A Deep Neural Network Approach for Fake News Detection using Linguistic and Psychological Features

Keshopan Arunthavachelvan[1], Shaina Raza[2], Chen Ding[1*]

[1]Computer Science, Toronto Metropolitan University, 350 Victoria St, Toronto, M5B 2K3, Ontario, Canada.

[2]AI Engineering, Vector Institute for Artificial Intelligence, 108 College St, Toronto, M5G 0C6, Ontario, Canada.

*Corresponding author(s). E-mail(s): cding@torontomu.ca;
Contributing authors: karunthavachelvan@torontomu.ca;
shaina.raza@vectorinstitute.ai;

Abstract

With the prominence of online social networks, news has become more accessible to a global audience. However, in the meantime, it has become increasingly difficult for individuals to differentiate between real and fake news. To reduce the spread of fake news, researchers have developed different classification models to identify fake news. In this paper, we propose a fake news detection system using a multilayer perceptron (MLP) model, which leverages linguistic and psychological features to determine the truthfulness of a news article. The model uses different features from the article's text content to detect fake news. In the experiment, we utilize a public dataset from the FakeNewsNet repository consisting of real and fake news articles collected from PolitiFact and BuzzFeed. We perform a meta-analysis to compare our model's performance with existing classification models using the same feature sets and evaluate the performance using the metrics such as prediction accuracy and F1 score. Overall, our classification model produces better results than existing baseline models, by achieving an accuracy and F1 score above ninety percent and performs three percent better than the best performing baseline method. The inclusion of linguistic and psychological features with a deep neural network allows our model to consistently and accurately classify fake news with ever-changing forms of news events.

Keywords: fake news classification, multilayer-perceptron model, linguistic features, psychological features, deep neural network

Self-assessment

1. **What is the main research question that your planned submission addresses?**
   How do linguistic and psychological features affect the performance of state-of-the-art fake news classification models?

2. **What makes your research results important and worth being reported in a top-ranked journal (as opposed to a conference)?**
   Our work is important and worth being reported in a top-ranked journal as opposed to a conference because we perform and share a detailed meta-analysis on the effects different linguistic and psychological features have in existing fake news classification models. Additionally, a detailed comparison is provided indicating the impact in performance BERT-based models have in fake news classification compared to standard bag of words (BOW).

3. **Why does your planned submission fit into the scope of UMUAI?**
   Our study contributes to the field of fake news research by studying the effects linguistic and psychological features have on fake news detection. This study aligns with the special issue of news personalization and analytics because we focus our research in reducing fake news dissemination on online news platforms.

4. **What are the main limitations of your approach?**
   The main limitation of our approach is the size of the dataset used for our experimentation. We utilize a sample of news articles from PolitiFact and BuzzFeed sources; however, the size of this dataset is smaller than the data available in existing news outlets.

5. **What is the relationship of your work to the closest 2-3 publications by others?**
   Similar work by Zhou et al [67] focuses on the effect linguistic features of a news article's text have in fake news detection. Our study builds upon this study by examining the effects linguistic and psychological features have on fake news detection. In addition, recent studies [16, 65] focus on the effect emotions have in fake news classification. In this paper, we expand on similar studies by including additional psychological features and emotions to examine their effect in fake news classification.

1 Introduction

Online social networks have encouraged the spread of news at a much faster rate in recent years. However, the easy access of these networks has led to an increase in fake news dissemination. It is imperative for these social networks to implement detection systems to help reduce the spread of fake news on their platforms. The term 'fake news' refers to the false or misleading information that appears as real news. It aims to deceive or mislead people. Fake news comes in many forms, such as clickbait (misleading headlines), disinformation (with malicious intention to mislead the public), misinformation (false information regardless of the motive behind), hoax, parody, satire, rumour, deceptive news, and other forms as discussed in the literature [70]. Fake news detection models are usually built from four perspectives [70]: false knowledge the news article carries, writing style, propagation patterns and the credibility of the source. Most existing research work in this domain focuses on the article's writing style and propagation patterns [19, 67]. However, in the early stage of fake news spread, there is a lack of information available in regards to the article's propagation patterns [26], so writing style, or content-based information is the more reliable source for fake news detection. In this work, we rely on content of the news article to build our detection model for fake news.

Various content-based features have been used in the fake news detection models, including writing style, textual complexity, and sentiment analysis [19]. Reza et al. [67] have tested a comprehensive list of news content features at lexicon-level, syntax-level, semantic-level and discourse-level and fed them into machine learning models, such as XGBoost or Random Forest, for fake news detection. In this work, we test whether adding psychological features into this list of linguistic features can further improve the detection accuracy of the fake news.

Psychological traits have noticeable effects when being used to differentiate between real and fake news. We group psychological features into three categories: emotions, swear words, and social behaviour. Fake news articles commonly contain emotional (positive, negative and neutral emotions) bias and swear words, while truthful articles are void (or lack) of these traits [16, 65]. Thus, these features provide value when identifying fake news. Additionally, fake news articles often lack informative and professional vocabulary, which is commonly exhibited in real news. As a result, we have chosen to include social behaviour in the model's feature set to help identify fake news. Features such as pro-social behaviour, politeness, interpersonal conflict, moralization and communication can be categorized as social behaviour [6]. Including these psychological traits aids in fake news detection and allows our model to perform better than baseline models.

Earlier efforts in fake news detection [17, 19] apply feature engineering to pick a set of features, such as content-based features, to feed into a classification model to perform the detection task. With the advent of deep neural network models and its wide success in various machine learning tasks, many recent fake news detection models have started building neural representations of news articles for their classification task.

In this paper, we propose a style-based fake news detection model, which analyzes the linguistic and psychological writing patterns of the article's text, and then applies

deep neural network for text-based classification task. In our model, we consider the lexical, semantic, syntactical features and psychological features using an article's textual content as the input to our detection model. We use deep neural networks to test whether we can further improve the detection accuracy of existing fake news classification models.

We perform an extensive set of experiments to compare the performance of our model with state-of-the-art models in the field of fake news research. We use fake news datasets from PolitiFact and BuzzFeed sources [53] and compare the performance of each model with an array of different feature sets and embedding systems by measuring both the accuracy of predictions and the F1 score. Overall, our model performs better than existing fake news classifiers while achieving an overall accuracy and F1 score of ninety percent. We also perform a comparative study between our model and a subset of the baseline models we used in the first set of experiments on a bigger fake news dataset PL-NCC [2] and get a similar conclusion.

Our contributions to the field of research can be summarized as the following:

1. We consider a mixture of content-based textual features. In addition to commonly used linguistic features, we also consider three psychological feature groups, including social behaviour, swear words and emotions. Research shows that psychological features in linguistic problems can help us better identify the veracity of news information [17, 65].
2. We use a deep neural network model (MLP) to build the fake news detection model. While the usage of deep neural networks is not novel in this field of research, we test several different model configurations by feeding the linguistic and psychological news features and report the best configurations for fake news research.
3. We perform extensive meta-analysis on different permutations of feature sets and test different configurations of the proposed model to identify the best setup of the model.
4. We test the effectiveness of using embedding vectors to represent news articles versus traditional bag of words (BOW) vectors.

The rest of the paper is organized as follows. We discuss related work conducted in fake news research in Section two. Section three focuses on the methodology of our model. We explain the experiment setup and analyze the results in Section four. Finally, we conclude the paper and address limitations and future work for this research in Section five .

2 Related work

State-of-the-art fake news detection methods can be categorized broadly into two types: (i) manual and (ii) automatic detection methods. Fact-checking websites, such as Reporterslab and PolitiFact, usually rely on human judgement to decide the truthfulness of the news. They can provide the ground truth (true/false labels) to determine the truthfulness of news. However, it is time-consuming to detect and report every fake news piece manually. The automatic detection methods are alternatives to the manual fact-checking ones. These detection models are broadly categorized

into knowledge-based, propagation pattern-based, source-based, content-based, and style-based methods.

Knowledge-based detection models [19, 56, 70] evaluate news authenticity by inferring the knowledge extracted from to-be-verified news content within a Knowledge Graph (KG).

Propagation-based detection models [32] analyze the spreading pattern of news across a platform to determine if the news is fake or real. These models [22, 24, 32, 44] utilize information related to the dissemination of fake news, e.g., how users spread it, and use techniques such as graphs and multi-dimensional points for fake news detection [22, 32].

