KeAi
CHINESE ROOTS
GLOBAL IMPACT

Contents lists available at ScienceDirect

International Journal of Cognitive Computing in Engineering

journal homepage: https://www.keaipublishing.com/en/journals/international-journal-of-cognitive-computing-in-engineering/

Fake News Classification using transformer based enhanced LSTM and BERT

Nishant Rai, Deepika Kumar*, Naman Kaushik, Chandan Raj, Ahad Ali

Department of Computer Science & Engineering, Bharati Vidyapeeth's College of Engineering, New Delhi 110063, India

ARTICLE INFO

Keywords:
Fake News
classification
transformer
Natural language processing

ABSTRACT

Fake News has been a concern all over the world and social media has only amplified this phenomenon. Fake News has been affecting the world on a large scale as these are targeted to sway the decisions of the crowd in a particular direction. Since manually verifying the legitimacy of news is very hard and costly, there has been a great interest of researchers in this field. Different approaches to identifying fake news were examined, such as content-based classification, social context-based classification, image-based classification, sentiment-based classification, and hybrid context-based classification. This paper aims to propose a model for fake news classification based on news titles, following the content-based classification approach. The model uses a BERT model with its outputs connected to an LSTM layer. Training and evaluation of the model were done on the FakeNewsNet dataset which contains two sub-datasets, PolitiFact and GossipCop. A comparison of the model with base classification models has been done. A vanilla BERT model has also been trained on the dataset under similar constraints as the proposed model has to evaluate the impact same using an LSTM layer. The results obtained showed a 2.50% and 1.10% increase in accuracy on PolitiFact and GossipCop datasets respectively over the vanilla pre-trained BERT model.

1. Introduction

Fake News is the misinformation disseminated among the public by mainstream sources like media outlets and social media. It is generally misleading to shape beliefs of the masses to one's favour. As Paskin stated "particular news articles that originate either on mainstream media (online or offline) or social media and have no factual basis, but presented as facts and not satire" (Paskin, 2018). It has been a topic of attention because this has been affecting our lives in various ways as there have been many incidents that have demonstrated the same very clearly. Research has indicated that fake news was a significant factor for Donald Trump's win as US president in the 2016 elections (Grinberg et al., 2019). Also, the Brexit polling decision of the UK citizens. During these events, fake news was targeted towards social media users depending upon their ideologies to persuade them to lean to a particular side and vote in their favour (Hopkin and Rosamond, 2018). The social media platform like Whatsapp realized the scale of the widespread culture of sharing fake news and consequently had to run awareness campaigns against it in India (Farooq, 2017). Fake news was ramping up following the COVID-19 breakout throughout the world and there was a profusion of rumors and fake news making rounds on social and mainstream media (O'Connor and Murphy, 2020, Kadam and Atre, 2020). It created chaos and problems to an extreme extent for people around the world and even resulted in a huge amount of fatalities.

The fake news ecosystem plays the basic human psychology a) *Naive Realism*: Believing the perception that only one's reality is correct and others are wrong. b)*Confirmation Bias*: Believing in the information that reinforces existing biases. Fake news has been a concern all over the world for the past several years. The concerns raised by it have only escalated with the constantly increasing time of people being spent on social media and thus being the main source of news for them. This attitude has just become much more common as social media serves its users the stories and posts that align with their ideologies. This gives rise to an echo chamber effect where the interactions of the user with others or the content he/she consumes match their existing ideologies and thought processes (Törnberg, 2018). Thus, most people never bother to check the trustworthiness of the news as it is supported by their confirmation bias (Moravec et al., 2018).

The solution to this problem is imperative, considering the above-stated facts about the explosion of fake news in the past decade and how it has been exploited in various areas. Most people interact with social media daily and share posts and news articles whose legitimacy is not determined which amplifies the effect of the fake news. Because of the widespread prevalence of fake news all around the world, there

* Corresponding author
E-mail address: deepika.kumar@bharatividyapeeth.edu (D. Kumar).

https://doi.org/10.1016/j.ijcce.2022.03.003
Received 29 October 2021; Received in revised form 15 March 2022; Accepted 15 March 2022
Available online 19 March 2022

are private media outlets that exclusively check the legitimacy of the news by verifying the facts presented in the report (Luengo and García-Marín, 2020). But as the world has moved towards social media, the amount of news content being published is colossal and classifying fake news is becoming increasingly difficult to assess the legitimacy of the news manually. This process is getting more tedious and more costly day by day, with the influx of huge amounts of information being generated every day. With the rise of DeepFake, checking media for manipulations from the source is essential for verifying the integrity and the information conveyed by it (Singh et al., 2020). Thus, recently computer researchers have been attempting to automate the process. This also motivated the authors to propose a model to automate the process and identify fake news patterns in news articles and media.

SVM and Naive Bayes (NB) classifiers have been used by various researchers in this field. These models differ in their functioning and structure but both produced similar results and were used as baseline models (Prasetijo et al., 2017, Granik and Mesyura, 2017). Various clustering algorithms and decision trees have been used extensively in the literature for experimentation (Goyal et al., 2016). Recurrent Neural Networks (RNNs) are very popular in this field, especially Long Short-Term Memory (LSTM) (Sundermeyer et al., 2012). However, RNNs usually face the problem of vanishing gradients, which hinders their capability of learning long data sequences which are solved by LSTMs. Word embedding is an important factor to be considered while designing a model for NLP problems (Mikolov et al., 2018). To improvise upon this, the proposed research used contextual word embedding using the BERT model (Devlin et al., 2018). BERT can learn contextualized word representations by utilizing a huge volume of unlabeled text corpora. BERT has performed well in the NLP tasks because of its intricate structure and excellent nonlinear representation learning capability. LSTMs effectively boost performance by memorizing and finding the pattern of key information. Therefore, contextualized word representations from BERT are employed in LSTM to improve fake news classification performance, thanks to their powerful ability to capture semantics and long-distance dependencies in news titles.

