Early Detection of Fake News "Before It Flies High"

Fantahun Bogale Gereme
Lab of Granular Computing & AI
Institute of Fundamental & Frontier Sciences
University of Electronic Science and Technology of China
Chengdu, China
fantishb@gmail.com

William Zhu
Lab of Granular Computing & AI
Institute of Fundamental & Frontier Sciences
University of Electronic Science and Technology of China
Chengdu, China
wfzhu@uestc.edu.cn

ABSTRACT
Currently, social media for news consumption is preferred over the conventional media and attracted many people due to its low cost, easy access, simplistic way of commenting & sharing, more timely nature, and rapid information sharing capabilities. On the other hand, it aggravates the prompt and wide spreading of fake news. Fake news may be fabricated for the purpose of, commercial gain, political propaganda, seeking attention, and intent of defamation. Interest of individuals, and various groups to influence events and policies around the globe is the other reason for fake news generation and dissemination. The extensive spread of fake news is progressively becoming a threat to individuals and society as a whole. It disrupts the authenticity balance of the news ecosystem; induces biased or false beliefs into consumers; creates real-life fears in the society and threatens freedom of speech, freedom of the press and democracy. The craving to mitigate the undesirable effects of fake news, recently makes fake news detection on social media an emerging research area attracting tremendous attention. Following this warm concern, various researches have been conducted and showed promising results. In this work, we propose a model for early detection of fake news using deep learning, and news content. Deep learning and heterogeneous dataset has been used to create a more generic model that could perform better in the real world. We conducted experiments on two real world datasets and a third dataset which is obtained by combining the two datasets and randomly shuffled them. Our experiment results have shown that early detection of fake news using news content and deep learning models, without waiting for news propagation, is achievable and should be given better attention to combat fake news effectively before it proliferates and misleads many people. The experimental results obtained are interesting.

CCS Concepts
• Computing methodologies→Machine learning • Computing methodologies→Natural language processing

Keywords
Fake News Detection; Social Media; Machine Learning; Deep Learning; Natural Language Processing.

ICBDT2019, August 28–30, 2019, Jinan, China

ACM ISBN 978-1-4503-7192-6/19/08…$15.00
https://doi.org/10.1145/3358528.3358567

1. CONTEXT
These days, social media is becoming the preferred way of news consumption over the traditional news transmission means. This may be attributed to its low cost, ease of access, simplistic way of commenting & sharing, more timely nature, and rapid dissemination of information [16], [22], [31]. Considering the well-known social media Facebook and Twitter for example, and analyzing the situation could simply witness the wide spread use of the same. In the United States, the number of adults who got news from social media has grown from 49% in 2012 to 62% in 2016 [4], [16]. An arbitrary user of social media with no track record or reputation can sometimes reach as many readers as CNN, BBC or the New York Times [4]. Hence, we don't need a hard proof to recognize the undeniable wide spread use and influence of social media. On the other hand, it becomes a suitable carrier for rapid and wide dissemination of fake news, that is, fabricated news with intentionally false information. The most popular fake news stories were more widely shared on social media than the most popular mainstream news stories [4]. It was estimated that over a million tweets were connected to fake news by the culmination of the 2016 US presidential election [16], [23]. In January 2017, a spokesman for the German government stated that they "are dealing with a phenomenon of a dimension that have not seen before", referring to the propagation of fake news [12], [13].

The reasons for crafting and disseminating fake news are varied. Sometimes it is created for commercial interest, to attract viewers and collect advertising revenue [12], [17], [21], [23], making political propaganda [17], [19], [21], [23], intent of defamation [23], influence events and policies around the world [21], seeking attention etc. [23]. The extensive spread of fake news is progressively becoming a threat to individuals and our society as a whole [16], [19], [21], [26].

(1) Fake news can disrupt the authenticity balance of the news ecosystem. For example, during the 2016 U.S. presidential election, the most popular fake news was even more extensively spread on Facebook than through the most popular and authentic mainstream news agents [16]. In a December poll, 64 percent of U.S. adults said that fake news has caused a huge confusion about the facts of existing events [27].