Source-based detection models [4, 19, 20, 43, 56] compare articles published by individuals or organizations by analyzing the writer's credibility to determine whether a news article is fake or real. The general idea behind this type of models is that publishers/users with low credibility have a higher chance of spreading fake news than individuals with much higher credibility. Users who frequently post fake news will have lower credibility than those who post truthful information.

The content-based models [18, 19, 42, 43, 58] use various types of content-related information from the news such as article content, headline, image/video, to capture differences between the writing style of a fake news text and the style of a truthful news text to build fake news detection classifiers. For example, Horne and Adalı [17] extract stylometry and psychological features from the news titles to differentiate fake news from real news. Przybyla et al. [43] develop a text classifier using bidirectional LSTMs to capture style-based features from news articles. Zellers et al. [66] develop a neural network model to determine the veracity of news based on the news text. Some other works [58, 69] consider features such as lexicons (frequency of words), syntax (frequency of verbs and nouns), disclosure (frequency of rhetorical words in sentences), latent topics, and semantics (frequency of multiple factors such as complexity of words, sentiment, etc.) for fake news detection.

A significant body of research has begun to focus on social contexts to detect fake news. The social context-based detection methods examine users' social interactions and extract relevant features representing users' posts (review/post, comments, replies) and network aspects (follower-followee relationships) from social media. For example, Liu and Wu [33] propose a neural network-based classifier that uses social media tweets, retweet sequences, and Twitter user profiles to determine the veracity of the news. Recently, transfer learning-based methods have been applied to detect fake news [25, 31]. Some works propose matrix factorization methods to model the relationships among the publishers, news stories and social media users for fake news detection [57].

With the advancement of deep learning and its successful application in multiple domains, more and more research work has started to use deep neural network models to build fake news detection systems. CNN and LSTM methods are used to combine various text-based features, such as from statements (claims) related to news data [28]. RNN and CNN methods are used to build propagation paths for detecting fake news at the early stage of its propagation [32]. An explainable fake news detection framework is proposed using LSTM networks [66]. A deep diffusive model is used to

learn the relationship among news content, news subjects and news creators in an augmented heterogeneous social network [64] for predicting the credibility of the news content, subjects and creators at the same time. Similar idea of multi-task learning is also used in [30] to detect fake news, novelty and emotion of the text simultaneously. Graph Neural Networks (GNN) have been used in many different ways for building the fake news detection systems [41]. For example, in [37], the graph learning framework is used to learn the representations of social contexts for fake news detection.

In recent years, there has been a great focus in NLP research on using transformer-based language models. BERT [10] and GPT [45] are two state-of-the-art such models. BERT has been used in a number of fake news detection models [25, 27, 31, 36, 61]. In [45], GPT-2 is used for the task of fake news detection. Training these language models can be time-consuming. So, a common practice is to use the pre-trained language models (PLM) with fine-tuning. Whitehouse et al. [63] argue that these PLMs are trained on general dataset and they don't have the up-to-date vocabulary on latest news events. A possible solution is to integrate the knowledge base. They test different integration methods and get mixed results (promising on one dataset, not so good on the other). Alghamdi et al. [1] evaluate the performance of adding other neural network structures on top of different BERT variations and see the improvement when BiGRU is used on COVID-Twitter-BERT. Hu et al. [21] find that a large language model (LLM) such as GPT-3.5 may not perform as well as a smaller, fine-tuned language model, however, it may serve as a good advisor to small PLMs by providing multi-perspective rationales.

Fake news detection can have merits outside of simple classification too. Recommendation system has become an important tool to alleviate the information overload problem in various domains. In the news domain, a news recommendation system recommends news articles based on user's past reading behavior. As the user engages with fake news sources, recommendation systems may suggest more similar articles to the user, thus increasing the exposure of fake news to the user [3]. This can become serious when content-based recommendation approaches are used. For collaborative filtering approaches that make recommendations based on other users' likes or diversity-integrated approaches that target at recommending dissimilar items to enhance diversity, it may not be that serious. Some studies [7, 52, 54] have explored combining recommendation systems with social context to perform early detection to control the news suggested to users. These detection models can leverage negative social engagement to suppress the propagation of fake news [50], subsequently reducing exposure to fake news.

Among the different directions of fake news research discussed, there have been promising studies [48] conducted which combine the different detection methods. Specifically, the combination of content-based and social contexts has proven useful in the field, allowing detection models to predict fake news early on in a news article's life cycle. By combining different fields of fake news research, detection models are capable of leveraging the benefits of each method with additional methods to remedy any drawbacks of other models.

In this work, we propose a content-based approach with different layers to perform fake news classification. Using the article's headline and text content, we extract a

wide selection of linguistic and psychological features and apply a deep neural network for classification. Additionally, we compare the efficacy of basic Bag of Words (BOW) vectors versus neural embedding vectors such as BERT for news article representation.

3 Methodology

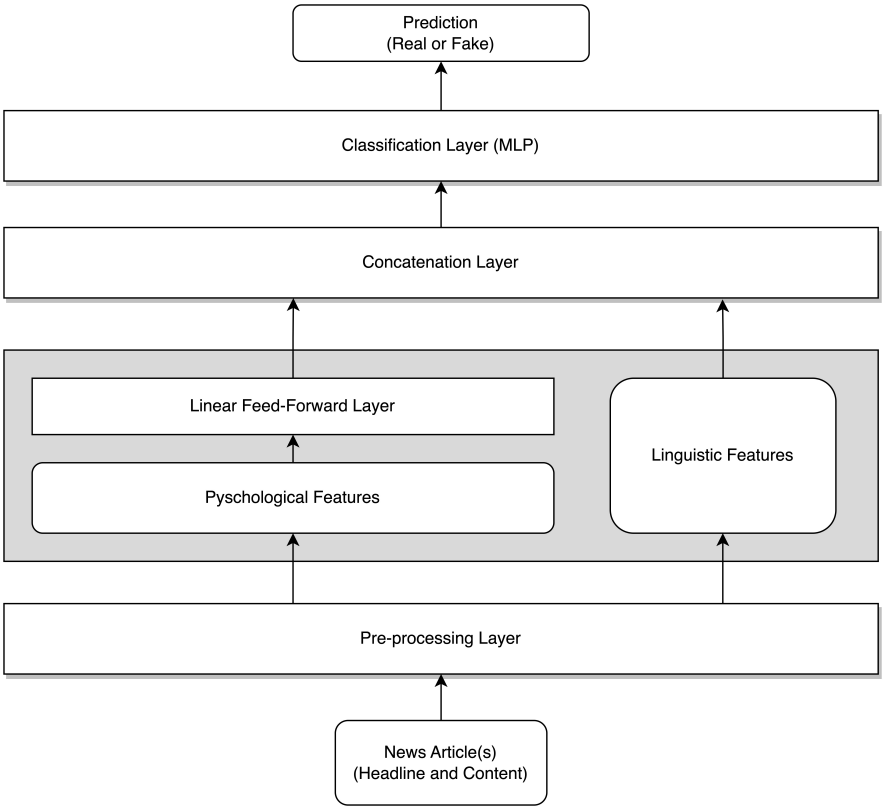

Fig. 1: Overview of proposed model

3.1 Problem definition

Our content-based classification model uses the textual content of news article(s) to detect fake news. Given N news articles, the problem defined in this work is to detect whether a news article is fake or real. We represent the output of our classification as a binary value of $Y = (0, 1)$, where 0 represents a true news article and 1 represents a fake news article.

3.2 Model architecture

Figure 1 provides an overview of our proposed model. We define each article in N number of news articles as a set P=(P_h, P_c), where P_h represents the headline of an article, and P_c represents the article's textual body content.

We want to improve the performance of the fake news classification system by extracting both the linguistic and psychological features from input P. We utilize natural language processing (NLP) to convert the textual content of the article into a numeric representation for linguistic features. Linguistic Inquiry and Word Count (LIWC) dictionary[1] is then used to further extract related psychological features. This diverse set of numeric features is then processed by a MultiLayer Perceptron model (MLP) for fake news classification. Before classification, we run our psychological features through a feed-forward linear layer and then concatenate both linguistic and psychological features for classification. The output from our concatenation layer is used as the input to the deep neural network. These features are used to train our classification model. Once trained, the model will perform its classification on the testing dataset. The output of our classification layer is the final binary prediction of Y which indicates the article's truthfulness. The grouping of the linguistic and psychological features we include in our model can be seen in Figure 2.