The research contributions have been summarized as follows:

- The proposed methodology classifies the news based on its linguistic features such as syntactical, grammatical and semantical aspects of the news reports and articles.
- To propose an approach for fake news classification by combining the BERT model with an LSTM which classifies news articles as either fake or legitimate.
- Accuracy, Precision, Recall, and F1 Score have been used as the evaluation criteria for checking the robustness of the proposed methodology.
- Empirical evaluation of the proposed methodology has been conducted with state-of-the-art methodologies such as conventional TCNN-URG, LIWC, CSI, HAN, SAFE etc. based on the various training and testing phases.

The paper is organized as follows: Section 2 discusses the literature done in the area of NLP and fake news detection Section 3. explains the dataset description, architecture of BERT and LSTM which is followed by the architecture of the proposed model Section 4. depicts the detailed Results & Analysis. The performance of the proposed methodology has been compared and analyzed with state-of-the-art methods which are followed by the conclusion section.

2. Literature Review

Earlier studies in the field of fake news detection revealed that various lexical features can be useful in understanding the differences between more and less trustworthy digital news sources. The authors combined LIWC (Linguistic Inquiry and Word Count) measurements with the original text and found that the linguistic features improved the F1 score on the test data for most models. The authors state that fact-checking is a challenging task but various lexical features can contribute to the understanding of the differences between more reliable and less reliable digital news sources (Rashkin et al., 2017). The dataset called FakeNewsNet was made publicly available for fake news detection. A comprehensive study has been done on the dataset and the results have been compared with classification models like SVM, Logistic Regression, NB, CNN, SAF/S (utilizing news content), and its variant SAF/A (utilizing social context). Amongst the compared models, a combined model of SAF/A and SAF/S with LSTM cells had achieved the best accuracy, i.e. 70.6% on the PolitiFact dataset; and 71.7% on the GossipCop dataset (Shu et al., 2020). A comprehensive study with hybrid models using N-Gram, Word-Embeddings, and Topic Models for content-based classification was proposed (Aggarwal et al., 2020, Walia et al., 2021). Hybrid models such as N-Gram, N-Gram + Topic, N-Gram + Word2Vec, Word2Vec + Topic and N-Gram + Word2Vec + Topic have been compared and analyzed, and an accuracy of 80%, 77%, 72%, 42% and 40% respectively and F1 score of 0.78, 0.76, 0.72, 0.39 and 0.46 respectively has been achieved. It was noted that the performance of machine learning models decreased as more models are combined, most probably due to high bias (Oriola, 8887).

Classification of fake news has been performed for news content features or social context features. Twitter data has been mined for social context and news content features extraction from the source, headline, body text of the news and images. The tweets can be targeted using specific words (Mittal et al., 2019). Retweet rate, the time difference between the retweets, users retweeting are some of the important features that provided social context along with the text content of the tweet. Another useful feature to mine is user comments on the tweet, which provided additional text content. Utilizing the text-content and user comments, a study compared fake news classification, models like RST, LIWC, text-CNN, and HAN which uses news-content. Models like HPA-BLSTM which relied only on user comments, i.e. social-context; and models like TCNN-URG and CSI utilised both news content and user comments for fake news classification. These models have been compared along with a proposed model called dEFEND, consisting of a word encoder, a sentence encoder, a user comments encoder, a sentence-comment co-attention layer and a fake news prediction component. The authors reported that dEFEND improved on other models and had an accuracy of 90.4% with an F1 score of 0.928 on the PolitiFact dataset; and an accuracy of 80.8% on the GossipCop dataset with an F1 score of 0.755. The authors observed a drop in the accuracy when either the co-attention for news contents or the user comments were removed. The results showed that user comments were necessary to guide fake news detection in dEFEND (Shu et al., 2019). A similar hybrid model utilizing GRU and RNN for word encoding, sentence encoding and user comments encoding was proposed with SVM as the classifier unit. It had an accuracy of 91.2% on PolitiFact with an F1 score of 0.932; and an accuracy of 80.2% on GossipCop with an F1 score of 0.762, but with the same limitation of relying on user comments (Albahar, 2021).

The possibilities of using hierarchical propagation networks (HPN) to perform temporal, i.e. time differences between post and user replies; and linguistic analysis, i.e. sentiment of the post and that of different levels of user replies were investigated. Hybrid frameworks utilizing HPNs paired with existing content-based classification models showed an overall improvement in results over the same existing content-based classification models not utilizing HPNs. Amongst the models compared by the authors, RST_HPFN (Rhetorical Structure Tree + Hierarchical propagation network) had the highest best accuracy, i.e. 87.5% on the PolitiFact dataset with an F1 score of 0.843; and LIWC_HPFN (Linguistic Inquiry and Word Count + Hierarchical propagation network) achieved the highest accuracy, i.e. 86.9% on the GossipCop dataset with an F1 score of 0.871. This model relied on the propagation data of the fake news including the temporal data of retweets and information of users sharing the tweet (Shu et al., 2020).An early detection method for fake news has been proposed which utilizes a pre-trained BERT summarization model for text summarization and GEAR, a fact verification model based on BERT. This method had an accuracy of 68.2% on the PolitiFact dataset with an F1 score of 0.725; and an accuracy of 73.8% on the GossipCop dataset with an F1 score of 0.525. The authors state that the model can be computationally expensive to use in real-world applications (Li and Zhou, 2020).GloVe word embeddings and a 1-D CNN for n-gram feature extraction, followed by an LSTM layer for temporal feature extraction has been used for content-based fake news classification (Agarwal et al., 2020). In another study, the use of LSTM cells with the Attention model (LSTM-ATT) for content-based classification was investigated, which had an accuracy of 83.3% on the PolitiFact dataset with an F1 score of 0.83; and an accuracy of 79.3% on the GossipCop dataset with an F1 score of 0.79. The authors concluded that the proposed model performed well compared with baseline models (Lin et al., 2019).