(2) Fake news deliberately persuades consumers to accept biased or false beliefs; it is usually manipulated by propagandists to take unfair political advantage. For example, some report shows that Russia has created bogus accounts and social bots to spread false stories [16].

(3) Fake news changes the way people understand and respond to real news. For example, some fake news was just created to trigger

people's distrust and make them confused, impeding their abilities to differentiate what is true from what is not[1].

(4) Fake news has created real world, real-life fears; a man carried an AR-15 rifle and walked in a Washington DC Pizzeria, because he recently read online that "this pizzeria was harboring young children as sex slaves as part of a child abuse ring led by Hillary Clinton"[2] [9].

(5) Fake news is now regarded as one of the ultimate threats to democracy, journalism, and freedom of expression [33]. It has played its role in weakening public trust in governments and news sources. These are only few example to explain the serious nature of fake news in negatively impacting individuals, events, institutions, societies and government systems around the globe.

To help mitigate the undesirable effects caused by fake news, both to benefit individuals, the public, governments and the news ecosystem, developing methods to automatically detect fake news on social media has become critical. This strong desire recently makes fake news detection on social media to be an emerging research area that is attracting tremendous attention. Various researches have been made with the intention of fake news detection. Some tried to use linguistic cues to detect fake news while others used the opinion of the crowd, source verification and social context methods. In addition to individual features few combinations of them have also been tried to serve the same purpose. Detail explanation is provided in the next section. But due to the challenging nature of the task, and the nascence of the phenomenon, the problem of fake news detection has not been yet sufficiently addressed. Further work has to be done to satisfy the problem of fake news detection.

Moreover, fake news detection is not sufficient, early detection of fake news is mandatory. Winston Churchill said, "A lie gets halfway around the world before the truth has a chance to get its pants on." For early detection and removal of fake news, we need to focus on the *content of the news* giving lesser attention to other auxiliary perspectives [33], because others like propagation based techniques should wait sometime for the news to propagate across the network to see the behavior.

The rest of this document is organized as follows. In section 2, we have presented literature. Our research methodology is explained in section 3. Section 4 is dedicated for experiments and evaluation while section 5 concludes our work.

2. LITERATURE

Although limited in number and the methods they used, there are researches made on fake news detection. Fake news detection techniques were classified by several researchers in different ways based on different criteria. For example, [23] classified fake news detection methods as detection via experts' verification, detection based on computational techniques and detection based on crowd sourcing. On the other hand, [16] categorized detection models as news content based and social context models and discuss several sub techniques under each. The text, the response, the source of the news, and the combinations of the above were considered by [19] to be the basic characteristics for categorizing detection methods. Considering the literature, to our best understanding, we can categorize the major existing methods as follows.

2.1 Content Based Methods

Content based methods are based on the content of the news itself. These methods mostly concentrate on mining specific linguistic hints, such as identifying anomalous patterns of pronouns, conjunctions, and words associated with sensitive, negative emotional word usage [8], [25], [28]. Fake news often contains an inflated number of swear words and personal pronouns [3].

Some have combined linguistic methods with machine learning techniques to label an article as factual or fabricated [3], [6], [24], [27], [30], [32]. According to [29], there are a variety of fake news, each with different potential textual indicators. Thus, existing works are designed upon hand-crafted features which are laborious and highly dependent on the specific dataset and the obtainability of domain knowledge to design appropriate features. To expand beyond the specificity of hand-crafted features, [15] proposed a model based on recurrent neural networks (RNNs) that uses mainly linguistic features; the CSI (Capture, Score and Integrate) model proposed by [19] captures text, response and user characteristics.

Moreover, fake content is not always generated in the form of text only format. Advances in graphics editing and manipulation tools have made it considerably easier to generate fake imagery [18]. Huh et al. [18] proposed a model that learns to detect visual manipulations from unlabeled data through self-supervision.

