Multi-modal Analysis of Misleading Political News

Anu Shrestha, Francesca Spezzano(✉), and Indhumathi Gurunathan

Computer Science Department, Boise State University, Boise, ID 83725, USA
{anushrestha,indhumathigurunathan}@u.boisestate.edu,
francescaspezzano@boisestate.edu

Abstract. The internet is a valuable resource to openly share information or opinions. Unfortunately, such internet openness has also made it increasingly easy to abuse these platforms through the dissemination of misinformation. As people are generally awash in information, they can sometimes have difficulty discerning misinformation propagated on these web platforms from truthful information. They may also lean too heavily on information providers or social media platforms to curate information even though such providers do not commonly validate sources. In this paper, we focus on political news and present an analysis of misleading news according to different modalities, including news content (headline, body, and associated image) and source bias. Our findings show that hyperpartisan news sources are more likely to spread misleading stories than other sources and that it is not necessary to read news body content to assess its validity, but considering other modalities such as headlines, visual content, and publisher bias can achieve better performances.

Keywords: Misinformation detection on the web · Multi-modal content analysis · Source bias

1 Introduction

The volume of misleading news present in current media has grown in popularity in recent years through social media and online news sources. In 2017, the Pew Research Center found that 67% of American adults (ages 18+) get news from social media, which was a 5% increase since 2016 [21]. An analysis of news leading up to the 2016 election conducted by BuzzFeed, found that there was more engagement with the leading misleading news stories than real news stories [24]. News is becoming more accessible and widespread than ever before. However, information proliferation has also contributed to the spread of misleading news, which has fostered the advancement of various methods to determine the validity of news. One such method is developed upon evaluating linguistic attributes such as features determining readability and lexical information [13,19,20]. These methods often mimic that of what would generally be considered the most effective of all: reading through the news with the purpose

M. van Duijn et al. (Eds.): MISDOOM 2020, LNCS 12259, pp. 261–276, 2020.
https://doi.org/10.1007/978-3-030-61841-4_18

of evaluating their accuracy. However, with the spread of misleading news, it is unlikely, if not impossible, for everyone to spend large quantities of time reading through multiple newspapers and sources. Of course, the news sharing process occurs rapidly, necessitating effective methods to recognize signals of misleading content. In fact, reading the news body content may be time-consuming, and often people are exposed to news through their snippet on social media, where only the news headline and images are shown.[1] This trend of showing only some flimsy cuts of news with catchy headline and visuals in social media news feeds has made people share such news frequently without having deep reading and monitoring. A recent study by Gabielkov et al. [10] found evidence that the number of news shares is an inaccurate measure of actual readership. Thus, people are immersed in information across social media, which is often shared without reading and validating the content, thus leading to possible consequences of its diffusion.

In this paper, we use machine learning and multi-modal content analysis to detect misleading political news. To the best of our knowledge, we present the *first* content-based study considering the headline, body content, visual, and source bias modalities together for misleading news detection. Because the news trends continuously evolve, we analyze news text (from body and headline) by focusing on linguistic style, text complexity, and psychological aspects of the text, rather than topic-dependent representations of documents (e.g., [7]). Moreover, we consider *new features* that have not been explored before such has to capture emotions in images and the political bias of the news publisher. Our analysis, conducted on two state-of-the-art political news datasets, namely FakeNewsNet [23] and BuzzFeedNews [20], reveals that:

- News headlines are more informative than news body content, suggesting that we can avoid to "read" the news excerpt and focus on other modalities to better detect misleading news.
- By comparing news headline and excerpt content, we observe that headline characteristics are more consistent than excerpt ones across datasets (e.g., punctuation features are the most important group of features in both datasets considered), and, in general, the headline focuses more on briefly drawing the attention of the reader, while a higher number of emotional/ psychological words is more a characteristic of an excerpt than the headline, for misleading news.
- Publisher bias is a strong predictor of news validity. In fact, by analyzing information collected from mediabiasfactcheck.com ("the most comprehensive media bias resource on the Internet"), we show that hyper-partisan news sources are more likely to spread misleading stories than other sources.
- Image features improve the automatic detection of misleading news with the most important features being the ones highlighting the expressions and emotions of depicted people.

[1] There are also some browser extensions that checks the source and further add the publisher bias to the news appearing in the social media feed [1].

- It is possible to detect misleading news from its snippet (news headline, image, and source bias) more accurately than looking into the body content: AUROC 0.91 vs. 0.78 on FakeNewsNet and 0.81 vs. 0.77 on BuzzFeedNews.

Overall, this paper contributes to determining effective and explicable multi-modal factors to recognize misleading news, that can be taught to people to recognize misleading news from its snippet and possibly decrease the unconscious spread of misinformation in social media [2].