SpotFake+, a multimodal framework utilizing XLNet for text processing and VGG-19 for image processing had an accuracy of 84.6% on PolitiFact and 85.6% on GossipCop. It was noticed that text and image classification on GossipCop had an accuracy of 83.6% and 80% respectively (Singhal et al., 2020). The importance of compound sentiment and retweet rate for the classification of fake news were explored and a simple feed-forward neural network used as a classifier. An accuracy of 64% over the datasets with an F1 score of 0.64 was achieved. A neural network with only compound sentiment was found to perform similar to one using both compound sentiment and retweet rate (Ezeakunne et al., 2020). In recent years, transformer-based models, like BERT has been explored for the task of fake news classification. One such proposed model utilizes three pre-trained BERT models for statements, metadata and justifications present in the LIAR PLUS dataset. The proposed triple-BERT framework had an accuracy of 74% on the dataset. Notably, on the LIAR dataset, consisting of statement and metadata, a double-BERT model had an accuracy of 72% (Mehta et al., 2021). In another study, the authors proposed a framework utilizing a pre-trained BERT and three parallel blocks of 1d-CNN having different kernel-sized convolutional layers with different filters for better learning. It gained high accuracy for the task of fake news classification on real-world fake news datasets as compared to state-of-the-art deep learning models, indicating that features output from BERT proved to be more useful for the task (Kaliyar et al., 2021). A study utilized BERT for sarcasm detection which focused on the context-based feature technique for sarcasm identification using deep learning, transformer learning, and conventional machine learning models on different datasets. It was observed that BERT performed better than the deep learning models consisting of GloVe embeddings, which shows BERT performed well in learning contextual features from the data (Eke et al., 2021). Yet another study discussed that BERT is best suited for fake news classification because of its deep contextualizing nature. Two models has been proposed i.e. BAKE, with weighted cross-entropy as training loss and exBAKE, both the models have been trained on CNN and Daily Mail data whereas FNC-1 dataset has been used for the fine-tuning. The authors reported F1 score of pre-trained BERT with cross-entropy as a training loss to be 0.656. F1 score of BAKE and exBAKE has been reported to be 0.734 and 0.746 respectively. The authors concluded that exBAKE was able to achieve better performance in majority categories and performed better in minority categories of the FNC-1 dataset. The use of weighted cross-entropy was found to be crucial for showing competitive results in fake news detection in FNC-1 dataset (Jwa et al., 2019).

A bi-directional transformer approach with a feed-forward classification layer discussed the benefits of utilizing transformer-based models over machine learning models. The authors reported that fine-tuned BERT had an accuracy of 97.02% compared to XGBoost, which had an accuracy of 89.37% on the NewsFN dataset (Aggarwal et al., 2020). A comparative analysis of deep learning approaches for fake news identification based on the COVID-19 fake news dataset has been done which resulted that a fine-tuned BERT performed better than other models including BiLSTM+ attention. LSTM model trained on word embeddings performed similarly to BiLSTM+ attention. The importance of pre-training on target domains like corpus has also been discussed in the paper. It was concluded that the transformer models performed much better than the non-transformer and word-based models (Wani et al., 2021).

3. Methodology

This section details the architecture of the proposed model. It also contains details about the dataset on which the models are trained and evaluated. A brief background about the BERT and LSTM architectures, dataset used and preprocessing methods has been explained in this section. Fig. 1 depicts the proposed framework for content-based fake news classification:

BERT with an LSTM layer has been used as the classification model to classify the news titles. BERT with an LSTM layer is employed as the classification model to fulfil this purpose. It uses large number of unlabeled text corpora and acquire contextualized word representations. Because of its complicated structure and high nonlinear representation learning power BERT scored well in the NLP tests. By memorizing and finding the pattern of vital information, LSTMs efficiently increased performance.

3.1. Data Preprocessing

Data preprocessing is the essential step for training the models, therefore news articles present in the dataset has been preprocessed. Following are the steps taken to preprocess the data:

- Lowercased every word in the sentence
- Changed “’t” to “not”. *For example, can’t be changed to can not*
- Removed “@name”
- Isolated and removed punctuations except "?"
- Removed other special characters
- Removed stop words except "not" and "can"
- Removed trailing whitespaces.
- Tokenization of the cleaned text-content

After the text contents of the datasets were preprocessed, tokenization vectors, attention masks and their binary category were obtained from Bert Tokenizer and packed together for training and classification.

3.2. BERT

BERT (Bidirectional Encoder Representations from Transformers) has been made up of a transformer attention mechanism that learns contextual relationships among words. The transformer consists of an encoder responsible for reading text input. It also consists of a decoder which is responsible for prediction based on the task. In contrast to directional models which read the text input in sequential order, the transformer encoder reads all of the words simultaneously, thus giving it a non-directional nature. This means that the model learns the context of the word from all of its surrounding words. Thus, it is termed bidirectional BERT architecture for sentence-level classification Fig. 2. represents the architecture of BERT sentence-level classification model:

BERT has many versions of pre-trained models for different use cases. Two of the most used models are-

- BERT-base: 12 encoder stack layers + 768 hidden units + 12 multi-head attention heads: 110M parameters.
- BERT-large: 24 encoder stack layers + 1024 hidden units + 16 multi-head attention heads: 340M parameters.

The input data needs to be converted into an appropriate format before using the pre-trained model. Relevant embeddings for each sentence has been obtained. Each encoder layer in these models takes list of token embeddings and their attention masks as input. The same number of embeddings with the same hidden size has been taken as the output.