2.2 Social Context Models

The nature of social media provides researchers with additional resources to complement and enhance News Content Models. Social context models include pertinent user social engagements in the analysis, capturing this auxiliary information from a variety of perspectives. Social context modeling approaches can possibly be classified into two categories: Stance-based and Propagation-based [16].

2.2.1 Stance-based

Stance detection is the task of automatically determining from a post whether the user is for, against or neutral towards some target entity, cause, or idea [11], [20].

2.2.2 Propagation-based

Propagation-based approaches for fake news detection reason about the interrelations of relevant social media posts to predict news trustworthiness [16]. The basic assumption is that the credibility of a news event is highly related to the credibility of relevant social media posts.

2.3 Source Reliability and Trustworthiness Checking Methods

The reliability of information depends heavily on the trustworthiness of the sources. Several user based features were utilized in spam detection and credibility assessment strategies. These features relate to information that can be extracted from the profile of a user. Features that are prevalent in various strategies include the age of the account, the number of tweets, the number of followers, the number of followees, or the ratio of followers to followees [1], [2], [5], [10], features related to the user profile description [5], [6], the registration age of the user, whether the account is a verified account and the presence of a URL on a user's profile [1], [2], [6].

[1] https://www.nytimes.com/2016/11/28/opinion/fakenews-and-the-internet-shell-game.html? r=0

[2] ttps://www.nytimes.com/2016/12/05/business/media/comet-ping-pong-pizza-shooting-fake-news-consequences.html

2.4 Crowd Sourcing Methods

Crowdsourcing-oriented methods exploit the "wisdom of the crowd" to enable ordinary people to annotate news content. Individual annotations are then aggregated to produce an overall assessment of the news veracity [16]. Facebook [9] recently introduced tools that enable users to flag fake news. Fiskkit[3] allows users to discuss and annotate the accuracy of specific parts of a news article. An anti-fake news bot named "For real" is a public account in the instant communication mobile application LINE[4], which allows people to report suspicious news content which is then further checked by editors.

2.5 Summary

In general, most existing approaches emphasize on extracting features manually, putting these features into traditional supervised classification models, such as naive Bayes, decision tree, logistic regression, k-nearest neighbor (KNN), and support vector machines (SVM), and then picking the classifier that achieves the best result. However, traditional classifiers alone are not strong enough to address the accuracy level the problem demands [16]. Deep neural networks like LSTM and Convolutional Neural Networks has to be exploited more for the accurate prediction of fake news. Moreover, detection of fake news is not sufficient; it has to be done earlier possibly as soon as its release time. For this purpose, researches should focus on the content of the news and less focus on other auxiliary features that are mostly obtained after the proliferation of the fake news and affect many people [33].

3. METHODOLOGY

3.1 Problem Definition

The term "fake news" is used in different contexts in various research works. Some considered hoaxes, and satires as fake news though satire is often entertainment-oriented. Concepts related to fake news are summarized in Table-2 adopted from [33] based on their authenticity (false or not), intention (bad or not) and whether they are news or not.

Table 1. Concepts Correlated with Fake News

Concepts	Authenticity	Intention	News?
Malicious false news	False	Bad	Yes
False news	False	Unknown	Yes
Satire news	Unknown	Not Bad	Yes
Disinformation	False	Bad	Unknown
Misinformation	False	Unknown	Unknown
Rumor	Unknown	Unknown	Unknown

Broadly speaking, fake news can be defined as false news [33]. But false news can be generated maliciously from bad intention or without planned bad intention. So adding the intention element clearly defined gives us the narrow definition of fake news. In our work, we define fake news in line with [4], [16], and [33]. Fake news is defined as news articles that are intentionally and verifiably false.