2 Related Work

To detect misleading news, many works have considered news content (headline, body, image), the social network between the users and their social engagement (share, comment, and discuss given news), or a hybrid approach that considers both [22]. Regarding misleading news detection from news content (which is the focus of our paper), Potthast et al. [20] attempted to classify news as real or fake based on its style as being part of hyperpartisan news, mainstream news, or satire. This study used a dataset composed of 1,627 articles from a Buzzfeed dataset. Features such as n-grams, stop words, parts of speech, and readability were considered in this study. Although there was higher F1-measure in determining the hyperpartisan vs. mainstream articles (0.78 F1-measure based on stylistic features and 0.74 for topic) the research was limited in deciphering between fake and real news (0.46 F1-measure for style-based features).

Horne and Adali [13] considered both news body and headline for determining the validity of news. They included three datasets: a dataset created by Buzzfeed leading to the 2016 U.S. elections, one created by the researchers containing real, fake and satire sources, and a third dataset containing real and satire articles from a previous study. Based on textual features extracted from body and headline, they found out that the content of fake and real news is drastically different as they were able to obtain a 0.71 accuracy when considering the number of nouns, lexical redundancy (TTR), word count, and the number of quotes. Further, the study found that fake titles contain different sorts of words (stop words, extremely positive words, and slang, among others) than titles of real news articles resulting in a 0.78 accuracy. Pérez-Rosas et al. [19] collected two new datasets, the FakeNewsATM dataset covering seven different news domains (education, business, sport, politics, etc.) and the Celebrity dataset regarding news on celebrities. They analyzed the news body content only and achieved an F1-measure up to 0.76 in detecting misleading content. They also tested cross-domain classification obtaining poor performances by training in one dataset and testing in the other one, but better accuracies (ranging from 0.51 to 0.91) in training on all but the test domain in the FakeNewsATM dataset.

Images in news articles also play a role in misleading news detection [3,12,14,25]. Fake images are used in news articles to provoke emotional responses from readers. Images are the most eye-catching type of content in the news; a reader can be convinced of a claim by just looking at the title of the news and the image

Table 1. Available datasets for misleading news detection.

Dataset	Size	Text	Images
BuzzFeedNews [20]	1,627	✓	
Horne and Adali DS1 [13]	71	✓	
Horne and Adali DS2 [13]	225	✓	
Pérez-Rosas et al. [19]	480	✓	
FakeNewsNet [23]	384	✓	✓

itself. So, it's crucial to include image analysis in fake news detection techniques. For instance, Jin et al. [15] showed that including visual and statistical features extracted from news images improves the results for microblogs news verification up to an F1-measure of 0.83 on a dataset collected from Sina Weibo on general news events and associated images. Wang et al. [27] proposed a deep-learning-based framework to extract features from both text and image of the tweets about news not related to specific events to detect misleading content. Results show an F1-measure ranging from 0.72 on Twitter to 0.83 on Sina Weibo.

In contrast with previous work, this paper provides a comprehensive study of four different content-based modalities to detect misleading political news. Other works have considered a single modality (e.g., either body content or images) or a subset of the modalities we considered (e.g., headline and body, or body, and image) but all these modalities together have not been investigated so far. Also, work involving image analysis [15,27] *focused on micro-blog content rather than proper news content.*

3 Datasets

In this section, we discuss the lack of a large scale misleading news dataset (especially in the political domain) and present the datasets we use in this paper, including a *new* dataset containing publisher bias and credibility we crawled from the MediaBias/FactCheck website.

Available Datasets and Limitations. There exist several datasets containing political news that have been used for fake news detection, as shown in Table 1.

Horne and Adali used two datasets in their paper [13]. The first dataset, DS1, contains 36 real news stories and 35 fake news stories, while the second one, DS2, contains 75 real, misleading, and satire news (75 for each category). The main drawback of these two datasets is that labels are assigned according to the credibility of the news source, instead of via fact-checking. However, a news source can have mixed credibility and publish both factual and misleading information. Pérez-Rosas et al. [19] collected a dataset of 480 news where 240 are fact-checked real news belonging to six different domains (sports, business, politics, etc.) and 240 are fake news collected via crowdsourcing, i.e., they asked

AMT workers to write a fake news item based on one of their real news item and by mimic journalist style (hence these are unrealistic news articles). In this paper, we use two datasets (described later in the section) to conduct our analysis, namely FakeNewsNet [23] and BuzzFeedNews [20] (the largest available dataset). FakeNewsNet is the only state-of-the-art dataset containing information beyond the news content modality and in the political domain.

As Table 1 shows, there is generally limited availability of large scale benchmarks for fake news detection as collecting labels requires fact-checking, which is a time-consuming activity. As reported in [22], other datasets have been used for related tasks, but they are not suitable for our analysis as they do not contain proper news articles. For instance, LIAR [26] contains human-labeled short statements, while CREDBANK [16] contains news events, where each event is a collection of tweets. Finally, the MediaEval Verifying Multimedia Use benchmark dataset [6] used in [27] contains images and tweets instead of news articles.

FakeNewsNet Dataset. This dataset consists of details about the news content, publisher information, and social engagement information [23]. The ground truth labels are collected from journalist experts such as Buzzfeed and the fact-checking website Politifact. The dataset is divided into two networks, Buzzfeed and Politifact, and the news contents are collected from Facebook web links. We downloaded all the available images related to the news in this dataset. The publishers' bias is retrieved from the dataset described in the next section. We merged together the news from both Politifact and Buzzfeed to have a larger dataset to work with. After cleaning the dataset from missing news bodies or headlines, we obtained a total of 384 news, 175 misleading and 209 factual.