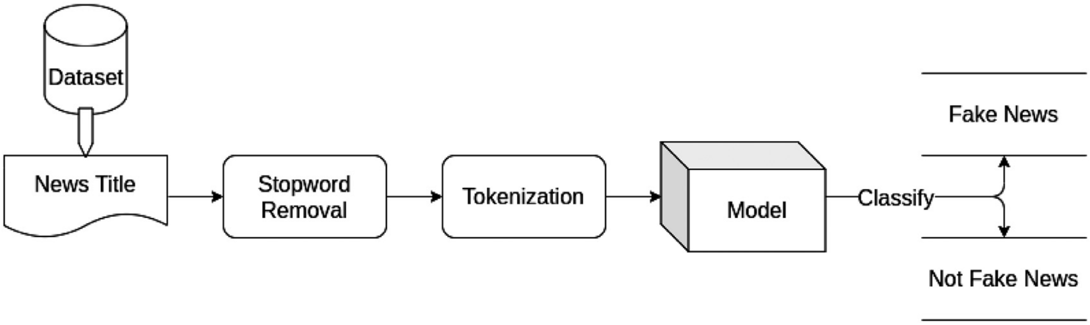

Fig. 1. Proposed framework for content-based fake news classification

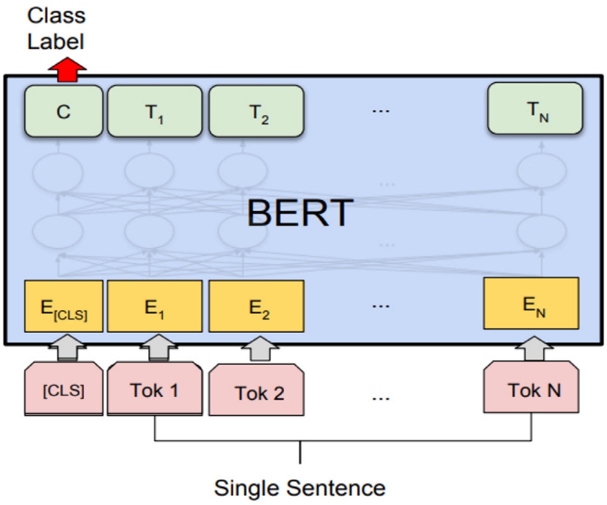

Fig. 2. The architecture of BERT sentence-level classification (Moravec et al., 2018)

A single vector representing the entire input sentence has been fed to the classifier, and the hidden state of the first token [CLS] of the model's output can be used to represent the entire sentence, which is used for classification purpose.

3.3. LSTM

Long short-term memory is a type of RNN that can learn long-term dependencies. The chain-like structure of LSTMs is similar to RNNs, but the base module that makes up the LSTM is structurally distinct from other RNNs. RNNs are good at learning small data sequences and excel at it (Luengo and García-Marín, 2020). The problem faced by RNNs is that they suffer from the vanishing gradient problem, which hampers their ability to learn and understand long sequences and context. LSTMs are special RNNs that do not face this problem and are well suited to learn long-term dependencies. All RNNs are made up of repeating modules but these generally have a very simple architecture such as a single *tanh* layer. LSTMs however are made up of repeating modules called cells containing four neural networks connected in a special manner which are shown in Fig. 3. Each cell passes two states to the next cell i.e. cell state and hidden state (c_t and h_t). Each cell is used to remember things and the manipulations to them are done using three mechanisms called gates namely being forgotten, input, and output gate. Forget Gate removes the information no longer useful for the LSTM which consists of a sigmoid layer that makes the decision. Input gate is responsible for the addition of relevant information to the current cell state and uses *tanh* and sigmoid layers. The output gate is responsible to show the relevant information from the current cell and employs a sigmoid layer Fig. 3 shows the LSTM architecture:

