unified dataset to incorporate news articles with user comments. Furthermore, we create a diverse set of linguistic and psychological features extracted from this unified dataset, and offer these feature values directly in our PL-NCC dataset. In this way, we can easily perform the task of fake news detection. A breakdown of the PL-NCC dataset is presented in Table I.

TABLE I
DATA SUMMARY

Total Articles	2,929
Total User Comments	41,867
Collection Year	2019
Fake News Percentage	69.9%
Real News Percentage	30.1%
Data Balance	Balanced
Data Components	Articles' headline, text, comments, classification label, and metadata
Data Sources	NELA (for articles' text and headlines), Fakeddit (for comment IDs)
Extra Data Retrieval	Reddit's API (for full user comments, date-time stamps, and upvotes)
Comment Data Window	First eight hours of user comment data after publication on Reddit
Dataset Combination	All the information is concatenated into the combined dataset

1) Data Pre-Processing: The NELA dataset consists of news articles collected from different news sources throughout 2019. Of these news sources, the dataset contains the news headline, the article's text, and metadata related to the time and source of publication. Fakeddit sources the dataset over a ten-year span of Reddit user comments from several selectively chosen subreddits. In the Fakeddit dataset, the user comments in the chosen subreddits, the publication timestamps and metadata related to each comment are extracted. The original news article titles from NELA and Reddit post headlines from Fakeddit are matched to concatenate the two datasets. After concatenation, we extract the article headline, article text, user comments, date and time of publication, and the binary classification label from NELA as features to create the pre-processed dataset.

The PL-NCC dataset leverages the news articles and user comments of the two datasets to extract psychological and linguistic characteristics of the textual content. Recent studies [13], [36], [37], indicate linguistic and psychological features provide valuable information for the classification task. Our dataset aims to improve on existing fake news datasets by including these features as part of the dataset.

2) Labeling: We inherit the two-way labels from NELA dataset to classify all news articles in the dataset. Although Fakeddit provides two-, three-, and six-way labelling in their dataset, these labels are based on a full subreddit's credibility score rather than a per-comment basis. As a result, these labels are much broader compared to the news source level NELA uses for labelling. Thus, we utilize the labelling system from the NELA dataset for our dataset, which obtains labelling based on the credibility of the publishing news source. Our final dataset will contain zero to many user comments for each news article in the dataset. For each article, one binary label is provided for evaluation of the classification model. A recent study by Raza and Ding [14] indicates that this form of dataset concatenation and labelling is effective for the classification of fake news.

C. Model Architecture

We express the classification outcome as a binary value, Y = (0, 1,), where zero indicates a truthful news article, and one signifies fake news. We define each article in N number of news articles as a set P=(P_h, P_c, P_c), where P_h represents the headline of an article, P_c represents the article's textual content, and P_u represents the related user comments.

To improve the classification task of existing fake news detection models, we developed the PL-NCC dataset, which includes the extracted linguistic and psychological features from input P. To obtain the linguistic features, we utilize natural language processing (NLP) to obtain the numerical embedding from the article's text content and user comments. We obtain the psychological features by feeding the article text and comments into the Linguistic Inquiry and Word Count (LIWC) dictionary [39] to obtain the respective feature scores. We test our dataset using our proposed NCC classification model, where the numeric embedding and feature scores are processed through a multilayer perception model (MLP). To improve the classification task of the MLP model, we introduce a feed-forward linear layer prior to model training to improve the efficacy of the extracted psychological features. After feature concatenation, our NCC model is trained using the concatenated features as an input, and after the classification task, the output of the model is the final binary prediction of our model Y, which identifies the veracity of the news articles in N.

D. PL-NCC Dataset

Our particular contribution to fake news research is on improving the effectiveness of textual features extracted from the NELA and Fakeddit datasets for fake news classification. To achieve this, we utilize natural language processing (NLP) techniques, including parts of speech (POS), context-free grammar (CFG), disinformation-related attributes (DIA), clickbait-related attributes (CBA) [25] and BERT embedding [38] and leverage the Linguistic Inquiry and Word Count (LIWC) dictionary [39] to extract linguistic and psychological features from the input data, specifically the PL-NCC dataset. This approach helps us successfully accomplish our objective of enhancing fake news classification. An overview of the feature breakdown for the PL-NCC dataset is presented in Fig. 2.

1) Linguistic Features: To improve the simplicity of classification models, we extract and store different sets of linguistic features from the articles' text and user comments in the PL-NCC dataset. Common features in most text-based classifica-

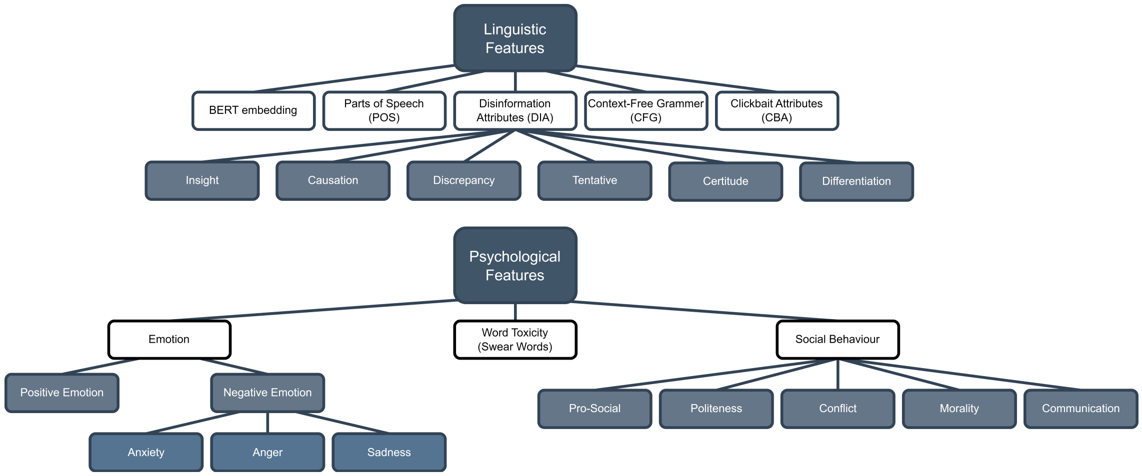


Fig. 2. Overview of PL-NCC Feature Breakdown

tion models include POS and CFG; thus, these features are included in the dataset.