3.3 Feature extraction

Table 1 illustrates the data model for the input and output of each layer in the proposed model.

Table 1: Datatype for layers in proposed model

		Datatype	
Layer	Feature	input	output
dataset input	headline	string	string
dataset input	text content	string	string
pre-processing	psychological features	string	float
pre-processing	text content	string	float
linear feed-forward	processed psychological features	float	float
concatenation layer	linguistic and psychological features	float	float
classification (MLP)	concatenated features	float	integer (0, 1)
prediction	classification prediction	integer (0, 1)	-

3.3.1 Linguistic features

Linguistics is a widely studied field in fake news research. Our proposed model uses headlines and textual content to extract linguistic feature groups from the article [67]. The weighted occurrences of bag of words (BOW), parts of speech (POS), and context-free grammar (CFG) from the input text are used as features for the proposed model and are structured as embedding vectors. Each feature in the

[1] https://www.liwc.app/help/howitworks

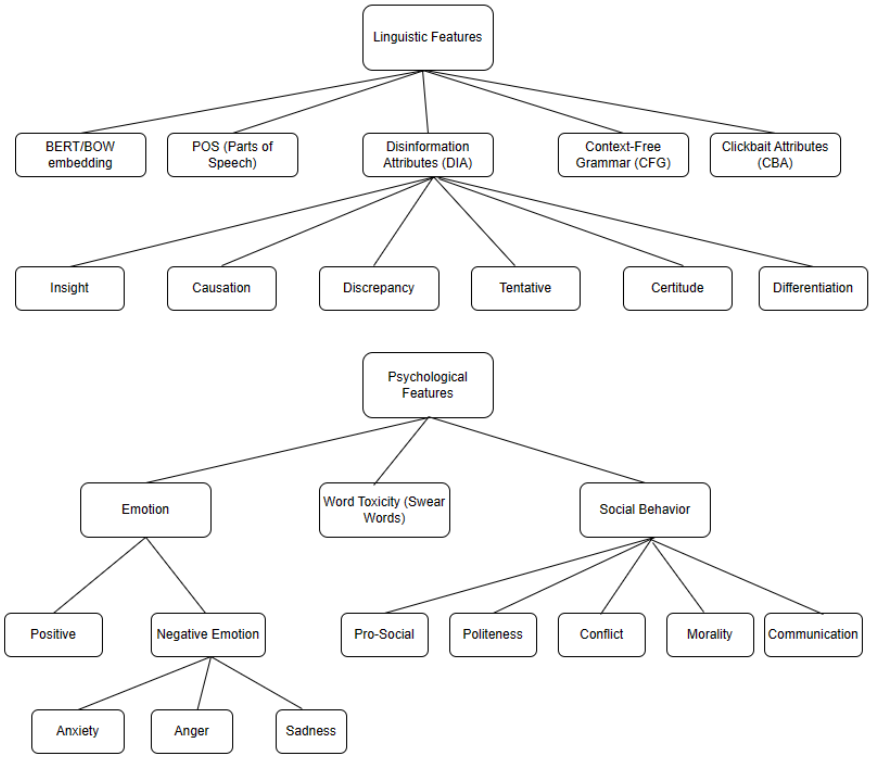

Fig. 2: Feature Breakdown

disinformation-related attributes (DIA) and clickbait-related attributes (CBA) feature groups is extracted from the input text using the TF-IDF scores of the terms.

These linguistic features provide valuable details about the article's writing style. DIA provides information about the deceptive, dis-informative content in the text. Clickbait-related attributes (CBA) are detected based on the content of the news headline. For example, phrases such as "this will blow your mind" or "can change your life", are often seen in fake news. We use a dictionary of common clickbait headlines containing forty-seven different clickbait feature groups for CBA detection. By comparing the headline of the news article against the dictionary, if the headline contains any of these clickbait titles, it is assigned a score of one; otherwise, a score of zero is assigned. DIA attributes include features related to insight (knowledge or thoughts), causation (explanatory terms), discrepancy (terms that provide reasoning), tentative (terms defining potential or conditions), certitude (terms defining actuality), and differentiation (terms comparing variance). The headline and textual content of each article are used as input to our data pre-processing layer. Our deep neural network requires an input of numeric features for classification; thus, the input text from the articles cannot be directly used as the input. We leverage natural language processing

(NLP) and term frequency-inverse document frequency (TF-IDF) to convert the article's content into numerical linguistic features.

3.3.2 Psychological features

Recent studies [16, 65] show that psychological features can be used to improve the performance of the fake news detection task. In this work, we consider three categories of psychological features: emotion, social behaviour, and swear words.

Our model includes positive, negative, and neutral emotional biases in our feature set. With emotional features, we can also extract tones of anxiety, sadness, and anger from an article. Several patterns have been identified with the use of emotion in the text of an article [16] when identifying fake news. Fake news articles commonly exhibit positive or negative biases, while true articles usually have a neutral emotional bias in their writing. These emotion features can further help extract traits such as toxicity and inflammatory attributes from the text[2]. The patterns and trends identified on these traits can aid in fake news detection.

Social behavioural traits include politeness, interpersonal conflict (disputes between individuals [12]), moralization, communication (subject conveys a topic or stance [6]), and pro-social behaviour (voluntary act to help others [5]). Morality is defined as the righteousness of actions taken by a subject and weighs the outcome on a scale [11]. Most news articles have a scale of morality in which they are written, and this information is vital when determining the veracity of fake news. Our study [2] indicates that the writing of real news articles is usually neutrally just, while fake news tends to lean towards either side of the scale. Traits such as politeness, communication and pro-social behaviour can be found in words related to gratitude, while interpersonal conflict depicts hostility towards a subject [6]. These traits are valuable in fake news research, as common terms of each feature group can be used to distinguish between real and fake news.

Finally, we include swear words to improve our model's detection of fake news. Real news articles are normally written in professional, unbiased language. Thus, swear words are not commonly found in true news. By identifying swear words in an article, our model can classify fake news more accurately.

Each psychological category extracted is treated as a feature group for our model. We obtain the occurrences of each term in the related feature group and cross-reference them with the LIWC dictionary to obtain a weighted value for each occurrence of every feature.

Similar to the linguistic features, our deep neural layer cannot process these psychological features as pure text values. Thus, we leverage the LIWC package to convert the textual psychological features into numerical representations.

LIWC. We use the Linguistic Inquiry and Word Count (LIWC) dictionary [6] to obtain these psychological features from input set P. LIWC compares the occurrences of each word or phrase in the input with its dictionary using TF-IDF weights[3] and outputs a numerical transformation for each psychological feature in the input. The

[2] https://developers.perspectiveapi.com/s/about-the-api-attributes-and-languages
[3] https://www.liwc.app/help/howitworks

output is the percentage of each term in the news article that exist within the LIWC dictionary. For example, a score of 5 on the psychological feature "social behavior" for one news article means that 5% of all the terms in this article are social behavior terms that can be found in the LIWC dictionary.

3.3.3 Model description

Linear feed-forward layer. Before feature concatenation, we include a feed-forward linear layer using psychological features as the input. Using a linear layer emphasizes the weights of psychological features common in fake news. This layer allows the proposed model to accurately identify psychological features unique to fake news from an article's text. The output is a numeric embedding of the psychological features.

Feature concatenation. After data pre-processing, our model returns a numeric representation of the linguistic and psychological features. We include a concatenation layer to combine the resulting numeric features from our pre-processing layer into one feature set. This feature set is used as the input to our neural network layer discussed in the next section.