$$i_t = \sigma\left(W_i.\left[h_{t-1,}x_t\right] + b_i\right) \tag{1}$$

$$f_t = \sigma\left(W_f.\left[h_{t-1,}x_t\right] + b_f\right) \tag{2}$$

$$o_t = \sigma\left(W_o.\left[h_{t-1,}x_t\right] + b_o\right) \tag{3}$$

$$C\prime_t = \tanh\left(W_c.\left[h_{t-1,}x_t\right] + b_c\right) \tag{4}$$

$$C_t = f_t * C_{t-1} + i_t * C\prime_t \tag{5}$$

$$h_t = o_t * \tanh\left(C_t\right) \tag{6}$$

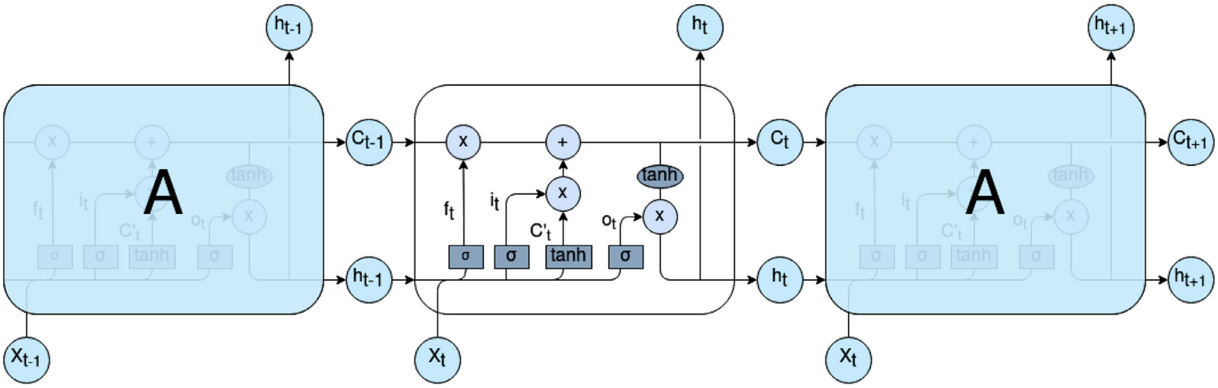

Fig. 3. LSTM architecture

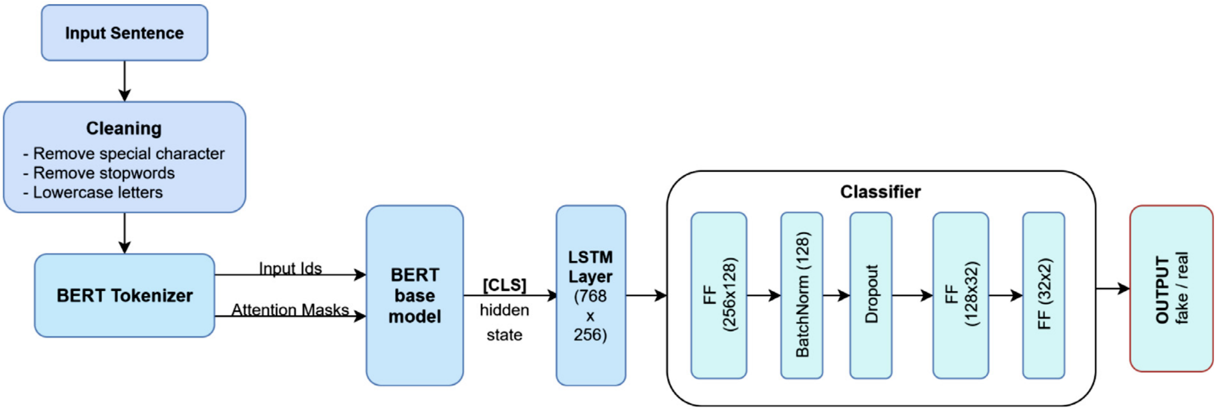

Fig. 4. Architecture of proposed methodology

3.4. Proposed Model Architecture

In the proposed methodology, the BERT-base-uncased model has been used with a feed-forward network of 768 hidden sizes. This BERT model takes inputs as input IDs and attention masks of the input sentence. BERT tokenizer has been used to perform this task which takes input sequence as [CLS] and [SEP] concatenated to sentence at the beginning and end respectively which gives the input ids and attention masks as output. These are fed into the BERT model which outputs a vector of embeddings for each token of hidden size 768. BERT provides contextualized sentence-level representations, which help LSTM to understand sentence semantics better. Recent research papers in this area discussed that if LSTM has been used as a combination with word embedding models, it gives a significant improvement in results (Deepak and Chitturi, 2020, Anand et al., 2021). Therefore, integrating LSTM with BERT can increase predictions even further, which implies that the proposed model has a better understanding of semantic meaning. Due to the bidirectional nature of BERT, [CLS] is encoded using a multi-layer encoding procedure that includes all representative information of all tokens and [CLS] serves as a “collective representation” for classification tasks. Therefore, for the classification task, the embeddings corresponding to the [CLS] token can be used to represent the whole sentence and fed to the classifier. The classifier has been built from scratch. A feed-forward linear layer with a size of 128 has been included in the classifier. A batch normalization layer has been used to standardize the inputs. Then, a dropout layer with a rate of 0.6 was added to avoid overfitting. Two feed-forward layers with an output size of 2 were added, which classifies the input news as fake or real. A threshold of 0.9 is chosen to one cell’s output over the other. The architecture of this methodology is depicted in Fig 4.

4. Results & Analysis

4.1. Dataset

In this paper, the FakeNewsNet dataset has used for training and testing purposes. The dataset consists of two sub-datasets, named PolitiFact and GossipCop. It was created by utilizing news content from fact-checking websites Politifact (Shu et al., 2020) and GossipCop (Shu et al., 2020). Politifact is a website for fact-checking political news articles labelled as fake or real. GossipCop fact checks news articles related to entertainment and labels them as fake or real. The overview of the data points present in the FakeNewsNet dataset can be seen in Table 1. Many fake news articles contain linguistic patterns that are unique to fake news. Based on this content-based classification on news titles has been opted. News titles were extracted from both datasets. The PolitiFact dataset has 432 data points that are labelled and 624 data points that are labelled Real. The GossipCop dataset has 5323 data points labelled Fake and 16817 data points labelled Real.

Table 1
No. of news articles in the dataset

Politifact		GossipCop	
Fake	**Real**	**Fake**	**Real**
432	624	5323	16817

In this section, the experimental results of training the models on the dataset and comparison of the proposed model with other classification models have been done. Training of the models was done on 80% of the data points selected at random and tested on 20% of the data. The performance of the models has been evaluated on various evaluation criteria like accuracy, precision, recall and F1 score. Comparison of the various models have been done, relevant comparison graphs, tables and confusion metrics of the proposed model were evaluated and visualized.

$$Accuracy = Number\ of\ instances\ classified\ correctly/Total\ number\ of\ instances \tag{7}$$

$$Precision = True\ Positives/(True\ Positives + FalsePositives) \tag{8}$$

$$Recall = True\ Positives/(True\ Positives + False\ Negatives) \tag{9}$$

$$F1\ Score = 2 * (Precison * Recall)/(Precision + Recall) \tag{10}$$

The proposed methodology performance has been compared and analyzed with the baseline models for Fake News Detection. Table 2 depicts the results of comparative analysis using PolitiFact and GossipCop datasets.

- **TCNN-URG (Qian et al., 2018):** It is a two-level convolutional neural network, which is used to train representations from news articles, and a conditional variational auto-encoder is used to capture features from user comments.
- **LIWC (Pennebaker et al., 2015):**Linguistic Inquiry and Word Count (LIWC) is a method for extracting lexicons that fall into psycholinguistic categories. It learns a feature vector from the perspective of psychology and deception.
- **CSI (Ruchansky et al., 2017):** CSI is a deep learning model that combines text, reaction, and source information. The news is represented using LSTM neural network with the Doc2Vec embedding on the news contents and user comments as input.

Table 2
Comparison of models for PolitiFact and GossipCop

Models	Dataset: PolitiFact (FakeNewsNet)				Dataset: GossipCop (FakeNewsNet)			
	Accuracy (%)	Precision	Recall	F1 Score	Accuracy (%)	Precision	Recall	F1 Score
TCNN-URG	71.20	0.71	0.94	0.81	73.60	0.71	0.52	0.60
LIWC	76.90	0.84	0.79	0.81	73.60	0.75	0.50	0.57
CSI	82.70	0.84	0.89	0.87	77.20	0.73	0.63	0.68
HAN	83.70	0.82	0.89	0.86	74.20	0.65	0.68	0.67
SAFE (Multimodal)	87.40	0.88	0.90	0.89	83.80	0.85	0.93	0.89
BERT	**86.25**	**0.90**	**0.87**	**0.88**	**83.00**	**0.89**	**0.89**	**0.89**
BERT + LSTM	**88.75**	**0.91**	**0.90**	**0.90**	**84.10**	**0.89**	**0.91**	**0.89**