Recent classification models have extracted newer forms of embedding models such as BERT [38], [40] to aid in text-based classification. Thus, we extract the BERT embeddings from the article's text, headline and user comments to include in the dataset. Detection models [41] have explored the use of clickbait and disinformation-related attributes from the article's text to aid in fake news classification. Disinformation traits such as discrepancy (terms which provide reasoning), causation (explanations), tentative (terms defining potential or conditions), insight (individual thoughts and knowledge), certitude (terms defining certainty), and differentiation (terms comparing variance) as well as click-bait attributes such as "this will blow your mind" and "can change your life" provide contrasting differences in fake and real news, which is beneficial in fake news classification. A higher score indicates that the feature is more likely to be real or fake news, respectively. We obtain a label for clickbait attributes by comparing the headline of the news article against a dictionary of forty-seven commonly used clickbait titles. If an article's headline contains any clickbait titles, it is assigned a score of one; otherwise, a score of zero is assigned. We illustrate these patterns in Table II.

2) Psychological Features: Recent studies [12], [13] in early fake news detection focus on different psychological features to improve the performance of the classification task. We classify psychological features into three feature groups: emotions, word toxicity, and social behavioural traits. Social behavioural traits such as morality, politeness, communication (addressing a topic or subject's stance), interpersonal conflict (conflict between two subjects), and pro-social behaviour (voluntary act to help others) show varying patterns in fake and real news, as shown in Table **??**. Additionally, we extract the negative and positive emotions within the article's text and user comments to assist with the classification task. With negative emotional features, tones such as anxiety, anger, and sadness can be extracted from the text and used for classifying fake news. Finally, word toxicity, such as the use of swear words is included as a feature in the PL-NCC dataset. We perform an in-depth analysis in Section IV to study the effects of the mentioned psychological feature groups in fake news classification.

Using the articles' text content and headlines, as well as user comment data from the combined NELA and Fakeddit datasets, we obtain the mentioned psychological features using the LIWC dictionary. LIWC utilizes TF-IDF weights to compare the frequency of words or phrases in the input text against the LIWC dictionary. LIWC then generates a numerical score out of one hundred for each psychological feature in the article's text and user comments. The resulting output indicates the percentage of terms in the input text included in the LIWC dictionary. Additionally, all linguistic and psychological features in our dataset, excluding BERT embedding, are represented as a percentage score out of 100.

E. NCC Classification Model

Figure 3 showcases the architecture of the proposed NCC classification model working in conjunction with the PL-NCC dataset.

1) Linear feed-forward layer: This work aims to explore the benefits of using psychological features with the classification of fake news. Before performing the classification task, we run the extracted psychological features from the PL-NCC dataset through a feed-forward linear layer. The inclusion of this layer emphasizes characteristics extracted from the psychological features which are unique to fake news articles. By emphasizing these traits, our classification model can better identify the veracity of fake news articles during the classification task. The output of this layer is an updated numeric embedding of the psychological features with emphasized fake news values.

2) Multilayer perceptron (MLP): For the classification task, we utilize a standard multilayer perceptron (MLP), acting as a deep neural network to train and classify both the linguistic