BuzzFeedNews Dataset. It contains news regarding the 2016 U.S. election published on Facebook by nine news agencies [20]. This dataset labels 356 news articles as left-leaning and 545 as right-leaning articles, while 1264 are mostly true, 212 are a mixture of true and false, and 87 are false.

MediaBias/FactCheck Dataset. To exploit the partisan information of the news source, we crawled the website mediabiasfactcheck.com, whose main goal is to educate the public on media bias and deceptive news practices. This website contains a comprehensive list of news sources, their bias, and their credibility of factual reporting scores. Here, the publisher's political bias is defined by using seven degrees of bias: *extreme-right, right, right-centered, neutral, left-centered, left, and extreme-left*. We collected the factual reporting score of all the news sources under five categories: *Left bias* (moderately to strongly biased toward liberal causes), *Left-center* (slight to moderate liberal bias), *Least* (minimal bias), *Right-Center* (slightly to moderately conservative in bias), and *Right bias* (moderately to strongly biased toward conservative causes). The credibility score of these publishers falls into three categories: *Very high* (which means the source is always factual), *High* (which means the source is almost always factual) and

Mixed (which means the source does not always use proper sourcing or sources to other biased/mixed sources). We also collected the publisher bias under the category *Questionable Sources*, which contains extremely biased publishers, mainly doing propaganda and/or writing misleading news. The number of publishers in each category considered is reported in Fig. 1. We retrieved a total of 1,783 publishers. The relationship between the source bias and its credibility is analyzed in Sect. 4.3.

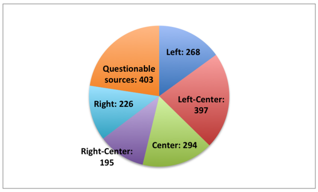

Fig. 1. Number of publishers per category in the MediaBias/FactCheck dataset.

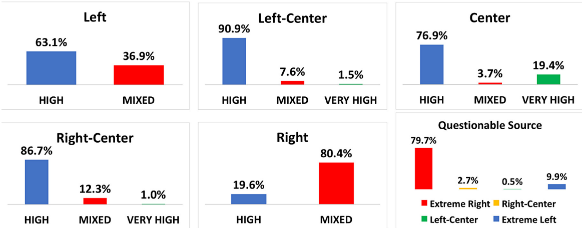

Fig. 2. Publisher credibility per bias and bias distribution within questionable sources in the MediaBias/FactCheck dataset.

4 Multi-modal Features

We now describe the set of features we used in the paper to analyze misleading political news. We consider four modalities, namely news content, and headline, images, and source bias.

4.1 Textual Features

Several approaches have been developed to extract features from text, from the widely used bag-of-words to the most recent BERT [7] deep learning-based approach. Although these approaches are popular in text analysis, they generate topic-dependent feature representation of documents that are not suitable for the dynamic environment of news where stories' topics change continuously. Therefore, in our analysis, we consider features that focus on linguistic style, text complexity, and psychological aspect to detect misleading news, such as Linguistic Inquiry and Word Count (LIWC) and text readability measures. Another approach is the Rhetorical Structure Theory (RST) which captures the writing style of documents [23]. However, as research has shown that the performance of LIWC is comparatively better than RST [23], we did not use RST in our analysis. Thus, to analyze the text of news body and headline, we consider the following groups of features (we also consider the number of stop words and upper case word count as additional features for news headline).

Linguistic Inquiry and Word Count (LIWC). LIWC is a transparent text analysis tool that counts words in psychologically meaningful categories. We use the LIWC 97 measures for analyzing the cognitive, affective, and grammatical processes in the text. To examine the difference between the factual and misleading news writing style, we divide the LIWC features into four categories [18]:

Linguistics features (28 features) refer to features that represent the functionality of text such as the average number of words per sentence and the rate of misspelling. This category of features also includes negations as well as part-of-speech (Adjective, Noun, Verb, Conjunction) frequencies.

Punctuation features (11 features) are used to dramatize or sensationalize a news story that can be analyzed through punctuation types used in the news such as Periods, Commas, Question, Exclamation, and Quotation marks, etc.

Similarly, *psychological features* (51 features) target emotional, social process, and cognitive processes. The affective processes (positive and negative emotions), social processes, cognitive processes, perceptual processes, biological processes, time orientations, relativity, personal concerns, and informal language (swear words, nonfluencies) can be used to scrutinize the emotional part of the news.

Summary features (7 features) define the frequency of words that reflect the thoughts, perspective, and honesty of the writer. It consists of Analytical thinking, Clout, Authenticity, Emotional tone, Words per sentence, Words more than six letters, and Dictionary words under this category.