We adopt the definition of [16] for the task of fake news detection. The details of mathematical formulation of fake news detection on social media, the definition of key components of fake news, the formal definition of fake news detection and the basic notations are defined as follows,

Let α represent a News Article with two major components: Publisher and Content. Publisher P[α] includes a set of profile features to describe the original author, such as name, domain, age, among other attributes. Content C[α] consists of a set of attributes that represent the news article and includes headline, body text, image, etc.

We define Social News Engagements as a set of tuples E={eit} to represent the process of how news propagates over time among n set of users depicted as $U = \{u_1, u_2, \ldots, u_n\}$ and their corresponding posts $P = \{ p_1, p_2, \ldots, p_n \}$ on social media regarding news article α. Each engagement eit = $\{u_i, p_i, t\}$ represents that a user ui spreads news article α using pi at time t. Note that we set t=Null if the article α does not have any engagement yet and thus ui represents the publisher.

Fake News Detection: Given the social news engagements E among n users for news article α, the task of fake news detection is to predict whether the news article α is a fake news piece or not, that is, $F : E \rightarrow \{0, 1\}$ such that,

$$\mathcal{F}(\alpha = \begin{cases} 1, & if\ \alpha\ is\ a\ fake\ article \\ 0, & otherwise \end{cases} \tag{1}$$

where $\mathcal{F}$ is the prediction function we are interested to learn.

3.2 Model Building Methods.

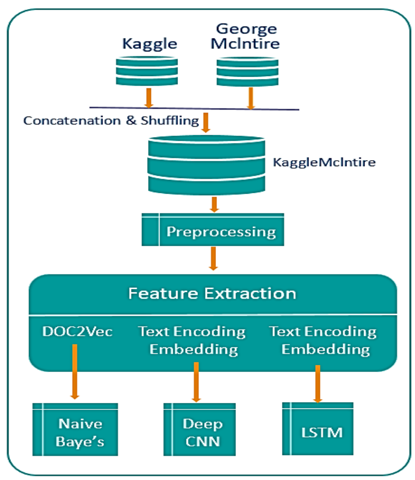


Figure 1. Our General Model for Early Detection of Fake News

Several different models were implemented; we provided a precise description of each including the base models.

3.2.1 Recurrent Neural Networks

The recurrent neural network (RNN) is a family of neural networks characterized by a graph of computational units forming a feedback loop. The internal state can be seen as a memory aiming to capture temporal dynamics in the input. When processing a text document, the internal state of the RNN allows holding information about the whole document. The basic RNN unit is described by the recursive function

$$h_t = f(W^h x_t + U^h h_{t-1} + b^h) \tag{2}$$

[3] http://fiskkit.com

[4] https://grants.g0v.tw/projects/588fa7b382223f001e022944

where f is an activation function, x_t is the input at time step t, h_{t-1} is the hidden state calculated at the previous time step and h_t is the new hidden state. The parameters of the RNN unit, weight matrices, W_h and U_h and bias vector b_h, are learned through backpropagation through time.

3.2.2 Long Short-Term Memory (LSTM)

LSTM is a variant of the RNN architecture aiming to evade the problem of vanishing or exploding gradients during training when backpropagating the errors through each time step of a sequence. LSTM uses so called long short-term memory cells in its internal state initially proposed by Hochreiter and Schmidhuber[14]. The idea of an LSTM unit is that it has a memory content and gates that regulate how much of the input is added to this and if to forget current memory content. The basic LSTM equation is shown below:

$$i_t = \sigma(W^i x_t + U^i h_{t-1} + b^i) \quad (3)$$

$$f_t = \sigma(W^f x_t + U^f h_{t-1} + b^f) \quad (4)$$

$$o_t = \sigma(W^o x_t + U^o h_{t-1} + b^o) \quad (5)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W^c x_t + U^c h_{t-1} + b^c) \quad (6)$$

$$h_t = o_t \odot \tanh(c_t) \quad (7)$$

where $\odot$ is the elementwise multiplication, σ is the sigmoid function and tanh is the hyperbolic tangent function. The memory of the LSTM unit at time step t is represented by c_t and the output is h_t. The input gates i_t, the forget gates f_t, and the output gates o_t, take values in the range of 0 to 1 representing how much information they let through at time step t. W* and U* are weight matrices and b* is a bias vector.