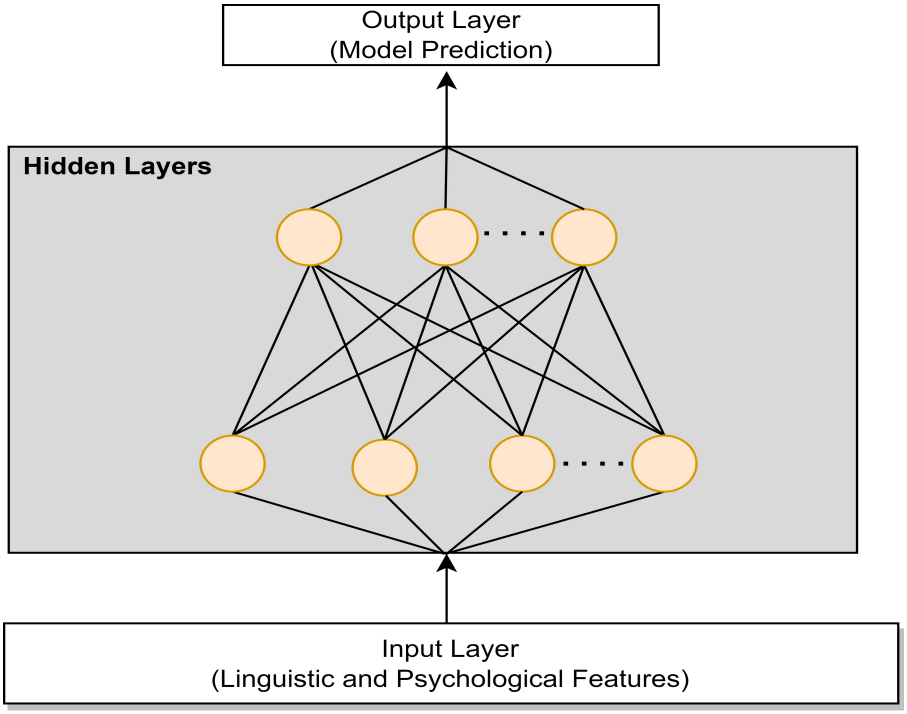

Fig. 3: Overview of Deep Neural Network (MLP)

3.3.4 Multilayer perceptron (MLP)

A standard multilayer perceptron (MLP) is used to model our deep neural network using the embedded psychological and linguistic features as the input. The architecture of our MLP classifier is presented in Figure 3. The neural model consists of one hundred layers, with ninety-eight hidden layers, one input layer, and one output layer. Both the number of hidden layers and the number of neurons in each layer are optimized using the training data. The resulting output of our neural network is the predicted classification of the input dataset N. We optimize our model using the Adam solver [35] with a ReLU activation function defined below where x is the value of the input.

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

We use the sparse categorical cross-entropy loss function defined below during the training phase:

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

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

Model training. Our model fits the training dataset into the MLP classifier. After the concatenation layer, seventy percent of our dataset is used to train the model, and the remaining thirty percent is used to test. Once our MLP classifier is initialized, it is trained using the training dataset.

4 Experiments

In this section, we present a series of experiments to showcase the efficacy of linguistic and psychological features in fake news classification and the effectiveness of the proposed model. We aim to answer the following questions through these experiments:

E1. How do linguistic and psychological features affect the performance of fake news classification?

E2. Can the proposed model leverage linguistic and psychological features to improve existing content-based classification?

E3. Which linguistic and psychological feature groups exhibit the best performance?

4.1 Experimental setup

We implement our model using NLTK [34], Keras [29] and scikit-learn [40] using a computer running on 64 GB of RAM, an eight-core sixteen thread processor and an RTX 2070 graphics card. We keep seventy percent of the dataset as the training set and use the remaining thirty percent as the test set. We then execute our model's training over one hundred epochs to reduce the training loss.

4.1.1 Datasets

We use public datasets of PolitiFact and BuzzFeed articles from the FakeNewsNet repository [53]. The datasets are collected specifically for fake news detection and are balanced between real and fake news articles. The PolitiFact and BuzzFeed datasets are collected from politifact.com and buzzfeed.com respectively. Each dataset consists of the textual news content and social media interactions from Twitter; however, we utilize only the news content features in this work, based on the research problem defined. The ground truth labels (fake or true) of news articles are provided by fact-checking experts, which guarantees the quality of news labels (fake or true). As illustrated in Figure 1, we extract the headline and text content of each article to feed into our proposed model to perform our content-based fake news classification. A breakdown of the FakeNewsNet dataset is illustrated in Table 2.

Table 2: Breakdown of FakeNewNet datasets

Dataset	Number of Real Articles	Number of Fake Articles
PolitiFact	120	120
BuzzFeed	91	91

In addition to these two small-sized datasets, we also perform a comparative study with a few selected baselines on a bigger dataset (with 2929 news articles) - PL-NCC [2]. This dataset merges the news article related information such as news headline, article text, as well as article meta-data from NELA-GT [15] and user comments from Fakeddit [47]. In this experiment, we only use the article part without including the comments for a fair comparison. The fake/non-fake ratio in this dataset is 70:30.

4.1.2 Experiment design

There are mainly two goals in our experiment. The first one is to test the effectiveness of the proposed model by comparing it with some baseline classification models. The second goal is to observe the effect linguistic and psychological features have on fake news classification. Since linguistic features have been widely used in previous models, here we mainly focus on testing psychological features. To understand these effects, we perform a series of experiments using different feature sets. Additionally, we perform two ablation studies to observe the performance and validate the necessity of different components in the proposed model.

- Our first set of experiments examine the effectiveness of the proposed model by comparing the model's performance with a few chosen baselines. By performing baseline comparisons, we can observe the impact psychological features have on fake news detection.
- Secondly, we test how different groups of psychological features impact the performance of the proposed model when used individually or paired with other linguistic and psychological feature groups. For this set of experiments, we remove one or more psychological feature groups from the complete feature set and observe its results.

- Thirdly, we perform a detailed ablation study on different components of the model to the necessity of including them in the proposed model.
  - Our first study observes how the model's performance is affected when a linear layer is introduced on top of psychological features. This experiment identifies if the linear layer is helpful for our fake news classification task.
  - Due to its prevalence in recent fake news research, we observe the performance using BERT-based representation in the proposed model over BOW-based text representation to identify which feature representation provides the best performance for our model.
- Finally, we fine-tune different hyperparameters of the proposed model and identify the parameters which have a noticeable impact on performance.

4.1.3 Evaluation metrics

We treat the task of fake news detection as a binary classification problem, where the detection result is either fake or real news. Based on commonly used practice in this line of research [57, 62], we assess the performance of our proposed model using accuracy and F1-score evaluation metrics. The information about actual and predicted class labels is determined by the confusion matrix as shown in Table 3.

Table 3: Confusion matrix

	Actual Fake	Actual Real
Predicted Fake	TP	FP
Predicted Real	FN	TN

Here, true positive (TP) refers to predicted fake news samples that are actually fake. False positive (FP) indicates that predicted fake news samples are indeed real. False negative (FN) means that predicted real news samples are indeed fake. True negatives (TN) indicate that predicted real news samples are actually real. For the F1 score (f) and accuracy, we perform the specific calculation as:

$$f = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{3}$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

4.1.4 Hyperparameters

In our experiments, we use a batch size of twenty-five. We keep the same batch size of twenty-five during the training process. The number of train epochs is one hundred. The model hyperparameters are shown in Table 4.

Table 4: Hyperparameters used for proposed model

Hyperparameter	Optimal value
Vocabulary size	12,072
Dimensionality size	50
Feed-forward layer dimensionality	64
Activation Function	relu
Number of labels	2
Batch size	25
Epochs	100
Learning rate	0.001
L2 Regularization	0.0001
Dropout	0.1
Optimizer	Adam
Loss function	sparse categorical cross entropy
Output layer	softmax

4.2 Model performance

4.2.1 Baselines

Table 5: Comparison of baseline models (complete model)

	PolitiFact		**BuzzFeed**	
Model	**Accuracy**	**F1 Score**	**Accuracy**	**F1 Score**
XGBoost	0.8010	0.8110	0.8181	0.8837
MLP	0.8130	0.8060	0.8194	0.8395
CNN	0.8472	0.8421	0.8333	0.8604
RNN	0.8054	0.8343	0.7993	0.8043
LSTM	0.6882	0.8154	0.6882	0.8154
FNDNet	0.8750	0.8890	0.8333	0.8604
Our Model	**0.9028**	**0.9091**	**0.8649**	**0.8889**

A. Overview of baseline models

Since the linguistic features we used in our model are mostly from Reza et al. [67], their model is used as one baseline. In their work, they have tested both XGBoost and Random Forest as the classification model. Since the performance from the two models are similar, in our paper, we only show the results from XGBoost (we re-implemented the model with all the linguistic features as the input). In many of the recent fake news detection models [14, 23, 25, 27, 48, 60], a neural representation (embedding) of the news article is used as the model input and a deep neural network is then used as the classification model. So here we choose four commonly used DNN models for fake new detection - CNN, MLP, RNN, and LSTM as our other baseline models, together with BERT embedding as the article representations. Additionally, we compare the results of our model against a state-of-the-art fake news classification model FNDNet [26].