Readability. Readability measures how easily the reader can read and understand a text. Text complexity is measured by using attributes such as word lengths, sentence lengths, and syllable counts. We use popular readability measures in our analysis: Flesh Reading Ease, Flesh Kincaid Grade Level, Coleman Liau Index, Gunning Fog Index, Simple Measure of Gobbledygook Index (SMOG), Automatic Readability Index (ARI), Lycee International Xavier Index (LIX), and Dale-chall Score. Higher scores of Flesch reading-ease indicate that the text is easier to read, and lower scores indicate difficult to read. Coleman Liau Index depends on characters of the word to measure the understandability of the text. The Gunning Fog Index, Automatic Readability Index, SMOG Index, Flesh Kincaid Grade Level are algorithmic heuristics used for estimating readability, that is, how many years of education is needed to understand the text. Dale-Chall readability test uses a list of words well-known for the fourth-grade students (easily readable words) to determine the difficulty of the text. We use this group of 9 readability features to measure news writing style complexity.

4.2 Image Features

To analyze the image associated with the news, we consider several tools, including (1) the ImageNet-VGG19 state-of-the-art deep-learning-based techniques to extract features from the images, (2) features describing face emotions, and (3) features referring to image quality such as noise and blur detection. Details regarding the features extracted to analyze images are reported in the following.

ImageNet-VGG19. We used a VGG19 pre-trained model from Keras for the visual feature extraction, which demonstrated a strong ability to generalize the images outside the ImageNet dataset via transfer learning [5]. We removed the classification layer of the VGG19 model and used the last fully connected layer of the neural network to generate a vector of latent features representing each input image. We used PCA to reduce the number of extracted features to 10.

Face Emotions. Images associated with factual news articles typically depict a figure speaking, whereas the misleading news articles contain more images of people with only expressions on their faces. Further, images in real news usually portray people with more positive expressions than people depicted in misleading news images. Thus, to capture face emotions in images, we used Microsoft Azure Cognitive Services API to detect faces in an image[2] which extracts several face attribute features. Among all the features extracted, we consider face emotion (anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise) and smile features. Each of these features ranges in [0,1] and indicates the confidence of observing the feature in the image.

Image Quality. Misleading news images are more likely to have been manipulated (e.g., via photoshop) and have a lower quality than factual news images

Table 2. Feature ablation for FakeNewsNet (left) and BuzzFeedNews (right) datasets.

Features	AUROC	F1	Avg. Prec.
News Content			
Readability	0.622	0.520	0.530
Punctuation (LIWC)	0.744	0.625	0.662
Linguistic (LIWC)	0.732	0.599	0.642
Psychological (LIWC)	0.728	0.623	0.634
Summary (LIWC)	0.666	0.550	0.542
All LIWC	0.751	0.615	0.666
All (Feature reduction (30))	0.784	0.663	0.697
Headline			
Upper Case WC	0.630	0.536	0.525
Stop Word Count	0.640	0.577	0.514
Readability	0.680	0.589	0.579
Punctuation (LIWC)	0.716	0.570	0.639
Linguistic (LIWC)	0.679	0.544	0.561
Psychological (LIWC)	0.604	0.520	0.503
Summary (LIWC)	0.674	0.557	0.596
All LIWC	0.675	0.547	0.639
All (Feature reduction (30))	0.801	0.657	0.756
Bias	0.868	0.739	0.670
Image			
Face Emotions	0.559	0.415	0.431
ImageNet-VGG19	0.534	0.420	0.419
Image Quality	0.551	0.430	0.400
All (Feature reduction (10))	0.595	0.479	0.466

Features	AUROC	F1	Avg. Prec.
News Content			
Readability	0.638	0.355	0.306
Punctuation (LIWC)	0.735	0.453	0.342
Linguistic (LIWC)	0.706	0.416	0.332
Psychological (LIWC)	0.741	0.446	0.400
Summary (LIWC)	0.675	0.399	0.302
All LIWC	0.762	0.477	0.410
All (Feature reduction (30))	0.771	0.477	0.410
Headline			
Upper Case WC	0.700	0.454	0.316
Stop Word Count	0.668	0.408	0.293
Readability	0.672	0.388	0.319
Punctuation (LIWC)	0.686	0.403	0.348
Linguistic (LIWC)	0.639	0.367	0.276
Psychological (LIWC)	0.631	0.357	0.298
Summary (LIWC)	0.621	0.347	0.265
All LIWC	0.734	0.445	0.386
All(Feature reduction (30))	0.794	0.520	0.420
Bias	0.708	0.563	0.386

[2] https://docs.microsoft.com/en-us/azure/cognitive-services/face/quickstarts/csharp.

typically. Thus, to capture news image quality to some extent, we computed the amount of blur in an image by using the OpenCV blur detection tool[3] implementing a method based on the Laplacian Variance [17] along with noise level of face pixels provided by Microsoft Azure Cognitive Service API.

4.3 Source Bias

Several studies in the field of journalism have theorized a correlation between the political bias of a publisher and the trustworthiness of the news content it distributes [8,11]. To validate this assumption, we examine the relationship between the political bias of a news source and its credibility by analyzing the information about 1,785 publishers in the MediaBias/FactCheck dataset.