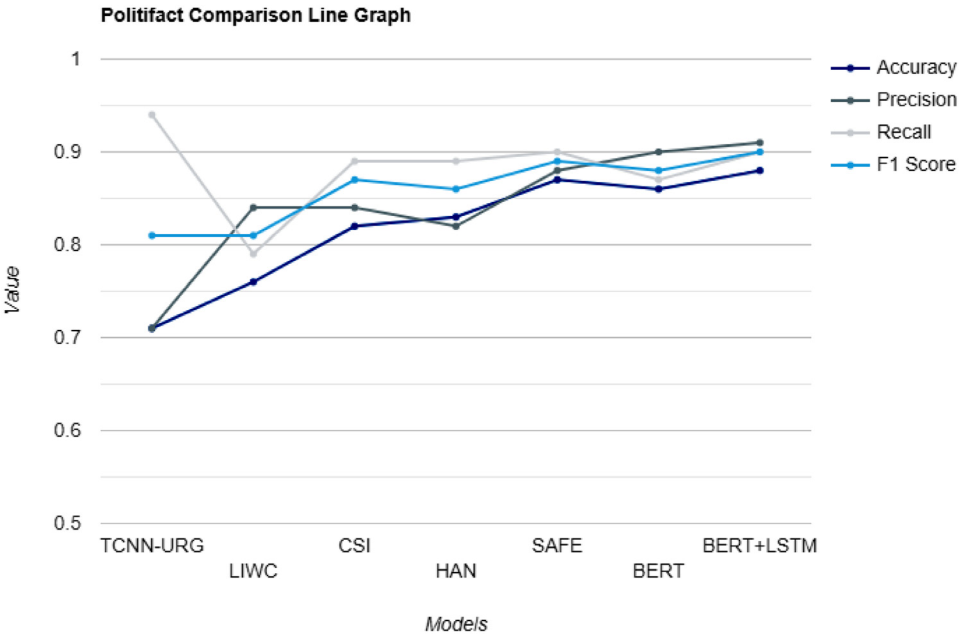

Fig. 5. PolitiFact Comparison Line Graph

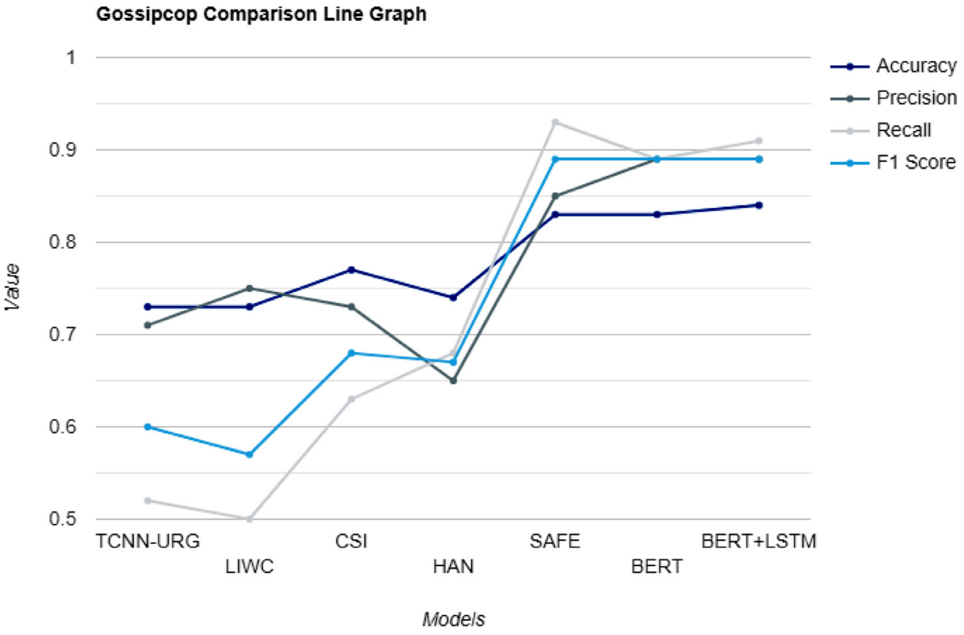

Fig. 6. GossipCop Comparison Line Graph

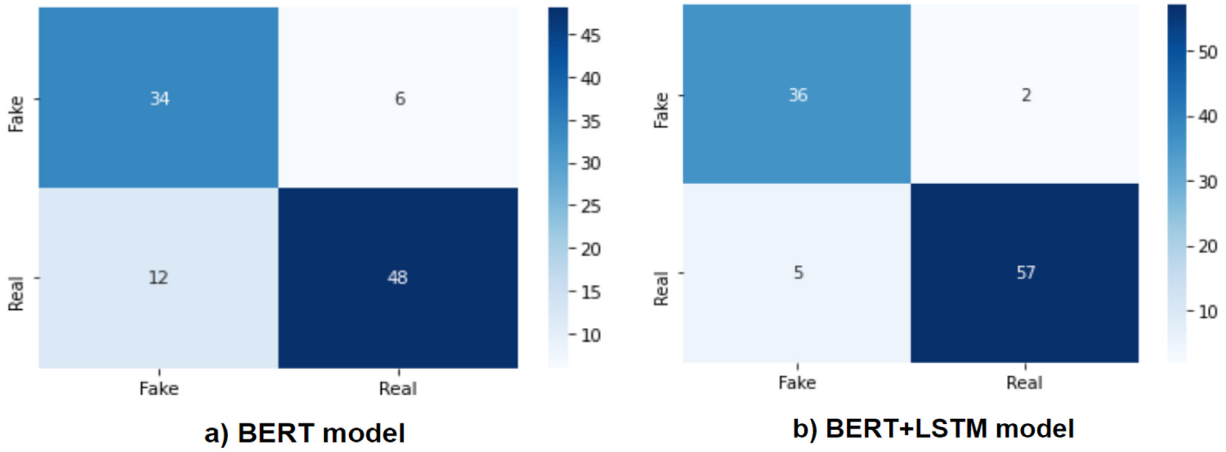

Fig. 7. Confusion Matrix of PolitiFact

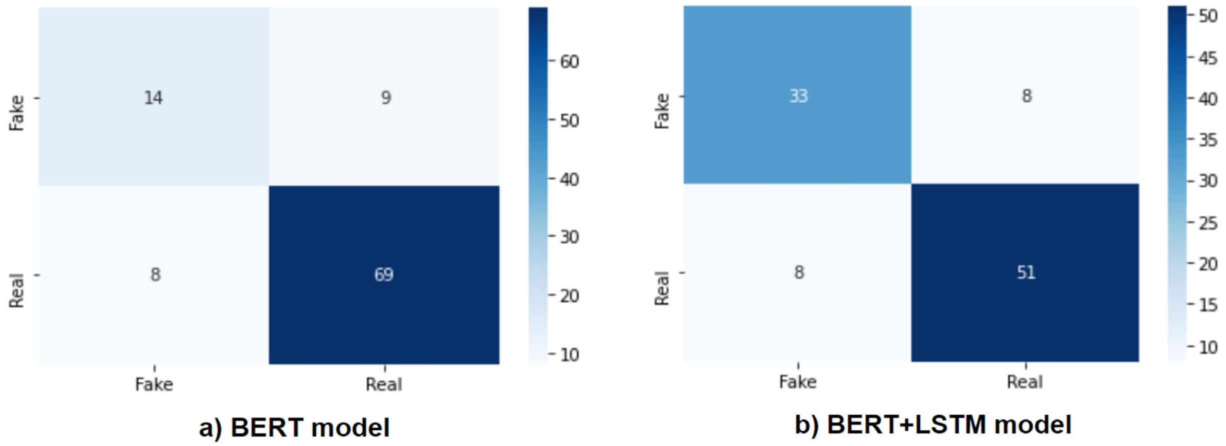

Fig. 8. Confusion Matrix of GossipCop

- **HAN** (Yang et al., 2016)**:** For false news detection, HAN applies a hierarchical attention neural network structure to news material. It uses word-level attention in each sentence and sentence-level attention in each text to encode news material.
- **SAFE(Multi-modal)** (Zhou et al., 2020)**:** Text-CNN is extended in Similarity-Aware Fake (SAFE) news detection by adding fully connected layer that automatically extracts textual information for each news article. A convolutional layer and a maximum pooling layer are included. Each word is first embedded in a piece of material with several words, and then a convolutional layer has been utilized to create a feature map from a succession of local inputs via a filter.

The proposed model has been implemented using (BERT+LSTM) on FakeNewsNet (PolitiFact and GossipCop) dataset and the results have been compared with baseline models such as TCNN-URG (Aggarwal et al., 2020), LIWC (Wani et al., 2021), CSI (Qian et al., 2018), HAN (Pennebaker et al., 2015), SAFE(Multimodal) (Ruchansky et al., 2017) and BERT. The proposed model achieved maximum accuracy of 88.75% as compared to other models. The proposed model got an increment of a minimum of 1.35% and a maximum of 17.55% accuracy for baseline models in the PolitiFact dataset. Also, an increment of a minimum of 0.3% and a maximum of 10.5% is seen in accuracy when compared to baseline models in the GossipCop dataset Fig. 6 and Fig. 7(a-b) shows a pictorial representation of the evaluated metrics over PolitiFact and GossipCop respectively.

A confusion matrix has been visualized for the correctness of predictions made by the model on the test set. By definition, C is a confusion matrix where $C_{i,j}$ is equal to the number of data points that belong to class i and are predicted to be in class j Fig. 8(a-b) represent the confusion matrix of the proposed model over PolitiFact and GossipCop datasets respectively.

5. Conclusion

Fake News information is broadcast to the public through mainstream sources such as media outlets and social media and is often deceptive with the intent of influencing public opinion in one's favour. The importance of fake news classification in the modern-day and work done towards the same has been discussed in the paper. A content-based classification model which classifies news as fake or real based on news titles has been proposed. To classify the news titles, BERT with an LSTM layer was utilized as the classification model. To accomplish this objective, BERT with an LSTM layer is used as the classification model. BERT can learn contextualized word representations from a wide number of unlabeled text datasets. BERT performed well in NLP testing due to its complex structure