XGBoost. XGBoost [8] is one of the leading traditional machine learning models used for classification problems. It is a scalable decision tree-based model.

MLP. MLP [46] models are feed-forward neural networks, which are ideal for classification problems where the input is labelled.

CNN. CNN [38] is a deep learning model which has become a popular choice for classification problems due to its capability of extracting features dynamically based on the input provided.

RNN. Recurrent neural networks (RNN), commonly used in language learning tasks, are trained models in series which leverage sequential data characteristics in the input text for the classification task.

LSTM. Long short-term memory (LSTM) models are used in a wide array of applications and are modified forms of RNN models used for longer-term memory retention for improved classification performance.

FNDNet. FNDNet [26] is a CNN-based model developed and trained using a GloVe word embedding model. The FNDNet model aims to automatically extract unique features, which are beneficial for identifying fake news.

BERT. BERT[4] is a bidirectional deep learning model which uses masking techniques to perform predictions on masked words. BERT embedding has gained popularity for text classification as of recent [27, 60] and is used in fake news classification models.

B. Performance comparison

In our first experiment, we compare the results from our model and from the baselines on both PolitiFact and BuzzFeed datasets. When implementing the baselines, we optimized the hyperparameters for each baseline separately and chose ones with best performance in the final comparison.

The proposed model outperforms all the baseline models as illustrated in Table 5. Compared to the best-performing baseline model FNDNet, it achieves a 2.8% increase in terms of accuracy and a 2.2% increase in terms of F1 score on PolitiFact dataset, 3.8% and 3.3% increase on BuzzFeed dataset. The original CNN model achieves the second-best result among the baselines, and LSTM has the worst performance among all models. We conclude that the proposed model performs better than the baseline models because we include both linguistic and psychological features as well as running the psychological features through a linear feed-forward layer, as illustrated in Section 4.2.3.

Since we observe similar patterns on both datasets, we only report the results from the PolitiFact dataset for the remaining experiments.

4.2.2 Effectiveness of linguistic and psychological features

We perform a detailed study on the effect individual psychological and linguistic features have on fake news classification. Our goal with this set of experiments is

[4] https://huggingface.co/blog/bert-101

to identify the impact psychological features have on fake news classification when paired with linguistic features. We begin our experiments with a complete model of features, including all linguistic and psychological feature groups. Each subsequent set of experiments removes different psychological feature groups from the feature set to analyze the resulting impact on performance. Within each set of experiments, we illustrate the performance of the psychological feature set using different permutations of linguistic features. The results of the experiments allow us to interpret the effectiveness of each psychological feature group used, although we have to emphasize that the finding we report here is based on the dataset we use in this research and more generalized conclusion can only be reached after we test on more datasets.

Experiment one. We illustrate the performance of our complete model in Table 6. The proposed model exhibits the best performance when using the complete feature set, which includes all linguistic and psychological features.

Table 6: Effectiveness of complete model

Note: Bold text indicates best performing feature set.

Feature groups	Accuracy	F1 score
BOW + POS + CFG + DIA + CBA + E + SW + SB	**0.9028**	**0.9091**

Legend: bag of words (BOW), parts of speech (POS), context-free grammar (CFG), disinformation-related attributes (DIA), clickbait related attributes (CBA), emotions (E), swear words (SW), social behaviour (SB)

Experiment two. From our complete model, we remove one psychological feature group to determine its effect in fake news classification, as illustrated in Table 7.

Our best performing models include the feature sets BOW, emotion, and social behaviour, as well as BOW, disinformation-related attributes, clickbait-related attributes, emotion, and social behaviour. These feature sets achieve an accuracy of eighty-nine percent. Similar feature sets with social behaviour and BOW removed show poorer performance than the better performing feature set, with up to a ten percent decrease in performance. Thus, we can infer that social behavioural features and BOW have positive effects in fake news classification.

Table 7: Effectiveness of linguistic and two psychological features

Note: Bold text indicates best performing feature set.

Feature groups	Accuracy	F1 Score
BOW + E + SW	0.8125	0.8421
POS + E + SW	0.7708	0.8136
CFG + E + SW	0.7292	0.7719
DIA + E + SW	0.8750	0.8889

CBA + E + SW	0.5405	0.4138
BOW + POS + CFG + E + SW	0.8125	0.8421
BOW + DIA + CBA + E + SW	0.8333	0.8621
POS + CFG + DIA + CBA + E + SW	0.7500	0.7857
BOW + POS + CFG + DIA + CBA + E + SW	0.8333	0.8621
BOW + E + SB	**0.8958**	**0.9123**
POS + E + SB	0.7917	0.8214
CFG + E + SB	0.7708	0.8070
DIA + E + SB	0.6667	0.6364
CBA + E + SB	0.5833	0.5238
BOW + POS + CFG + E + SB	0.8125	0.8421
BOW + DIA + CBA + E + SB	**0.8958**	**0.9123**
POS + CFG + DIA + CBA + E + SB	0.7917	0.8276
BOW + POS + CFG + DIA + CBA + E + SB	0.8542	0.8727
BOW + SW + SB	0.8125	0.8421
POS + SW + SB	0.7708	0.8254
CFG + SW + SB	0.8125	0.8421
DIA + SW + SB	0.8333	0.8667
CBA + SW + SB	0.5833	0.6154
BOW + POS + CFG + SW + SB	0.8750	0.8966
BOW + DIA + CBA + SW + SB	0.8125	0.8364
POS + CFG + DIA + CBA + SW + SB	0.8333	0.8621
BOW + POS + CFG + DIA + CBA + SW + SB	0.8750	0.8929

Legend: bag of words (BOW), parts of speech (POS), context-free grammar (CFG), disinformation-related attributes (DIA), clickbait related attributes (CBA), emotions (E), swear words (SW), social behaviour (SB)

Experiment three. We then compare the impact of each individual psychological feature group by removing two psychological feature groups from the complete model. Table 8 shows our best performing feature set from this experiment, which includes all linguistic features and swear words only.

Although our best performing model including swear words performs better than the comparable social behaviour and emotion feature groups, there is higher variability in the results when using different permutations of linguistic features. Feature sets including swear words have an average performance of roughly seventy-five percent. However, feature groups including social behaviour perform consistently better than swear words and emotion, with an average performance of eighty percent. We identify the variability in performance to correlate to the number of individual features used in each psychological feature group. Swear words have much fewer individual features compared to the social behaviour feature group, which has the most number of features comparatively. By introducing more features into the feature set, the model has more information to train with, allowing the model to perform more consistently.

Table 8: Effectiveness of linguistic and one psychological features

Note: Bold text indicates best performing feature set.

Feature groups	Accuracy	F1 Score
BOW + E	0.8125	0.8421
POS + E	0.7292	0.7636
CFG + E	0.6667	0.6800
DIA + E	0.8542	0.8814
CBA + E	0.5676	0.4667
BOW + POS + CFG + E	0.8120	0.8421
BOW + DIA + CBA + E	0.8333	0.8571
POS + CFG + DIA + CBA + E	0.7917	0.8214
BOW + POS + CFG + DIA + CBA + E	0.8750	0.8929
BOW + SW	0.8125	0.8421
POS + SW	0.7292	0.7636
CFG + SW	0.6875	0.6809
DIA + SW	0.8542	0.8679
CBA + SW	0.3956	0.3704
BOW + POS + CFG + SW	0.8542	0.8772
BOW + DIA + CBA + SW	0.7708	0.8070
POS + CFG + DIA + CBA + SW	0.7292	0.7636
BOW + POS + CFG + DIA + CBA + SW	**0.8958**	**0.9123**
BOW + SB	0.8542	0.8772
POS + SB	0.8125	0.8475
CFG + SB	0.6667	0.6800
DIA + SB	0.8333	0.8462
CBA + SB	0.7083	0.7308
BOW + POS + CFG + SB	0.8333	0.8571
BOW + DIA + CBA + SB	0.8333	0.8571
POS + CFG + DIA + CBA + SB	0.7083	0.7308
BOW + POS + CFG + DIA + CBA + SB	0.8125	0.8302

Legend: bag of words (BOW), parts of speech (POS), context-free grammar (CFG), disinformation-related attributes (DIA), clickbait related attributes (CBA), emotions (E), swear words (SW), social behaviour (SB)

Experiment four. Experiment four is conducted to obtain a baseline for our previous experiments. The results of this experiment, as illustrated in Table 9, exclude all psychological features from the feature sets, and only showcase the performance of the proposed model using linguistic features. We can compare the results of this experiment with previous experiments to analyze how the inclusion of psychological features affects the performance of fake news classification.