Figure 2 shows the distribution of the credibility score per political bias category (from Left to Right) and the bias distribution in the questionable sources. The plots show that when the news source is moderate to strongly biased (either conservative or liberal), then the source is more likely to publish misleading news than other news sources that are more moderate and declared as left-centered, right-centered, or neutral. Also, we see that *Extreme-right* (or strongly conservative) is the predominant bias among the questionable sources. Thus, we also use the news source bias as another modality in our analysis.

Table 3. Top-30 most important news body content features and their corresponding logistic regression coefficients for the FakeNewsNet (left) and BuzzFeedNews (right).

FakeNewsNet				BuzzFeedNews			
Factual		**Misleading**		**Factual**		**Misleading**	
-0.97	assent	1.77	death	-1.08	affect	0.97	posemo
-0.87	hear	1.02	discrep	-0.71	fleschkincaid	0.86	negemo
-0.86	interrog	0.85	sexual	-0.61	dalechallknown	0.77	smog
-0.84	risk	0.82	informal	-0.61	nonflu	0.62	ari
-0.83	sad	0.81	motion	-0.55	dalechallscore	0.48	bio
-0.83	Parenth	0.69	shehe	-0.46	Dash	0.46	male
-0.61	relativ	0.68	family	-0.44	percept	0.45	filler
-0.54	compare	0.68	swear	-0.43	SemiC	0.43	female
-0.54	gunningfog	0.67	bio	-0.43	body	0.36	see
-0.52	auxverb	0.65	QMark	-0.43	ingest	0.35	affiliation
-0.51	i	0.54	colon	-0.41	gunningfog	0.34	anx
-0.51	drives	0.53	they	-0.40	swear	0.33	relig
-0.50	cogproc	0.51	netspeak	-0.29	shehe	0.28	Colon
-0.45	social	0.51	tentat	-0.25	friend	0.26	adverb
-0.45	you	0.51	adj	-0.25	netspeak	0.26	assent

[3] https://www.pyimagesearch.com/2015/09/07/blur-detection-with-opencv/.

5 Multi-modal Analysis

We used each group of features described in the previous section in input to a logistic regression classifier with L2 regularization (with 5-fold cross-validation) to compute the performance of these features in classifying factual vs. misleading stories. We also tried other classifiers such as Support Vector Machine (SVM) and Random Forest, but Logistic Regression achieved the best results. Hence, we report in the paper Logistic Regression results only. We used class weighting to deal with class imbalance. The results for logistic regression are reported in Table 2 according to the area under the ROC curve (AUROC), F1-measure (F1), and average precision (AvgP) and discussed in the following.

News Body Content. The first modality we analyze is the news body content. Here, we see that the LIWC features are better than the readability features for both the datasets: 0.75 vs. 0.62 AUROC for FakeNewsNet and 0.76 vs. 0.64 for BuzzFeedNews. Also, performances are comparable for both the dataset, according to AUROC. One difference between the two datasets is the most important group of features within the LIWC features: punctuation features are the most important ones for FakeNewsNet (0.74 AUROC, 0.63 F1, 0.66 AvgP) whereas psychological features (0.69 AUROC, 0.40 F1, 0.35 AvgP) are the best predictors for the BuzzFeedNews dataset. As the latter has a higher class imbalance than FakeNewsNet (19% vs. 45% of misleading news), we obtain lower values of F1-measure and average precision.

Combining both readability and LIWC features (and by performing feature reduction to avoid overfitting) classification results improve with respect to each group of features individually: AUROC of 0.78 for FakeNewsNet and 0.77 for BuzzFeedNews. Feature reduction consists of the most informative features in the news body content computed by using the coefficients of a logistic regression model (30 features in total, 15 for factual news, and 15 for misleading ones). Table 3 shows these most important features for FakeNewsNet and BuzzFeedNews and the corresponding coefficients from the logistic regression model. We see that readability features appear within the most important features in both datasets. By comparing the readability of factual and misleading news, we observe

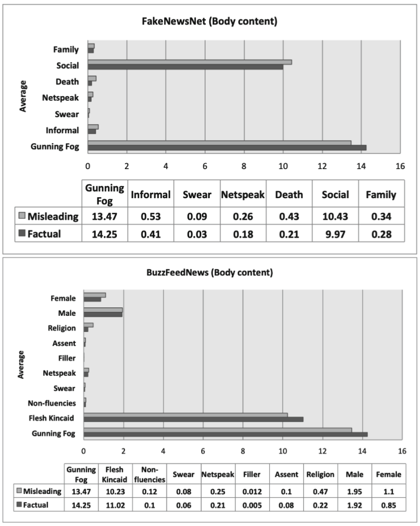

Fig. 3. Most important features for news body content with average values for factual and misleading news: FakeNewsNet (top) and BuzzFeedNews (bottom).

Table 4. Top-30 most important headline features and their corresponding logistic regression coefficients for FakeNewsNet (left) and BuzzFeedNews (right) datasets.