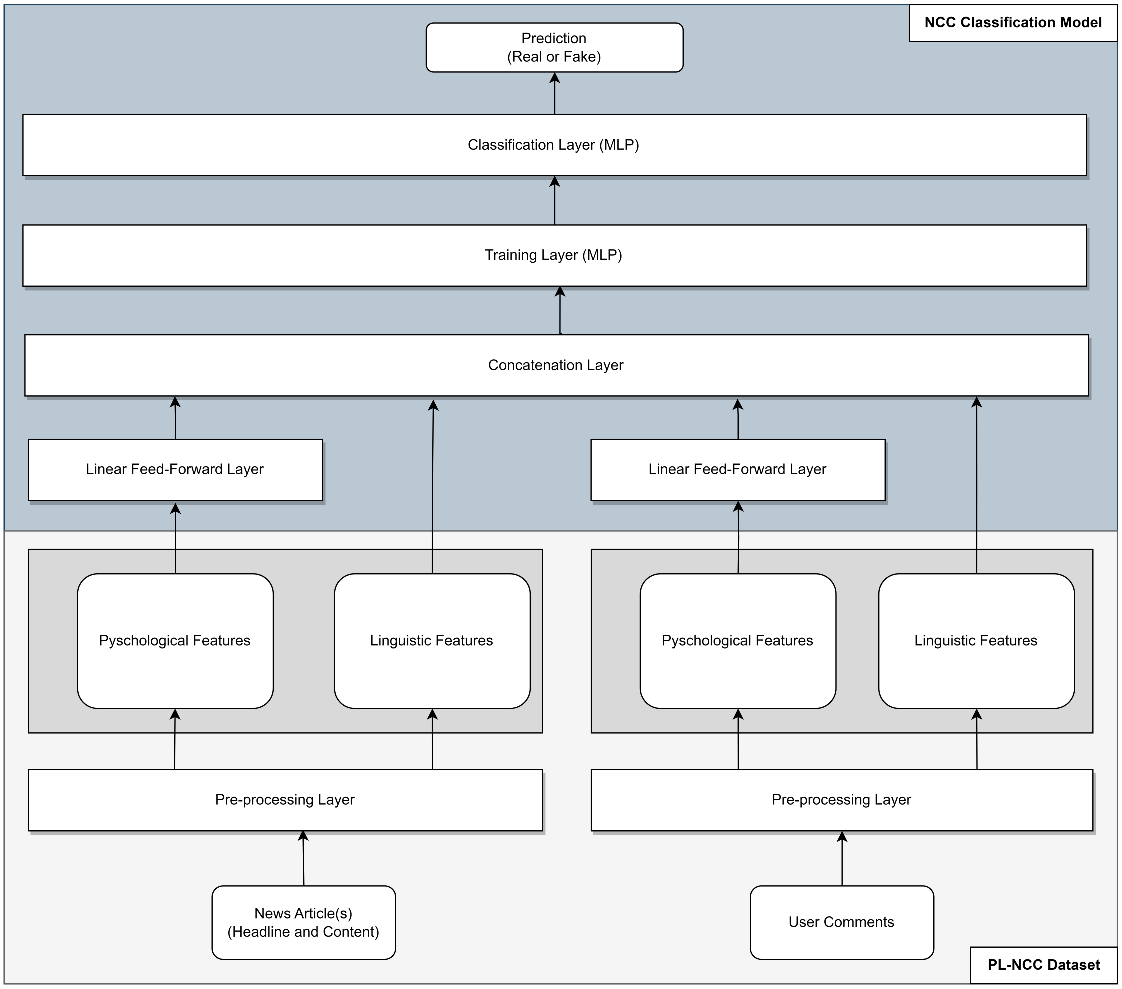


Fig. 3. Architecture of Proposed NCC Classification Model using PL-NCC Dataset

and psychological features extracted from the news content and user comments. We fit our neural model with one hundred layers, consisting of one input layer, one output layer, and ninety-eight hidden layers. After training and classification, our neural network returns the binary prediction Y of the input dataset N. Our model is optimized using the Adam optimizer and utilizes the rectified linear unit (ReLU) activation function defined below, where x defines the value of the input.

$$f(x) = max(0, x) \tag{1}$$

When training our model, we define the following sparse categorical cross-entropy loss function to evaluate the performance of our model's training:

$$L = -(ylog(p) + (1 - y)log(1 - p)) \tag{2}$$

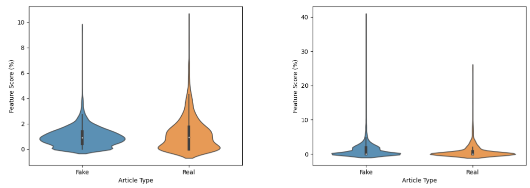
where p represents the predicted value of our model and y represents the actual value.

IV. Experiments

We conduct a series of experiments to showcase the effectiveness of linguistic and psychological features in fake news classification, as well as the efficacy of our proposed NCC classification model.

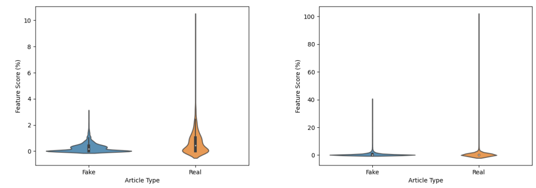
A. *Experimental setup*

Various tools and libraries including NLTK, Keras, and Scikit-Learn are used for data processing and classification tasks. The experiments are run on a computer with 64 gigabytes of RAM, a 16-core 32-thread processor, and an RTX 2070 graphics card. For the classification task, 70% of the dataset is used for training, while the remaining 30% is used as the testing set. To minimize the training loss, each model is trained for fifteen epochs and optimized using the Adam optimizer. The linguistic features extracted using Scikit learn are embedded with a dimensionality size of 768, and the scores extracted using LIWC for the psychological features are used as is for the model's classification. Each model uses a ReLU activation layer, a softmax output layer, and utilizes the sparse categorical cross-entropy loss function.



(a) Article (b) Comments

Fig. 4. Score Distribution of **DIA Discrepancy** Feature for All News Articles and Comments



(a) Article (b) Comments

Fig. 5. Score Distribution of **DIA Certitude** Feature for All News Articles and Comments

We test the effectiveness of the proposed linguistic and psychological features and user comments in our PL-NCC by running the dataset against state-of-the-art classification models, including XGBoost, convolutional neural networks (CNN), multilayer perceptron models (MLP), and decision tree classifiers (DTC). Recent work in the field of fake news research [40], [42] have adapted BERT embedding models to improve the performance of the fake news classification task. In our analysis, we experiment with different language models, including BERT embedding and Bag of Words models, to showcase the benefits of BERT embedding models in fake news classification.

We consider the task of identifying fake news as a binary classification problem, where the resulting output is either a real (0) or fake (1) integer. To evaluate the effectiveness of our dataset against various classification models, we rely on standardized metrics [37], [43] such as accuracy and F1 Score as evaluation metrics.

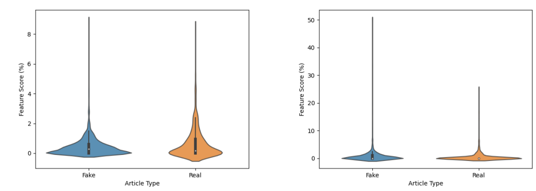
Our dataset provides the embedded values of each linguistic and psychological feature so it can be inputted into a classification model without any pre-processing steps. To train each model, we feed all embedded features as the input for the model's training from the training set and finally use the embeddings from the testing set to obtain the evaluation results from the model.