Similar to previous experiments, feature sets including BOW exhibit increased performance from the model. Our best performing feature set includes all linguistic

features in the feature group, obtaining an accuracy of eighty-five percent and an F1 score of eighty-seven percent.

We also note that disinformation and clickbait-related attributes perform poorly when used alone during classification. However, when we pair disinformation-related attributes with other linguistic features, we see an improvement in performance. Unlike BOW, POS, and CFG, fewer features are extracted from the text for disinformation and clickbait-related attributes. Linguistic features such as BOW, POS, and CFG can have hundreds of numeric embedding for each term in the text, while disinformation and clickbait-related attributes have six and forty-seven main feature groups respectively. When these attributes are used in isolation, the proposed model cannot accurately identify individual characteristics that are unique to fake news and noise is produced during training. This leads to poorer performance. However, when disinformation-related attributes are included with other linguistic features, the model can utilize unique features of disinformation-related attributes common in fake news articles to supplement the trained linguistic features. Doing so allows the proposed model to identify fake news more accurately.

Table 9: Effectiveness using linguistic features only

Note: Bold text indicates best performing feature set.

Feature groups	Accuracy	F1 Score
BOW	0.8125	0.8421
POS	0.7917	0.8276
CFG	0.6875	0.7059
DIA	0.7500	0.7391
CBA	0.5556	0.4068
DIA + CBA	0.7292	0.7234
BOW + POS	0.8125	0.8421
BOW + CFG	0.8125	0.8421
BOW + POS + CFG	0.8125	0.8421
BOW + DIA + CBA	0.8333	0.8571
POS + CFG + DIA + CBA	0.7708	0.8070
BOW + POS + CFG + DIA + CBA	**0.8542**	**0.8727**

Legend: bag of words (BOW), parts of speech (POS), context-free grammar (CFG), disinformation-related attributes (DIA), clickbait related attributes (CBA)

Experiment five. Finally, we perform a fifth experiment to determine the impact each psychological feature group has on fake news classification without including linguistic features. Results shown in Table 10 indicate very poor performance for these experiments; however, we continue to see the social behaviour feature group performs better than emotions and swear words.

Table 10: Effectiveness using psychological feature groups only

Note: Bold text indicates best performing feature set.

Feature groups	Accuracy	F1 Score
Emotions	0.4583	0.2353
Swear words	0.3958	0.2353
Social behaviour	0.5625	0.5532
Swear words	0.7500	0.7391
Emotions + Social behaviour	0.6875	0.7059
Swear words + Social behaviour	0.6458	0.6909
Emotions + Swear words + Social behaviour	**0.7708**	**0.7755**

4.2.3 Effectiveness of linear layer

The proposed model introduces a feed-forward linear layer to refine and enhance the efficacy of psychological features during classification. To analyze the effectiveness of the linear layer, we execute our model with and without a linear layer using the PolitiFact dataset and a complete feature set. The results of this experiment is presented in Table 11. The standard MLP model without a linear layer performs similarly to the baseline MLP models presented in Table 5; however, the introduction of the linear layer improves the model's performance by seven percent.

The linear layer amplifies the weight of psychological features frequent in fake news, such as swear words and pro-social behaviour, while reducing the weight of psychological features common in true news. By focusing on psychological traits common in fake news, our proposed model can accurately identify fake news better than baseline MLP models.

Table 11: Effectiveness of linear layer

	(Complete Model)	
Has Linear Layer	Accuracy	F1 Score
Yes	**0.9028**	**0.9091**
No	0.8333	0.8421

4.2.4 Effectiveness of neural representation with linguistic features

Recent studies in the field of fake news research [13, 60] have focused on deep learning based methods to improve the performance of fake news classification; however, studies such as Dacrema et. al [9] have proven simpler machine learning or deep learning models can be just as effective or better than more complicated models. We perform an in-depth study on the effects BOW and BERT embeddings have on

fake news classification. The results of our experiments, as indicated by Table 12, coincide with Dacrema et al's research finding.

Table 12: Effectiveness of neural representation with linguistic features (BOW/SBERT)

Note: Bold text indicates best performing feature set.

Neural representation	**BOW**		**BERT**	
Additional features	**Accuracy**	**F1 Score**	**Accuracy**	**F1 Score**
-	0.8125	0.8421	0.7083	0.7342
POS + CFG	0.8125	0.8421	0.7222	0.7436
DIA + CBA	0.8333	0.8571	0.7500	0.7750
E	0.8125	0.8421	0.6806	0.7013
POS + CFG + E	0.8120	0.8421	0.7222	0.7436
DIA + CBA + E	0.8333	0.8571	0.7222	0.7500
SW	0.8125	0.8421	0.6806	0.7013
POS + CFG + SW	0.8542	0.8772	0.7222	0.7436
DIA + CBA + SW	0.7708	0.8070	0.6806	0.7013
SB	0.8542	0.8772	0.6806	0.7013
POS + CFG + SB	0.8333	0.8571	0.7778	0.8000
DIA + CBA + SB	0.8333	0.8571	0.7083	0.7342
E + SW	0.8125	0.8421	0.6806	0.7013
POS + CFG + E + SW	0.8125	0.8421	0.7222	0.7436
DIA + CBA + E + SW	0.8333	0.8621	0.7917	0.8148
E + SB	0.8958	0.9123	0.7917	0.8235
POS + CFG + E + SB	0.8125	0.8421	0.7778	0.8049
DIA + CBA + E + SB	0.8958	0.9123	0.7361	0.7595
SW + SB	0.8125	0.8421	0.6806	0.7013
POS + CFG + SW + SB	0.8750	0.8966	0.7639	0.7848
DIA + CBA + SW + SB	0.8125	0.8364	0.7083	0.7342
All features	**0.9028**	**0.9091**	**0.8333**	**0.8462**

Legend: bag of words (BOW), parts of speech (POS), context-free grammar (CFG), disinformation-related attributes (DIA), clickbait related attributes (CBA), emotions (E), swear words (SW), social behaviour (SB)

Although BERT produces numerical embedding like BOW, BERT performs its own predictions on the article's text to identify masked words, which it uses to generate embeddings. Since BERT is a large language model with many parameters, its performance on the smaller datasets in our experiments were not as good as expected.

We suspect this is due to overfitting. Unlike BERT embeddings, BOW generates embedding values based on the number of occurrences of words in the text, and does not perform internal classification. This allows BOW to be used for a small dataset like PolitiFact more reliably compared to BERT embedding. Thus, the proposed model performs better classification using BOW rather than BERT embedding.

4.2.5 Sensitivity analysis

The hyperparameters used in the proposed model are fine-tuned to obtain the best performance of the model. We test various configurations of different hyperparameters and illustrate the effects of only the most impactful hyperparameters in Tables 13 and 14. We tested other hyperparameter configurations; however, we do not see a significant impact on performance. We performed the experiments on both datasets and got similar results. Here, we only report the results from PolitiFact and keep all other hyperparameters constant.

Table 13: Effectiveness of batch size

Batch Size	Accuracy	F1 Score	Runtime (seconds)
5	0.8333	0.8462	25.3233
15	0.8472	0.8608	11.8215
25	**0.9028**	**0.9091**	**8.0181**
50	0.8611	0.8684	6.1011
75	0.8333	0.8462	5.8151
150	0.8194	0.8312	4.6675

Note: Higher accuracy and F1 score is better. Lower runtime is better.

Our best-performing model has a batch size of twenty-five, which equates to roughly ten percent of the total dataset. Through experimentation, larger batch sizes allow the model to converge faster, thus reducing runtime, but consume more memory as a result. In contrast, a smaller batch size takes longer to converge but requires less memory when training. It is imperative that a good batch size is chosen to balance an efficient runtime and usage of resources.