FakeNewsNet			
Factual		**Misleading**	
-1.13	colemanliau	1.47	ari
-1.12	Parenth	1.10	friend
-1.10	affiliation	1.04	we
-0.89	negate	0.67	Exclam
-0.83	fleschkincaid	0.94	sexual
-0.76	# stopwords	0.79	motion
-0.60	shehe	0.60	tentat
-0.48	relativ	0.57	family
-0.43	lix	0.55	space
-0.39	i	0.46	netspeak
-0.38	home	0.46	differ
-0.33	male	0.45	they
-0.33	nonflu	0.45	reward
-0.32	bio	0.41	time
-0.30	Colon	0.37	body

BuzzFeedNews			
Factual		**Misleading**	
-0.62	dalechallknown	0.35	# uppercase words
-0.42	swear	0.22	ari
-0.39	nonflu	0.17	informal
-0.36	# stopwords	0.17	fleshkincaid
-0.32	assent	0.15	WPS
-0.22	netspeak	0.15	Exclam
-0.20	dalechallscore	0.15	health
-0.18	colemanliau	0.14	hear
-0.11	home	0.13	relig
-0.10	drives	0.13	female
-0.09	time	0.12	they
-0.08	i	0.12	affiliation
-0.08	WC	0.10	ingest
-0.08	Apostro	0.09	male
-0.08	social	0.08	power

that factual news is harder to understand. We have, on average, higher values of readability scores in factual than misleading news, indicating higher text complexity (cf. Fig. 3). On the other hand, misleading news uses more informal language and tentative words evoking uncertainty than factual ones. As we see in Fig. 3, on average, misleading news has higher scores for these language features on both datasets: higher frequency of informal words (e.g., 'thnx', 'hmm', 'youknow'), swear words, and netspeak (words frequently used in social media and text messaging in FakeNewsNet, and higher frequencies of non-fluencies (e.g. 'er', 'umm', 'uh', 'uh-huh'), swear words, netspeak, filler words and assent words in BuzzFeedNews. The above analysis clearly shows that factual news in both datasets is written with complex constructions of texts, which is mostly seen in the field of journalism [4], unlike the misleading ones which are written informally showing non-professional character.

Also, misleading news in both datasets has higher frequencies of psychology related words such as personal concerns (death in FakeNewsNet and religion-related words in BuzzFeedNews) and social words (e.g., social and family-related words in FakeNewsNet and male and female related words in BuzzFeedNews).

News Headline. Among all the features we considered to analyze the news headline, we see in Table 2 that, LIWC punctuation features are the best group of features in both datasets achieving an AUROC of 0.72 (resp. 0.69), an F1-measure of 0.57 (resp. 0.40) and an average precision of 0.64 (resp. 0.35) on FakeNewsNet (resp. BuzzFeedNews) dataset. This shows that the headline's features are more consistent across datasets than news body content. Similarly to the news body content, by combining both readability and LIWC features (and by performing feature reduction to avoid overfitting as we did for excerpt features), classification results improve with respect to each group of features individually: AUROC of 0.80 for FakeNewsNet and 0.79 for BuzzFeedNews.

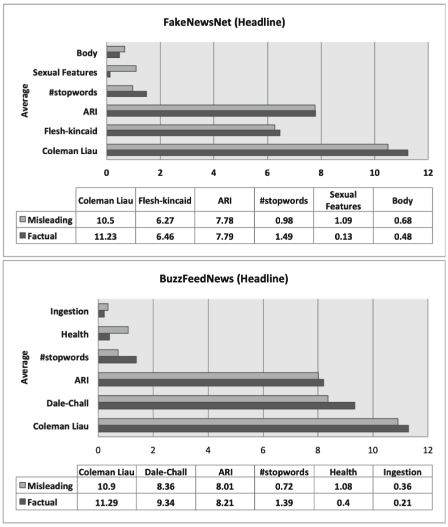

Fig. 4. Most important features for news headline with average values for factual and misleading news: FakeNewsNet (top) and BuzzFeedNews (bottom).

Table 4 shows the most important headline features in our datasets. Figure 4 shows the average values for factual vs. misleading news of the best features discussed in the following. Again, readability measures appear among the most important features in both datasets. Comparing the average values of readability features between factual and misleading news provides evidence that factual news headlines are written professionally than misleading ones. Also, factual news headlines of both datasets have a higher average value of stopwords count, while BuzzFeedNews misleading news headlines are written using more capital letters.

In addition, we see that the misleading news headlines have higher frequency of words related to biological processes (e.g., 'eat', 'blood', 'pain'), namely sex (e.g., 'love', 'incest', 'beauty') and body lexicon (e.g., 'cheek', 'hands', 'lips') in FakeNewsNet, and health related words (e.g., 'clinic', 'pill', 'ill') and ingestion (e.g., 'eat', 'dish') in BuzzFeedNews.

This analysis shows that the orientation towards the feelings, body, and health lexicon is a very strong characteristic of a misleading news headline. Observing such biological words occurring significantly more in misleading news than in factual ones indicates that the former is made more sensational along with more uppercase letters for exaggerations to catch the reader's attention.