To assess the impact of each additional feature group, we conduct a series of experiments to determine the efficacy of the PL-NCC dataset using the proposed features. The conducted experiments offer valuable insights into the impact of linguistic and psychological features in fake news detection.

B. Patterns of Linguistic and Psychological Features

Firstly, we analyze the effects of each linguistic and psychological feature in fake and real news and observe the following patterns. The results of our analysis are presented in Table II, where we illustrate the score distribution for each linguistic and psychological feature between real and fake news.

Additionally, we show examples of the score distribution for some features in Fig. 4 to 7. In each figure, the feature score of all articles in the testing corpus are represented in the Y-axis of the violin plot, while the X-axis represents the news category (Fake and Real). The orange plot of the figure indicates fake news articles, while the blue plot represents real news. In our analysis, we identify that fake news articles are more likely to



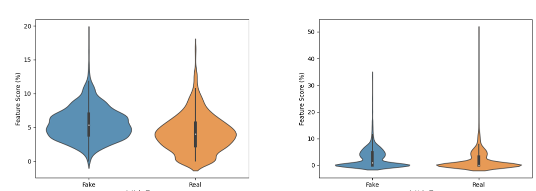
(a) Article (b) Comments

Fig. 6. Score Distribution of **Negative Emotion** Feature for All News Articles and Comments

have clickbait titles compared to real news articles, as there is a higher average of fake news articles with a higher feature score compared to real news.

Through our analysis, we identify that user comments exhibit higher scores of DIA (disinformation-related attributes) in fake news than in real news, such as discrepancy in Fig. 4. However, the article text has varying results for the same traits in both fake and real news. This analysis signifies the importance of using user comments with news content, as different information can be extracted from the two data types, aiding in fake news detection. Differentiation in the article's text and headline is equally distributed between real and fake news, while certitude is much more frequent in real news than fake, as illustrated in Fig. 5. This result highlights the benefits of this feature in fake news classification.

Recent work with cognitive NLP [12], [13] has shown promising results in the classification task when emotions are extracted from the dataset. Their work analyzed the patterns of emotions in fake news research and identified that fake news is emotionally biased compared to real news. We see similar patterns in our analysis of these features. Fig. 6 illustrates



(a) Article (b) Comments

Fig. 7. Score Distribution of **Social Behaviour** Feature for All News Articles and Comments

TABLE II
FEATURE COMPARISON BETWEEN FAKE AND REAL NEWS (HIGHER AVERAGE SCORE MEANS MORE PROMINENT)

	Article				**Comments**			
	Fake News		**Real News**		**Fake News**		**Real News**	
Feature	Avg.	Range	Avg.	Range	Avg.	Range	Avg.	Range
Clickbait	0.008	0.0 - 1.0	0.024	0.0 - 1.0	-	-	-	-
Insight	1.738	0.0 - 7.71	2.160	0.0 - 18.18	1.580	0.0 - 33.33	1.234	0.0 - 50.0
Causation	1.387	0.0 - 5.98	1.194	0.0 - 11.11	1.088	0.0 - 20.0	0.795	0.0 - 25.0
Discrepancy	0.986	0.0 - 9.52	1.233	0.0 - 10.0	1.264	0.0 - 40.0	0.852	0.0 - 25.0
Tentative	1.348	0.0 - 10.28	1.678	0.0 - 9.09	1.892	0.0 - 33.33	1.109	0.0 - 33.33
Certitude	0.268	0.0 - 2.99	0.701	0.0 - 10.0	0.526	0.0 - 40.0	0.638	0.0 - 100.0
Differentiation	2.471	0.0 - 7.94	2.006	0.0 - 20.0	2.381	0.0 - 22.22	1.558	0.0 - 25.0
Toxicity	0.01	0.0 - 2.04	0.27	0.0 - 10.53	0.576	0.0 - 100.0	0.396	0.0 - 25.0
Pos. Emotion	0.292	0.0 - 7.14	0.498	0.0 - 13.33	0.519	0.0 - 33.33	0.415	0.0 - 20.0
Neg. Emotion	0.430	0.0 - 8.89	0.667	0.0 - 8.33	0.804	0.0 - 50.0	0.510	0.0 - 25.0
Soc. Behaviour	5.557	0.0 - 18.87	4.147	0.0 - 16.67	2.760	0.0 - 33.33	1.828	0.0 - 50.0
Pro Social	0.608	0.0 - 6.06	0.48	0.0 - 7.14	0.351	0.0 - 20.0	0.233	0.0 - 50.0
Politeness	0.344	0.0 - 4.9	0.157	0.0 - 4.76	0.288	0.0 - 16.67	0.129	0.0 - 10.0
Conflict	0.646	0.0 - 9.09	0.612	0.0 - 11.76	0.269	0.0 - 25.0	0.181	0.0 - 9.09
Morality	0.439	0.0 - 5.54	0.503	0.0 - 14.29	0.351	0.0 - 33.33	0.224	0.0 - 25.0
Communication	2.701	0.0 - 14.29	1.562	0.0 - 11.11	1.218	0.0 - 25.0	0.742	0.0 - 50.0

Average represents the **Average Embedding Score** obtained for each feature from LIWC. **Range** represents the lowest and highest **Feature Embedding Score** obtained for each feature from LIWC. **Clickbait attributes** relate to the news headlines only, thus, do not have a comment score.

that emotions are more present in fake news, as indicated by the significantly higher scores in fake news compared to real news. While real news articles have higher negative emotional scores, the difference is not as significant when compared to fake news.