and great nonlinear representation learning capability. LSTMs effectively improve performance by memorizing and finding the pattern of crucial information. Contextualized word representations from BERT could be employed in LSTM to improve false news classification performance because of their high ability to capture semantics and long-distance relationships in news titles. It has been compared with other classification methods and a vanilla BERT model. A slight improvement has been seen with the proposed model which is indicative of the model learning linguistic patterns of news titles and their connection with fake news. A few setbacks that the model may have faced is the thin line between fake news and real news titles. Often, the titles do not appear to have any difference as the writers of fake news begin to use language similar to that used by real news. To overcome this problem, the news title has to be manually fact-checked, which may be a prospect. The performance of deep learning models increases with more data and more data, in this case, would mean more instances where fake news titles can be seen and a comprehensive study on the language used in it can be done. Social media is a mass spreader of fake news where fake news is often tweeted and shared multiple times. Future work may include training the model on linguistic and lexical patterns of fake news as seen on social media sites. This architecture can be tested in the future on a variety of application domains, and it may even improve existing benchmarks. The proposed model can be studied and tested with various set-ups in the hopes of achieving greater performance than the current state. We also intend to tune the hyperparameters of both the BERT and following layers, as well as to conduct a thorough analysis of their impacts.

Compliance with Ethical Standards

The authors declare that they do not have any conflict of interest. This research did not involve any human or animal participation. All authors have checked and agreed on the submission.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Aggarwal, A., Chauhan, A., Kumar, D., Mittal, M., & Verma, S. (2020). Classification of fake news by fine-tuning deep bidirectional transformers-based language model. *EAI Endorsed Transactions on Scalable Information Systems, 7*(27).

Aggarwal, A., Chauhan, A., Kumar, D., Mittal, M., Roy, S., & Kim, T. H. (2020). Video caption based searching using end-to-end dense captioning and sentence embeddings. *Symmetry, 12*(6), 992.

Agarwal, A., Mittal, M., Pathak, A., & Goyal, L. M. (2020). Fake news detection using a blend of neural networks: an application of deep learning. *SN Computer Science, 1*(3), 1–9.

Albahar, M. (2021). A hybrid model for fake news detection: Leveraging news content and user comments in fake news. *IET Information Security*.

Anand, I., Negi, H., Kumar, D., Mittal, M., Kim, T. H., & Roy, S. (2021). Residual U-Network for Breast Tumor Segmentation from Magnetic Resonance Images. *Computers Materials & Continua, 67*(3), 3107–3127.