News Source Bias. The news source bias is a strong predictor for news credibility in both the datasets considered, and it achieves AUROC of 0.87 (resp. 0.71), F1-measure of 0.74 (resp. 0.56), and average precision of 0.67 (resp. 0.39) in the FakeNewsNet (resp. BuzzFeedNews) dataset. This result further confirms the correlation between source bias and the credibility of the news it distributes. It is worth noting that the publisher's information is independent of the news labels as the former is collected from MediaBias/FactCheck, while the latter from Buzzfeed and Politifact.

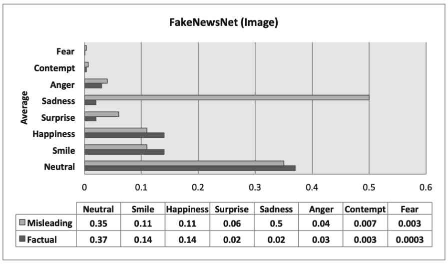

Fig. 5. Most important features for news image and average values for factual and misleading news.

Table 5. Top-10 most important image features and corresponding logistic regression coefficients for FakeNewsNet.

Factual		Misleading	
−0.16	Happiness	1.02	Surprise
−0.16	Smile	0.61	Sadness
−0.14	Noise	0.29	Anger
−0.07	Neutral	0.09	Contempt
−0.03	VGG19	0.08	Fear

News Image. Image features are not as good as other modalities in detecting misleading news in the FakeNewsNet dataset. However, when we use the image associated with the news to determine the news validity, we see that features describing face emotions achieve best results according to AUROC (0.56) and average precision (0.43), while image quality features are the best according to F1-measure. Moreover, by combining all the image features (and performing feature reduction by considering only the top-10 most important features according to the coefficients of the logistic regression), we improve the classification results up to 0.60 AUROC, 0.48 F1-measure, and 0.47 average precision. The top-10 most important image features are reported in Table 5. As expected, we see the face emotion-based features to be the most important ones. Figure 5 shows the average values for factual vs. misleading news of the best image features. Here, we see that, on average, images associated with factual news depict people with more neutral-positive emotions (neutral, smile, happiness) than images associated with misleading news. On the other hand, misleading news is paired with more provocative images showing people expressing, on average, more surprise, sadness, anger, contempt, and fear. Also, only one ImageNet-VGG19 feature appears in the top-10, where we find the noise level of face pixel feature as well.

5.1 Do We Need to "Read"?

Here, we address the question of whether we need to look at the news body content to detect misleading news, or we can achieve better results by using other modalities. Fairbanks et al. [9] posed and investigated this question for the first time and found that exploiting web links within news articles' bodies outperforms body text-based features for misleading news detection. To address the question

Table 6. Results comparing news snippet feature combination (headline, image, and source bias) with news body content for FakeNewsNet (left) and BuzzFeedNews (right).

Features	AUROC	F1	Avg. Prec.
Headline	0.801	0.657	0.756
Headline + Image	0.821	0.678	0.725
Headline + Image + Bias	**0.908**	**0.783**	**0.817**
News Content	0.784	0.663	0.697

Features	AUROC	F1	Avg. Prec.
Headline	0.794	0.520	0.420
Headline + Bias	**0.812**	**0.534**	**0.462**
News Content	0.771	0.477	0.410

in our case, we can refer to the first part of our analysis and Table 2. We see that, in both datasets, we get better information from the news headline to determine whether it is factual or not: AUROC of 0.80 vs. 0.78 in FakeNewsNet and 0.79 vs. 0.77 in BuzzFeedNews. This result confirms and generalizes by using larger datasets the finding of Horne and Adali [13] that the news title is more informative than the body content. Moreover, in the case of the FakeNewsNet dataset, considering the publisher bias achieves a better AUROC of 0.87.

5.2 Can We Detect Misleading News from Its Snippet?

Next, we address the question of whether combining headline, bias and image features, hence considering the news snippet and mimic how news is distributed on social networks, can further improve misleading news detection results. Table 6 report the combined results for FakeNewsNet (left) and BuzzFeedNews (right). For headline, image, and content, we consider the most important features previously computed via feature reduction (30, 10, 30 features, respectively). The first observation is that, even if the image features alone are not enough to differentiate between factual and misleading news (AUROC of 0.60 in the FakenewsNet, cf. Table 2), we see from Table 6 (left) that they help in improving classification results when combined with the headline features (2% improvement for AUROC and F1-measure). Moreover, adding the source bias further improves up to 0.91 AUROC, 0.78 F1-measure, and 0.82 average precision. In the case of the BuzzFeedNews dataset, we do not have image information, but Table 6 (right) shows that adding the bias to the headline features achieves 0.81 AUROC, 0.53 F1-measure, and average precision 0.46, which is better that only consider the news body content. It is worth noting that, as reported in Sect. 2, Potthast et al. [20] addressed the problem of automatically detecting misleading stories in the BuzzFeedNews dataset achieving an F1-measure of 0.46. They only analyzed news content with a different set of style-based features. However, their experimental setting was different from the one of this paper. Thus, for a fair comparison with the methods used in this paper, we reproduced their setting (considering only the left-wing articles and the right-wing articles of the corpus and balancing the dataset via oversampling) and computed classification results. We achieve an F1-measure of 0.58 with the news body content (best 30 features from readability and LIWC) and F1-measure of 0.61 when we consider the combination of the best 30 headline features and source bias. In both cases, we improve their proposed method.