Additionally, user comments in fake news show higher emotional scores, while real news articles show increased scores in negative emotions. As outlined previously, analyzing the different patterns between the news article and user comments provides valuable insight into the importance of user comments in fake news classification. Real news articles have elevated levels of word toxicity, such as swear words, compared to fake news. When analyzing real news articles which contain a high toxicity score, we discover that specific news sources, such as the Onion, are recognized as reliable news sources; however, their writing style contains a substantial amount of offensive language in their headlines, such as ”depressed monkey throwing sh*t at himself”. Sources such as the Onion provide news articles based on current events; however, they are written in the form of satire. Although the content of the news source can sometimes be truthful, these news articles are considered outliers in the data. Additionally, user comments with high levels of toxicity are often correlated with fake news.

C. Effectiveness of User Comments

While articles represent the viewpoint of a single writer, user comments provide valuable perspectives from the public. Negative emotions in comments indicate conflicting views on the article's topic, while positive emotions can indicate support. By examining user comments and word toxicity, we observe a correlation between an increase in toxic language in user comments and news articles that receive higher negative emotional scores from users. As fake news tends to drive controversial topics [3] to generate user engagement and promote news propagation, this can result in users expressing heightened emotional views about the news topic, leading to an increase in toxic language within their comments. Our research also reveals that real news has a higher average score in negative emotions within the article content, as indicated in Table II. Since news sources cover topics like natural disasters, violence, and politics, negative bias may be present in the writing, signifying the higher negative emotional score shown in Fig. 6.

We examine the effects of various social behavioural traits in fake and real news, and present our results in Table II. Traits such as morality and interpersonal conflict are more prevalent in the article content in fake news. Scores closer to 0.5 for morality indicate the news article is morally just, while higher scores for conflict indicate the news contains more conflict-related information. Additionally, pro-social behaviour, politeness and communication have elevated scores in user comments and the article's text in real news. A lower score for the three features indicates that the news is more pro-social, polite, and communicative. Our analysis indicates that using these linguistic and psychological features and user comments provides valuable information in the classification task and improves the performance of existing detection models. By leveraging these patterns, classification models can better differentiate between fake and real news.

D. Effectiveness of PL-NCC Dataset Against Baseline Models

After a thorough analysis of each feature's behaviour in fake and real news, we aim to determine the efficacy of our proposed PL-NCC dataset by comparing its performance against leading datasets such as NELA-GT and Fakeddit.

TABLE III
FAKE NEWS DETECTION ACCURACY ON OUR DATASET COMBINED OVER 5-FOLDS

PL-NCC Dataset						
	Text-Only Input		*With Proposed Features*		**Change in Accuracy**	
Model	**Accuracy**	**F1 Score**	**Accuracy**	**F1 Score**	**Accuracy**	**F1 Score**
BERT	0.868	0.908	0.972	0.980	+ 10.4%	+ 7.2%
DistilBERT	0.871	0.908	0.972	0.979	+ 10.1%	+ 7.1%
MLP	0.957	0.969	0.975	0.982	+ 1.8%	+ 1.3%
XGBoost	0.945	0.961	0.975	0.982	+ 3.0%	+ 2.1%
CNN	0.949	0.963	0.958	0.964	+ 0.9%	+ 0.1%
DTC	0.855	0.897	0.917	0.941	+ 6.2%	+ 4.4%
NCC	**0.957**	**0.969**	**0.978**	**0.984**	+ 2.1%	+ 1.5%

Proposed Features include User Comments, Linguistic and Psychological Features.
BOW embedding of article headline and texts are the input to the models.
The text-only input excludes user comments, linguistic, and psychological features.

TABLE IV
FAKE NEWS DETECTION ACCURACY ON ORIGINAL DATASET COMBINED OVER 5-FOLDS

	NELA-GT Dataset		**Fakeddit Dataset**	
Model	**Accuracy**	**F1 Score**	**Accuracy**	**F1 Score**
BERT	0.920	0.942	0.789	0.847
DistilBERT	0.948	0.962	0.767	0.833
MLP	0.951	0.965	0.656	0.754
XGBoost	0.952	0.966	0.706	0.811
CNN	0.948	0.965	0.679	0.779
DTC	0.853	0.894	0.673	0.770

We compare each dataset against a series of state-of-the-art classification models. To determine the effectiveness of our proposed features, we compare the performance of only the article's text and headline, followed by another set of experiments using all features in our PL-NCC dataset.

To test the performance of our PL-NCC dataset and NCC classification model, we execute the detection task using the selected baseline deep learning and mahcine learning models commonly used for the classification task of fake news [14], [40], [42], including Multilayer Perception (MLP), Convolutional Neural Networks (CNN), XGBoost and a standard Decision Tree (DTC) based classifier, and BERT models. The results of our experiments are shown in Tables III and IV.

Our results showcase that the PL-NCC dataset and NCC classification model outperform baseline models by up to twelve percent, as shown in Tables III and IV. For our work, we exclude the performance of our NCC classification model in the baseline comparisons illustrated in Table IV, as these tests only use the basic text from each respective dataset, and excludes the proposed linear layer of our NCC classification model. Thus, the results for this experiment are identical to the basic MLP classification model as illustrated in the table. In our analysis, the decision tree model is the least performant compared to all other baselines, and our model outperforms all the other models when including the proposed features. In our analysis, we also compare the baseline and PL-NCC datasets against different learning models, including BERT embedding. In our comparison, we include both BERT and DistilBERT embedding models to test our model's performance. As illustrated in our baseline comparison, the DistilBERT language model usually performs better compared to the BERT language model (only worse on Fakeddit dataset). Based on this set of experiments, we conclude that the proposed PL-NCC dataset and NCC classification model perform better than baseline models due to the inclusion of both linguistic and psychological features, user comments and our proposed linear feed-forward layer. When comparing the performance of baseline models on different datasets, we can see that their performance is better when using the PL-NCC dataset. In addition, with the same feature sets, we notice better results when utilizing the article's text and user comments from the PL-NCC dataset, compared to relying solely on NELA-GT's article content or Fakeddit's comment data. The results of our baseline dataset comparison, as illustrated in Table IV indicates that the Fakeddit dataset performs worse for the classification tasks compared to the NELA-GT dataset. Further analysis reveals that the performance drop can be a result of the missing news article content, as Fakeddit only provides the raw user comments related to each news article.