Table 14: Effectiveness of dropout

Dropout	Accuracy	F1 Score
0.01	0.8730	0.8770
0.1	**0.9028**	**0.9091**
0.3	0.8333	0.8500
0.5	0.8056	0.8205
0.8	0.8056	0.8205

Similar to the previous experiment, it is important to fine-tune thedropout rate to avoid the loss of smaller details in our input data. Smaller dropout values exhibit less uncertainty in the dataset compared to larger values [59]. When using larger dropout values, minute details from the input data are lost during training. Throughout our analysis, we identify a smaller dropout value of ten percent to yield the most accurate classification for the proposed model, ensuring details in the data are not lost.

4.2.6 Comparative study on PL-NCC dataset

To test our model on a bigger dataset, we choose PL-NCC (news article part), which has close to 3000 news articles. To do the comparison, we use a select subset of baselines, including XGBoost, MLP and CNN. The results are shown in Table 15. From the table, we can see that our model also performs the best.

Table 15: Fake news detection results on PL-NCC dataset

Model	Accuracy	F1 Score
XGBoost	0.952	0.966
MLP	0.951	0.965
CNN	0.948	0.965
Our model	**0.954**	**0.967**

4.3 Discussion of results

In this section, we analyze the results of the conducted experiments to address the discussion questions presented in Section 4. The experiments conducted provide valuable insight into the impact linguistic and psychological features have on fake news detection. Although psychological features perform poorly when used alone, the features complement well when used in conjunction with other linguistic features. Psychological features have unique characteristics which allow fake news detection models to classify fake news more accurately. Our experiments indicate psychological features positively affect the performance of text-based fake news classification. Including a linear layer allows the proposed model to leverage the benefits of psychological features in fake news classification, allowing our model to perform better than baselines. Finally, social behaviour is favourable when identifying fake news as it provides a noticeable improvement in performance in all experiments conducted.

One limitation of our research is the size of the dataset used for our experiments. We use a sample of the FakeNewsNet repository; however, the datasets available in this repository are smaller than the data available in existing news outlets. This limits the experiments conducted in our study, though it is not a limitation of the model itself. Also, because of the small size of the dataset, we only have divided the dataset into training and test sets, without having a validation set for better hyperparameter tuning. The second limitation is that the proposed model requires a large training sample to perform its news classification and is hindered by newer forms of fake

news implementations. Thus, the model requires frequent training using new articles. Thirdly, there is a level of uncertainty in the model performance. It is possible that with entirely new data (in terms of the size, format, language, writing style, etc.) the model may perform differently. Our reported results can be treated as upper bounds on the performance. Lastly, although the dataset we used from Fakenewsnet [53] provides the article-level labels chosen by human experts, there is a risk that the manual annotation process may not yield the accurate class labels for some news articles.

For future work, we plan to extend our study to investigate other linguistic and psychological features which can provide improved performance to existing fake news classification models. Secondly, we look to incorporate social user engagement in our research to improve the efficacy of our model with newer forms of fake news classification. User feedback is a great tool for fake news classification, as it provides valuable information on the veracity of news events from the readers' perspective [26] and can be adapted to perform early fake news detection. Thirdly, we could try to fine-tune the BERT model on a fake news dataset to see whether it improves the performance. Currently, the BERT embedding performs worse than the BOW embedding, with fine-tuning, maybe we can see different results. Fourthly, we would like to investigate whether the performance of the model is affected by the article length. If we have a dataset with long news articles (e.g., regular news articles published by a newspaper), or a dataset with short articles (e.g., short news updates on a social media site), will we observe the similar performance? Finally, we will test our model on more datasets, such as Covid19 [39], ReCOVery [68], etc., and we will also test some recent large language models to see whether they can capture the psycho-linguistic features implicitly or whether it is still necessary to explicitly extract and include these features in the detection model.

5 Conclusion

Fake news detection has been a growing field of research over recent years due to the prominence of social media. It is imperative for social media outlets to incorporate detection systems to limit the exposure of fake news to its users.

In this paper, we make contributions to the field of fake news research by: 1) leveraging linguistic and psychological characteristics from articles' text to perform content-based fake news classification; and 2) performing extensive meta-analysis on state-of-the-art classification models using different permutations of linguistic and psychological feature groups to report the best configuration for fake news classification. We propose a style-based fake news detection system which uses deep neural networks, linguistic and psychological features to perform content-based fake news classification. We execute several experiments to demonstrate the effectiveness of our proposed model and to showcase the efficacy psychological features exhibit in fake news research when paired with linguistic features.

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

References

[1] Alghamdi, J., Lin, Y. and Luo, S. (2023). Towards COVID-19 fake news detection using transformer-based models. Knowledge-Based Systems, 274, p.110642.

[2] Arunthavachelvan, K., Raza, S., and Ding, C. (2023) PLNCC: Leveraging new data features for enhanced accuracy of fake news detection. In Proceedings of the International Conference on Advances in Social Networks Analysis and Mining (pp. 144-148).

[3] Balcar, S., Skrhak, V., and Peska, L. (2022). Rank-sensitive proportional aggregations in dynamic recommendation scenarios. User Modeling and User-Adapted Interaction, 1-62.

[4] Baly, R., Karadzhov, G., Alexandrov, D., Glass, J., and Nakov, P. (2018). Predicting factuality of reporting and bias of news media sources. arXiv preprint arXiv:1810.01765.

[5] Bierhoff, H. W. (2002). Prosocial behaviour. Psychology Press.

[6] Boyd, R. L., Ashokkumar, A., Seraj, S., and Pennebaker, J. W. (2022). The development and psychometric properties of LIWC-22. Austin, TX: University of Texas at Austin.

[7] Cheng, L., Guo, R., Shu, K., and Liu, H. (2021). Causal understanding of fake news dissemination on social media. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 148-157).

[8] Chen, T., and Guestrin, C. (2016). XGboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794).

[9] Dacrema, F.M., Cremonesi, P., and Jannach, D. (2019). Are we really making much progress? A worrying analysis of recent neural recommendation approaches. In Proceedings of the 13th ACM Conference on Recommender Systems (pp. 101-109).

[10] Devlin, J., Chang, M. W., Lee, K., and Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[11] Effron, D. A., and Raj, M. (2020). Misinformation and morality: Encountering fake-news headlines makes them seem less unethical to publish and share. Psychological Science, 31(1), 75-87.

[12] Emerson, R. M. (2015). Everyday troubles: The micro-politics of interpersonal conflict. University of Chicago Press.

[13] Farokhian, M., Rafe, V., and Veisi, H. (2022). Fake news detection using parallel BERT deep neural networks. arXiv preprint arXiv:2204.04793.

[14] Farzad, A., Mashayekhi, H. & Hassanpour, H. A comparative performance analysis of different activation functions in LSTM networks for classification. *Neural Computing And Applications*. **31** pp. 2507-2521 (2019)

[15] Gruppi, M., Horne, B.D., Adalı, S. (2023). NELA-GT-2022: A large multi-labelled news dataset for the study of misinformation in news articles. arXiv preprint arXiv:2203.05659.

[16] Guo, C., Cao, J., Zhang, X., Shu, K., and Yu, M. (2019). Exploiting emotions for fake news detection on social media. arXiv preprint arXiv:1903.01728.

[17] Horne, B. D., and Adali, S. (2017). 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 Proceedings of the Eleventh International AAAI Conference on Web and Social Media, Workshop on News and Public Opinion (pp. 759-766).

[18] Horne, B. D., Dron, W., Khedr, S., and Adali, S. (2018). Assessing the news landscape: A multi-module toolkit for evaluating the credibility of news. In Companion Proceedings of the The Web Conference 2018 (pp. 235-238).

[19] Horne, B. D., Nørregaard, J., and Adali, S. (2019). Robust fake news detection over time and attack. ACM Transactions on Intelligent Systems and Technology (TIST), 11(1), 1-23.

[20] Hu, L., Yang, T., Zhang, L., Zhong, W., Tang, D., Shi, C., Duan, N., and Zhou, M. (2021). Compare to the knowledge: Graph neural fake news detection with external knowledge. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (pp. 754-763).

[21] Hu, B., Sheng, Q., Cao, J., Shi, Y., Li, Y., Wang, D., and Qi, P. (2024). Bad actor, good advisor: Exploring the role of large language models in fake news detection. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 38, No. 20, pp. 22105-22113).