Thus, our analysis reveals that looking at the news snippet by considering the headline characteristics from Table 4, checking the publisher bias and putting more attention on the associated images provides user-friendly tools that can be taught to people via media literacy to warn them about possible misleading news and can hopefully prevent people from massively spreading non-factual news through online social media.

6 Conclusion

We presented an analysis of the relative importance of different news modalities (body, headline, source bias, and visual content) in detecting misleading political news. In particular, our findings demonstrate a strong correlation between political bias and news credibility and the importance of image emotion features. Moreover, we showed that it is not necessary to analyze the news body to assess its validity, but comparable results can be achieved by looking at alternative modalities, including headline features, source bias, and visual content.

References

1. Mediabias/factcheck apps/extensions. https://mediabiasfactcheck.com/appsextensions/
2. Pen America: Faking news: Fraudulent news and the fight for truth (2018). https://pen.org/faking-news/
3. Antol, S., et al.: VQA: visual question answering. In: ICCV, pp. 2425–2433. IEEE (2015)
4. Barton, D.: Literacy: An Introduction to the Ecology of Written Language. Wiley, Hoboken (2017)
5. Chou, C.N., Shie, C.K., Chang, F.C., Chang, J., Chang, E.Y.: Representation learning on large and small data. arXiv:1707.09873 (2017)
6. Boididou, C., et al.: Verifying multimedia use at mediaeval 2015. In: MediaEval (2015)
7. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: NAACL-HLT Volume 1 (Long and Short Papers), pp. 4171–4186 (2019)
8. Entman, R.M.: Framing bias: media in the distribution of power. J. Commun. **57**(1), 163–173 (2007)
9. Fairbanks, J.P., Knauf, N., Georgia, E.B.: Credibility assessment in the news: do we need to read? In: MIS2: Misinformation and Misbehavior Mining on the Web Workshop (2018)
10. Gabielkov, M., Ramachandran, A., Chaintreau, A., Legout, A.: Social clicks: what and who gets read on Twitter? ACM SIGMETRICS Perform. Eval. Rev. **44**(1), 179–192 (2016)
11. Gentzkow, M., Shapiro, J.M., Stone, D.F.: Media bias in the marketplace: theory. In: Handbook of Media Economics, vol. 1, pp. 623–645. Elsevier (2015)

12. Gupta, M., Zhao, P., Han, J.: Evaluating event credibility on Twitter. In: SDM, pp. 153–164. SIAM (2012)
13. Horne, B.D., Adali, S.: This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. arXiv preprint arXiv:1703.09398 (2017)
14. Jin, Z., Cao, J., Guo, H., Zhang, Y., Luo, J.: Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In: ACMMM, pp. 795–816 (2017)
15. Jin, Z., Cao, J., Zhang, Y., Zhou, J., Tian, Q.: Novel visual and statistical image features for microblogs news verification. IEEE Trans. Multimed. **19**(3), 598–608 (2017)
16. Mitra, T., Gilbert, E.: CREDBANK: a large-scale social media corpus with associated credibility annotations. In: ICWSM, pp. 258–267 (2015)
17. Pech-Pacheco, J.L., Cristóbal, G., Chamorro-Martinez, J., Fernández-Valdivia, J.: Diatom autofocusing in brightfield microscopy: a comparative study. In: ICPR, vol. 3, pp. 314–317. IEEE (2000)
18. Pennebaker, J.W., Boyd, R.L., Jordan, K., Blackburn, K.: The development and psychometric properties of LIWC2015. Technical report (2015)
19. Pérez-Rosas, V., Kleinberg, B., Lefevre, A., Mihalcea, R.: Automatic detection of fake news. In: ACL, pp. 3391–3401 (2018)
20. Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., Stein, B.: A stylometric inquiry into hyperpartisan and fake news. In: ACL, pp. 231–240 (2018)
21. Shearer, E., Gottfried, J.: News use across social media platforms 2017 (2017)
22. Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H.: Fake news detection on social media: a data mining perspective. ACM SIGKDD Explor. Newslett. **19**(1), 22–36 (2017)
23. Shu, K., Wang, S., Liu, H.: Beyond news contents: the role of social context for fake news detection. In: WSDM, pp. 312–320 (2019)
24. Silverman, C.: This analysis shows how viral fake election news stories outperformed real news on Facebook. BuzzFeed News **16** (2016)
25. Ping Tian, D., et al.: A review on image feature extraction and representation techniques. Int. J. Multimed. Ubiquit. Eng. **8**(4), 385–396 (2013)
26. Wang, W.Y.: "Liar, liar pants on fire": a new benchmark dataset for fake news detection. In: ACL, pp. 422–426 (2017)
27. Wang, Y., et al.: EANN: event adversarial neural networks for multi-modal fake news detection. In: SIGKDD, pp. 849–857. ACM (2018)