TABLE V
ABLATION STUDY USING ARTICLE TEXT AND USER COMMENTS

Model	**Accuracy**	**F1 Score**
BOW Only	0.723	0.778
BERT Only	0.868	0.908
BERT + POS + CFG	0.956	0.965
BERT + DIA + CBA	0.944	0.957
Linguistic + Emotion	0.963	0.960
Linguistic + Social Behaviour	0.965	0.952
Linguistic + Swear Words	0.968	0.964
Linguistic + Psychological	0.972	0.979

E. *Effectiveness of Linguistic and Psychological Features*

In the next set of experiments, we compare different feature groups to examine the effects of linguistic and psychological features, as depicted in Table V. Throughout these experiments, we utilize the MLP classifier and keep all hyperparameters consistent while modifying the feature sets used. As we gradually include more of the proposed features in the classification task, the model's performance consistently improves, with an increase of up to twenty-five percent from the first experiment in the series. By incorporating user comments and linguistic and psychological features into our dataset, fake news classification models are better equipped to predict fake news compared to traditional text-based datasets.

F. Effectiveness of linear layer

TABLE VI
EFFECTIVENESS OF LINEAR LAYER USING NCC CLASSIFICATION MODEL

Has Linear Layer	(Complete Model) Accuracy	F1 Score
Yes	**0.978**	**0.984**
No	0.957	0.969

In our work, we include a feed-forward linear layer to the NCC classification model to improve the accuracy of classification for the psychological features extracted. To test the efficacy of this linear layer, we run the classification task against the PL-NCC dataset with and without the linear layer. The results of these experiments are shown in Table VI. The standard MLP model without the linear layer performs similarly to the baseline MLP models presented in Tables III and IV. The results of this experiment indicate that there is better performance with the use of a linear layer, as we see up to a two percent increase in performance accuracy.

The linear layer amplifies the weights of psychological features often found in fake news, such as word toxicity and pro-social behaviour, whilst reducing the significance of psychological features common in true news. By emphasizing the psychological traits associated with fake news, our NCC model can more accurately identify the veracity of fake news compared to the baseline MLP models.

G. Effectiveness of neural representation with linguistic features

Recent work on fake news detection [40] has focused on using deep learning techniques to improve the classification of fake news detection. As part of our research, we aim to identify the significance of deep learning language models, such as BERT, and how they compare to traditional linguistic features such as bag-of-words (BOW). Work by Dacrema et al. showcases how BERT embeddings may not be as effective in the task of fake news classification; however, we present the results of our experiments in Table VII, which showcases the unique patterns obtained from BERT and BOW features.

While both BERT and BOW produce numerical embeddings of the input text, BERT performs its prediction using masked words within the article to create its embeddings. However, since BERT is a large model with numerous parameters, it does not perform well when using smaller datasets, such as PolitiFact, as seen in our results. We conclude this is a result of overfitting done within the embedding model. In contrast,

TABLE VII
EFFECTIVENESS OF NEURAL REPRESENTATION WITH LINGUISTIC FEATURES (BOW/BERT) USING NCC CLASSIFICATION MODEL

	BOW		BERT	
Dataset	**Accuracy**	**F1 Score**	**Accuracy**	**F1 Score**
PolitiFact	**0.9028**	**0.9091**	0.8333	0.8462
PL-NCC	0.9682	0.97674	**0.9731**	**0.9803**

BOW generates embeddings by counting word occurrences and does not perform its own classification of the input text. However, when we analyze the results of BERT embedding using a larger dataset, such as our PL-NCC dataset, we see much better performance using BERT embeddings compared to BOW. Our analysis showcases that BERT embedding performs much better with larger datasets as there is a larger vocabulary from the input dataset for the BERT embedding model to train on. Thus, although BERT may underperform with smaller datasets compared to BOW, it still proves to be a novel tool in the classification of fake news.

V. DISCUSSION

The Reddit API limits the number of API calls which can be made at one time. Due to these time limitations, we have only collected the first eight hours of user comments for each news post in the PLNCC dataset. While sufficient for early fake news detection models, general text-based classification models are limited by the eight-hour timespan of the user comments. Future work on this dataset can entail extracting more user comment data using Reddit's API, providing a larger timespan of user comments for text-based classification. To label our dataset, we are reliant on NELA and Fakeddit's source-based labelling. Although effective, alternative labelling methods such as on a per-article basis can provide better training for classification models. Thus, leveraging manual fact-checking sources for each news article can improve the effectiveness of our dataset.

VI. CONCLUSION

As fake news becomes more prominent in modern day, it is imperative that research is focused on the mitigation of fake news propagation, both through the development of robust news datasets as well as accurate and efficient classification models. In this work, we proposed a novel approach to consolidate two state-of-the-art fake news datasets NELA-GT and Fakeddit, by introducing a set of linguistic and psychological features extracted from the news dataset, as well as creating a consolidated dataset PL-NCC to include both news articles and their respective user comments. We then leverage this dataset to develop an improved fake news detection model, using MLP with a feed-forward linear layer to enhance the efficacy of psychological characteristics extracted from the news text and user comments.