Deepak, S., & Chitturi, B. (2020). Deep neural approach to Fake-News identification. *Procedia Computer Science, 167*, 2236–2243.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Eke, C. I., Norman, A. A., & Shuib, L. (2021). Context-Based Feature Technique for Sarcasm Identification in Benchmark Datasets Using Deep Learning and BERT Model. *IEEE Access, 9*, 48501–48518.

Ezeakunne, U., Ho, S. M., & Liu, X. (2020). Sentiment and retweet analysis of user response for fake news detection. In *Proceedings of the 2020 International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation*.

Farooq, G. (2017). Politics of Fake News: how WhatsApp became a potent propaganda tool in India. *Media Watch, 9*(1), 106–117.

Goyal, L. M., Mittal, M., & Sethi, J. K. (2016). Fuzzy model generation using Subtractive and Fuzzy C-Means clustering. *CSI transactions on ICT, 4*(2-4), 129–133.

Granik, M., & Mesyura, V. (2017). Fake news detection using naive Bayes classifier. In *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)* (pp. 900–903). IEEE.

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 US presidential election. *Science, 363*(6425), 374–378.

Hopkin, J., & Rosamond, B. (2018). Post-truth politics, bullshit and bad ideas:'Deficit Fetishism'in the UK. *New political economy, 23*(6), 641–655.

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

Kadam, A. B., & Atre, S. R. (2020). Negative impact of social media panic during the COVID-19 outbreak in India. *Journal of travel medicine, 27*(3), taaa057.

Kaliyar, R. K., Goswami, A., & 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.

Li, Q., & Zhou, W. (2020). Connecting the Dots Between Fact Verification and Fake News Detection. *arXiv preprint arXiv:2010.05202*.

Lin, Jun & Tremblay-Taylor, Glenna & Mou, Guanyi & You, Di & Lee, Kyumin. (2019). Detecting Fake News Articles. 3021-3025. 10.1109/BigData47090.2019.9005980.

Luengo, M., & García-Marín, D. (2020). The performance of truth: politicians, fact-checking journalism, and the struggle to tackle COVID-19 misinformation. *American Journal of Cultural Sociology, 8*(3), 405–427.

Mehta, D., Dwivedi, A., Patra, A., & Kumar, M. A. (2021). A transformer-based architecture for fake news classification. *Social Network Analysis and Mining, 11*(1), 1–12.

Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., & Joulin, A. (2018). Advances in Pre-Training Distributed Word Representations. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* Opgehaal van https://aclanthology.org/L18-1008 .

Mittal, M., Kaur, I., Pandey, S. C., Verma, A., & Goyal, L. M. (2019). Opinion Mining for the Tweets in Healthcare Sector using Fuzzy Association Rule. *EAI Endorsed Transactions on Pervasive Health and Technology, 4*(16).

Moravec, P., Minas, R., & Dennis, A. R. (2018). *Fake news on social media: People believe what they want to believe when it makes no sense at all* (pp. 18–87). Kelley School of Business Research Paper.

O'Connor, C., & Murphy, M. (2020). Going viral: doctors must tackle fake news in the covid-19 pandemic. *bmj, 369*(10.1136).

Oriola, O. (2022). Exploring N-gram, Word Embedding and Topic Models for Content-based Fake News Detection in FakeNewsNet Evaluation. *International Journal of Computer Applications, 975*, 8887.

Paskin, D. (2018). Real or fake news: who knows? *The Journal of Social Media in Society, 7*(2), 252–273.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015.

Prasetijo, A. B., Isnanto, R. R., Eridani, D., Soetrisno, Y. A. A., Arfan, M., & Sofwan, A. (2017). Hoax detection system on Indonesian news sites based on text classification using SVM and SGD. In *2017 4th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)* (pp. 45–49). IEEE.

Qian, F., Gong, C., Sharma, K., & Liu, Y. (2018). Neural User Response Generator: Fake News Detection with Collective User Intelligence. *In IJCAI, 18*, 3834–3840.

Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., & Choi, Y. (2017). Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 2931–2937).

Ruchansky, N., Seo, S., & Liu, Y. (2017). Csi: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 797–806).

Singh, A., Saimbhi, A. S., Singh, N., & Mittal, M. (2020). DeepFake Video Detection: A Time-Distributed Approach. *SN Computer Science, 1*(4), 1–8.

Sundermeyer, M., Schlüter, R., & Ney, H. (2012). LSTM neural networks for language modeling. *Thirteenth annual conference of the international speech communication association*.

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2020). Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data, 8*(3), 171–188.

Shu, K., Cui, L., Wang, S., Lee, D., & Liu, H. (2019). defend: Explainable fake news detection. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 395–405).

Shu, K., Mahudeswaran, D., Wang, S., & Liu, H. (2020). Hierarchical propagation networks for fake news detection: Investigation and exploitation. *Proceedings of the International AAAI Conference on Web and Social Media, 14*, 626–637.

Singhal, S., Kabra, A., Sharma, M., Shah, R. R., Chakraborty, T., & Kumaraguru, P. (2020). Spotfake+: A multimodal framework for fake news detection via transfer learning (student abstract). *Proceedings of the AAAI Conference on Artificial Intelligence, 34*(10), 13915–13916.

Törnberg, P. (2018). Echo chambers and viral misinformation: Modeling fake news as complex contagion. *PloS one, 13*(9), Article e0203958.

Walia, I. S., Kumar, D., Sharma, K., Hemanth, J. D., & Popescu, D. E. (2021). An Integrated Approach for Monitoring Social Distancing and Face Mask Detection Using Stacked ResNet-50 and YOLOv5. *Electronics, 10*(23), 2996.

Wani, A., Joshi, I., Khandve, S., Wagh, V., & Joshi, R. (2021). Evaluating deep learning approaches for covid19 fake news detection. In *International Workshop on Combating On line Ho st ile Posts in Regional Languages during Emergency situation* (pp. 153–163). Cham: Springer.

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016). Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies* (pp. 1480–1489).

Zhou, X., Wu, J., & Zafarani, R. (2020). Similarity-Aware Multi-modal Fake News Detection. *Advances in Knowledge Discovery and Data Mining, 12085*, 354.