[22] Huang, Q., Zhou, C., Wu, J., Liu, L., and Wang, B. (2020). Deep spatial–temporal structure learning for rumor detection on Twitter. Neural Computing and Applications, 1-11.

[23] Jehad, R., and Yousif, S. A. (2021). Classification of fake news using multi-layer perceptron. In AIP Conference Proceedings (Vol. 2334, No. 1, p. 070004). AIP Publishing LLC.

[24] Jiang, S., Chen, X., Zhang, L., Chen, S., and Liu, H. (2019). User-characteristic enhanced model for fake news detection in social media. In CCF International Conference on Natural Language Processing and Chinese Computing (pp. 634-646). Springer, Cham.

[25] Jwa, H., Oh, D., Park, K., Kang, J. M., and Lim, H. (2019). exbake: Automatic fake news detection model based on bidirectional encoder representations from transformers (bert). Applied Sciences, 9(19), 4062.

[26] Kaliyar, R. K., Goswami, A., Narang, P., and Sinha, S. (2020). FNDNet–a deep convolutional neural network for fake news detection. Cognitive Systems Research, 61, 32-44.

[27] Kaliyar, R. K., Goswami, A., and Narang, P. (2021). FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. Multimedia tools and applications, 80(8), 11765-11788.

[28] Karimi, H., Roy, P., Saba-Sadiya, S., and Tang, J. (2018). Multi-source multi-class fake news detection. In Proceedings of the 27th International Conference on Computational Linguistics (pp. 1546-1557).

[29] Ketkar, N. (2017). Introduction to keras. In Deep learning with Python (pp. 97-111). Apress, Berkeley, CA.

[30] Kumari, R., Ashok, N., Ghosal, T., and Ekbal, A. (2021). Misinformation detection using multitask learning with mutual learning for novelty detection and emotion recognition. Information Processing and Management, 58(5), p.102631.

[31] Liu, C., Wu, X., Yu, M., Li, G., Jiang, J., Huang, W., and Lu, X. (2019). A two-stage model based on BERT for short fake news detection. In International Conference on Knowledge Science, Engineering and Management (pp. 172-183). Springer, Cham.

[32] Liu, Y., and Wu, Y. F. (2018). Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 32, No. 1).

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

[34] Loper, E., and Bird, S. (2002). NLTK: The natural language toolkit. arXiv preprint cs/0205028.

[35] Loshchilov, I., and Hutter, F. (2017). Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101.

[36] Mosallanezhad, A., Karami, M., Shu, K., Mancenido, M. V., and Liu, H. (2022). Domain adaptive fake news detection via reinforcement learning. In Proceedings of the ACM Web Conference 2022 (pp. 3632–3640).

[37] Nguyen, V. H., Sugiyama, K., Nakov, P., and Kan, M. Y. (2020). Fang: Leveraging social context for fake news detection using graph representation. In Proceedings of the 29th ACM International Conference on Information and Knowledge Management (pp. 1165-1174).

[38] O'Shea, K., and Nash, R. (2015). An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458.

[39] Patwa, P., Sharma, S., Pykl, S., Guptha, V., Kumari, G., Akhtar, M.S., Ekbal, A., Das, A., and Chakraborty, T. (2021). Fighting an infodemic: Covid-19 fake news dataset. In Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021, Collocated with AAAI 2021, (pp. 21-29). Springer.

[40] Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. the Journal of Machine Learning Research, 12, 2825-2830.

[41] Phan, H.T., Nguyen, N.T., and Hwang, D. (2023). Fake news detection: A survey of graph neural network methods. Applied Soft Computing, p.110235.

[42] Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., and Stein, B. (2017). A stylometric inquiry into hyperpartisan and fake news. arXiv preprint arXiv:1702.05638.

[43] Przybyla, P. (2020, April). Capturing the style of fake news. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 01, pp. 490-497).

[44] Qian, F., Gong, C., Sharma, K., and Liu, Y. (2018, July). Neural User Response Generator: Fake News Detection with Collective User Intelligence. In IJCAI (Vol. 18, pp. 3834-3840).

[45] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., and Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI blog, 1(8), 9.

[46] Ramchoun, H., Ghanou, Y., Ettaouil, M., and Janati Idrissi, M. A. (2016). Multilayer perceptron: Architecture optimization and training. International Journal of Interactive Multimedia and Artificial Intelligence, 4(1); 26-30.

[47] Nakamura, K., Levy, S., and Wang, W.Y. (2019). r/fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection. arXiv preprint arXiv:1911.03854.

[48] Raza, S., and Ding, C. (2022). Fake news detection based on news content and social contexts: a transformer-based approach. International Journal of Data Science and Analytics, 13(4), 335-362.

[49] Reimers, N., and Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084.

[50] Schmitt, M. F., and Spinosa, E. J. (2022). Scalable stream-based recommendations with random walks on incremental graph of sequential interactions with implicit feedback. User Modeling and User-Adapted Interaction, 1-31.

[51] Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena*. **404** pp. 132306 (2020)

[52] Shu, K., Bernard, H. R., and Liu, H. (2019). Studying fake news via network analysis: detection and mitigation. In Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining (pp. 43-65). Springer, Cham.

[53] Shu, K., Mahudeswaran, D., Wang, S., Lee, D., and Liu, H. (2020). Fakenewsnet: A data repository with news content, social context, and spatio-temporal information for studying fake news on social media. Big data, 8(3), 171-188.

[54] Shu, K., Sliva, A., Wang, S., Tang, J., and Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22-36.

[55] Shu, K., Wang, S., and Liu, H. (2019, January). Beyond news contents: The role of social context for fake news detection. In Proceedings of the twelfth ACM International Conference on Web Search and Data Mining (pp. 312-320).

[56] Shu, K., Zheng, G., Li, Y., Mukherjee, S., Awadallah, A. H., Ruston, S., and Liu, H. (2021). Early detection of fake news with multi-source weak social supervision. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 650-666). Springer, Cham.

[57] Shu, K., Zhou, X., Wang, S., Zafarani, R., and Liu, H. (2019, August). The role of user profiles for fake news detection. In Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (pp. 436-439).

[58] Silva, R. M., Santos, R. L., Almeida, T. A., and Pardo, T. A. (2020). Towards automatically filtering fake news in Portuguese. Expert Systems with Applications, 146, 113199.

[59] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., andSalakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 15(1), 1929-1958.

[60] Szczepański, M., Pawlicki, M., Kozik, R., and Choraś, M. (2021). New explainability method for BERT-based model in fake news detection. Scientific Reports, 11(1), 1-13.

[61] Vijjali, R., Potluri, P., Kumar, S., and Teki, S. (2020). Two stage transformer model for COVID-19 fake news detection and fact checking. arXiv preprint arXiv:2011.13253.

[62] Vo, N., and Lee, K. (2019, July). Learning from fact-checkers: analysis and generation of fact-checking language. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 335-344).

[63] Whitehouse, C., Weyde, T., Madhyastha, P., and Komninos, N. (2022). Evaluation of fake news detection with knowledge-enhanced language models. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 16, pp. 1425-1429).

[64] Zhang, J., Dong, B., and Philip, S.Y. (2020). Fakedetector: Effective fake news detection with deep diffusive neural network. In Proceedings of the 36th IEEE International Conference on Data Engineering (ICDE) (pp. 1826-1829).

[65] Zhang, X., Cao, J., Li, X., Sheng, Q., Zhong, L., and Shu, K. (2021, April). Mining dual emotion for fake news detection. In Proceedings of the Web Conference 2021 (pp. 3465-3476).

[66] Zellers, R., Holtzman, A., Rashkin, H., Bisk, Y., Farhadi, A., Roesner, F., and Choi, Y. (2019). Defending against neural fake news. Advances in Neural Information Processing Systems, 32.

[67] Zhou, X., Jain, A., Phoha, V. V., and Zafarani, R. (2020). Fake news early detection: A theory-driven model. Digital Threats: Research and Practice, 1(2), 1-25.

[68] Zhou, X., Mulay, A., Ferrara, E., and Zafarani, R. (2020). Recovery: A multimodal repository for covid-19 news credibility research. In Proceedings of the 29th ACM International Conference on Information and Knowledge Management (pp. 3205-3212).

[69] Zhou, X., Wu, J., and Zafarani, R. (2020). Safe: similarity-aware multi-modal fake news detection (2020). Preprint. arXiv, 200304981.

[70] Zhou, X., and Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys (CSUR), 53(5), 1-40.