In this work, we make contributions to fake news research by: 1) performing an in-depth analysis of the effects of different linguistic and psychological features in both real and fake news; 2) creating a compiled dataset which includes the articles' text content with the related user comments to incorporate user propagation data with text-based classification; and 3) develop an improved fake news dataset which includes pre-embedded values of effective linguistic and psychological features to improve the efficacy of existing fake news classifiers; 4) determining the effectiveness of our proposed PL-NCC dataset using our NCC classification model, which introduces a feed-forward linear layer into the MLP classifier to better identify the veracity of fake news. We execute a series of experiments to demonstrate the effectiveness of our PL-NCC dataset and NCC classification model, and our results indicate that the proposed models provide significant improvements over existing state-of-the-art classification models and datasets, showing up to a twelve percent increase in classification accuracy.

Analysis of our experiments indicates significant benefits to the user of linguistic and psychological features in fake news classification when used in conjunction with user comments and the article text. While psychological features are not as effective for the classification task when used in isolation, we identify that these features excel when used in conjunction with other linguistic features. Specifically, our experiments showcase that social behavioural traits in particular demonstrate the best performance for the classification task. In our set of experiments, the psychological features improve the performance of the classification task. Thus, We include these features as part of our dataset so they can be integrated into newer classification models as deemed fit. In addition, we identify that the inclusion of the feed-forward linear layer in our work enhances the efficacy of the proposed psychological features, improving the classification of fake news even further.

VII. Acknowledgments.

This work is partially sponsored by Natural Science and Engineering Research Council of Canada (grant 2020-04760).

References

[1] S. Vosoughi, D. Roy, and S. Aral, 'The spread of true and false news online', science, vol. 359, no. 6380, pp. 1146–1151, 2018.

[2] X. Zhang and A. A. Ghorbani, 'An overview of online fake news: Characterization, detection, and discussion', Information Processing & Management, vol. 57, no. 2, p. 102025, 2020.

[3] B. D. Horne and S. Adali, 'This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news', in Eleventh international AAAI conference on web and social media, 2017.

[4] N. Ruchansky, S. Seo, and Y. Liu, 'Csi: A hybrid deep model for fake news detection', in Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 2017, pp. 797–806.

[5] Q. Huang, C. Zhou, J. Wu, L. Liu, and B. Wang, 'Deep spatial–temporal structure learning for rumor detection on Twitter', Neural Computing and Applications, no. August, 2020.

[6] X. Zhou and R. Zafarani, 'Network-based fake news detection: A pattern-driven approach', ACM SIGKDD explorations newsletter, vol. 21, no. 2, pp. 48–60, 2019.

[7] F. Qian, C. Gong, K. Sharma, and Y. Liu, 'Neural user response generator: Fake news detection with collective user intelligence', in IJCAI International Joint Conference on Artificial Intelligence, 2018, vol. 2018-July, pp. 3834–3840.

[8] Y. Liu, and Y.-F. B. Wu, 'FNED: a deep network for fake news early detection on social media', ACM Transactions on Information Systems (TOIS) 38:3, pp. 1-33, 2020.

[9] S. Ahmed, K. Hinkelmann, and F. Corradini, "Development of fake news model using machine learning through natural language processing," arXiv preprint arXiv:2201.07489, 2022.

[10] N. R. de Oliveira, P. S. Pisa, M. A. Lopez, D. S. V. de Medeiros, and D. M. Mattos, "Identifying fake news on social networks based on natural language processing: trends and challenges," Information, vol. 12, no. 1, p. 38, 2021.

[11] A. Acerbi, "Cognitive attraction and online misinformation," Palgrave Communications, vol. 5, no. 1, 2019.

[12] X. Zhang, J. Cao, X. Li, Q. Sheng, L. Zhong, and K. Shu, 'Mining dual emotion for fake news detection', in Proceedings of the Web Conference 2021, 2021, pp. 3465–3476.

[13] C. Guo, J. Cao, X. Zhang, K. Shu, and M. Yu, 'Exploiting emotions for fake news detection on social media', arXiv preprint arXiv:1903. 01728, 2019.

[14] S. Raza and C. Ding, 'Fake news detection based on news content and social contexts: a transformer-based approach', International Journal of Data Science and Analytics, vol. 13, no. 4, pp. 335–362, 2022.

[15] S. Gaillard, Z. A. Oláh, S. Venmans, and M. Burke, "Countering the cognitive, linguistic, and psychological underpinnings behind susceptibility to fake news: A review of current literature with special focus on the role of age and digital literacy," Frontiers in Communication, vol. 6, p. 661801, 2021.

[16] M. Gruppi, B. D. Horne, and S. Adalı, 'NELA-GT-2019: A Large Multi-Labelled News Dataset for The Study of Misinformation in News Articles', arXiv [cs.CY]. 2020.

[17] K. Nakamura, S. Levy, and W. Y. Wang, 'r/fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection', arXiv preprint arXiv:1911. 03854, 2019.

[18] G. Pennycook and D. G. Rand, "The psychology of fake news," Trends in cognitive sciences, vol. 25, no. 5, pp. 388–402, 2021.

[19] E. C. Tandoc Jr, Z. W. Lim, and R. Ling, "Defining 'fake news' A typology of scholarly definitions," Digital journalism, vol. 6, no. 2, pp. 137–153, 2018.

[20] B. D. Horne, W. Dron, S. Khedr, and S. Adali, 'Assessing the News Landscape: A Multi-Module Toolkit for Evaluating the Credibility of News', in The Web Conference 2018 - Companion of the World Wide Web Conference, WWW 2018, 2018, pp. 235–238.

[21] T. Mitra and E. Gilbert, 'Credbank: A large-scale social media corpus with associated credibility annotations', in Proceedings of the international AAAI conference on web and social media, 2015, vol. 9, pp. 258–267.

[22] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, 'Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media', Big data, vol. 8, no. 3, pp. 171–188, 2020.

[23] W. Y. Wang, '" liar, liar pants on fire": A new benchmark dataset for fake news detection', arXiv preprint arXiv:1705. 00648, 2017.

[24] R. Baly, G. Karadzhov, D. Alexandrov, J. Glass, and P. Nakov, 'Predicting factuality of reporting and bias of news media sources', arXiv preprint arXiv:1810. 01765, 2018.

[25] X. Zhou, A. Jain, V. V. Phoha, and R. Zafarani, 'Fake news early detection: A theory-driven model', Digital Threats: Research and Practice, vol. 1, no. 2, pp. 1–25, 2020.

[26] X. Zhou, J. Wu, and R. Zafarani, 'SAFE: Similarity-Aware Multi-modal Fake News Detection', 2020.

[27] P. K. Verma, P. Agrawal, I. Amorim, and R. Prodan, 'WELFake: Word embedding over linguistic features for fake news detection', IEEE Transactions on Computational Social Systems, vol. 8, no. 4, pp. 881–893, 2021.

[28] Y. Wu, P. Zhan, Y. Zhang, L. Wang, and Z. Xu, 'Multimodal fusion with co-attention networks for fake news detection', in Findings of the association for computational linguistics: ACL-IJCNLP 2021, 2021, pp. 2560–2569.

[29] V. Christlein, C. Riess, J. Jordan, C. Riess, and E. Angelopoulou, 'An evaluation of popular copy-move forgery detection approaches', IEEE Transactions on information forensics and security, vol. 7, no. 6, pp. 1841–1854, 2012.

[30] S. Heller, L. Rossetto, and H. Schuldt, 'The ps-battles dataset-an image collection for image manipulation detection', arXiv preprint arXiv:1804. 04866, 2018.

[31] S. Dadkhah, X. Zhang, A. G. Weismann, A. Firouzi, and A. A. Ghorbani, 'TruthSeeker: The Largest Social Media Ground-Truth Dataset for Real/Fake Content', 2023.
[32] E. Kochkina, M. Liakata, and A. Zubiaga, 'All-in-one: Multi-task learning for rumour verification', arXiv preprint arXiv:1806. 03713, 2018.
[33] L. Derczynski, K. Bontcheva, M. Liakata, R. Procter, G. W. S. Hoi, and A. Zubiaga, 'SemEval-2017 Task 8: RumourEval: Determining rumour veracity and support for rumours', arXiv preprint arXiv:1704. 05972, 2017.
[34] M. F. L. Schmitt and E. J. Spinosa, 'Scalable stream-based recommendations with random walks on incremental graph of sequential interactions with implicit feedback', User Modeling and User-Adapted Interaction, pp. 1–31, 2022.
[35] J. Thorne, A. Vlachos, C. Christodoulopoulos, and A. Mittal, 'Fever: a large-scale dataset for fact extraction and verification', arXiv preprint arXiv:1803. 05355, 2018.
[36] 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.
[37] K. Shu, S. Wang, and H. Liu, 'Beyond news contents: The role of social context for fake news detection', WSDM 2019 - Proceedings of the 12th ACM International Conference on Web Search and Data Mining, vol. 9, pp. 312–320, 2019.
[38] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, 'BERT: Pre-training of deep bidirectional transformers for language understanding', arXiv preprint arXiv:1810. 04805, 2018.
[39] J. W. Pennebaker, M. E. Francis, and R. J. Booth, 'Linguistic inquiry and word count: LIWC 2001', Mahway: Lawrence Erlbaum Associates, vol. 71, no. 2001, p. 2001, 2001.
[40] M. Szczepański, M. Pawlicki, R. Kozik, and M. Choraś, 'New explainability method for BERT-based model in fake news detection', Scientific Reports, vol. 11, no. 1, pp. 1–13, 2021.
[41] X. Zhou and R. Zafarani, 'A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities', 2020.
[42] R. K. Kaliyar, A. Goswami, and P. Narang, 'FakeBERT: Fake news detection in social media with a BERT-based deep learning approach', Multimedia tools and applications, vol. 80, no. 8, pp. 11765–11788, 2021.
[43] N. Vo and K. Lee, 'Learning from fact-checkers: Analysis and generation of fact-checking language', SIGIR 2019 - Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 335–344, 2019.
[44] R. Zafarani, X. Zhou, K. Shu, and H. Liu, "Fake news research: Theories, detection strategies, and open problems," in Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery data mining, 2019, pp. 3207–3208.
[45] L. Chao, X. Wu, M. Yu, G. Li, J. Jiang, W. Huang, and X. Lu. "A two-stage model based on BERT for short fake news detection." In Knowledge Science, Engineering and Management: 12th International Conference, KSEM 2019, Athens, Greece, August 28–30, 2019, Proceedings, Part II 12, pp. 172-183. Springer International Publishing, 2019.
[46] Fang, X., Wu, H., Jing, J., Meng, Y., Yu, B., Yu, H., Zhang, H. (2024). NSEP: Early fake news detection via news semantic environment perception. Information Processing Management, 61(2), 103594.
[47] Gaillard, S., Oláh, Z.A., Venmans, S. and Burke, M., 2021. Countering the cognitive, linguistic, and psychological underpinnings behind susceptibility to fake news: A review of current literature with special focus on the role of age and digital literacy. Frontiers in Communication, 6, p.661801.
[48] Yang, H., Zhang, J., Zhang, L., Cheng, X. and Hu, Z., 2024. MRAN: Multimodal relationship-aware attention network for fake news detection. Computer Standards Interfaces, 89, p.103822.
[49] Chen, C. and Shu, K., 2023. Can llm-generated misinformation be detected?. *arXiv preprint arXiv:2309.13788*.