3.2.3 Convolutional Neural Network (CNN)

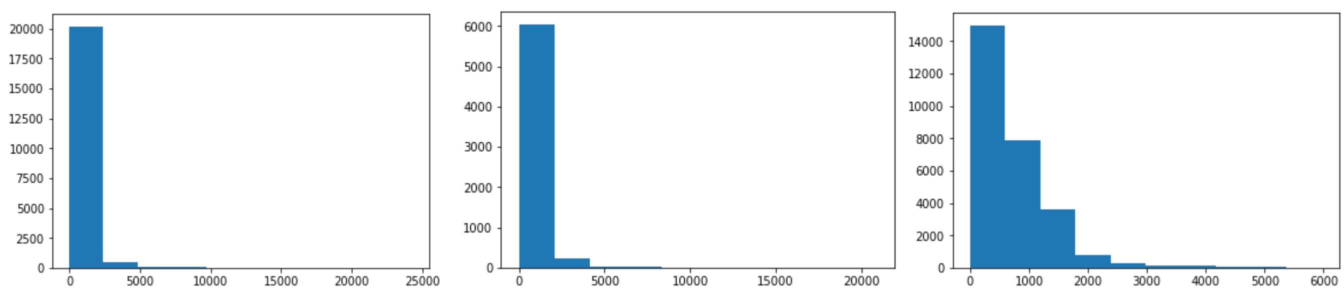
In machine learning, a convolutional neural network (CNN, or ConvNet) is a class of deep, feed-forward artificial neural networks, most commonly applied to analyzing visual imagery. CNNs use a variation of multilayer perceptrons designed to require minimal preprocessing. A CNN consists of an input and an output layer, as well as multiple hidden layers. The hidden layers of a CNN typically consist of convolutional layers, pooling layers, fully connected layers and normalization layers. Even though CNNs were most famous in analysis of visual imagery, they are currently being used in NLP as well.

3.2.4 Very Deep Convolutional Networks (VDCNN)

Very Deep Convolutional Networks (VDCNN) is a novel architecture presented by Conneau et al., [7] for text processing which operates directly at the character level and uses only small convolutions and pooling operations. They have showed that the performance of this model increases with the depth.

3.3 Data Set

The fake news detection problem has no agreed upon publicly available benchmark dataset [16]. The available datasets do have their own limitations. Early detection of fake news should focus on content based techniques giving little attention to other auxiliary features [33]. For this, collecting news samples from different domains, topics and websites plays significant role in maintaining generality of deceptive writing styles. For the purpose of detecting fake news at its early stage, we are going to use the text content of news articles.



(a)Kaggle (b) **Geaorge Mcnltire** (c) **KaggleMclntire**

Figure 2. Histogram of the number of articles against word count in each article, where the vertical axis shows the number of articles and the horizontal axis shows word count in each article.

Table 2. Dataset composition

Dataset	Source	#Fake News	#Real News	Total
Kaggle	Scraped from 244 websites by BS-Detecter	10,349	10,369	20,718
George Mclntire dataset	The New York Times, WSJ, Bloomberg, NPR, the Guardian and Kaggle	3,151	3,160	6,311
KaggleMclntire	The union of Kaggle and Mcnltire sources	13,500	13,529	27,029

Most of the researches have used specific publicly available datasets. Using a combination of datasets from different sources, topic and domains can help train a more generic model that could perform well in the real world [16]. For this reason, two well-known datasets, Kaggle dataset, and George Mclntire dataset by George Mclntire were used. We concatenated both datasets and shuffled them randomly to create the combined dataset. This combined dataset as illustrated in Table-2, came from diversified sources and is heterogeneous. For reference we named the combined dataset as KaggleMclntire dataset. Figure 2 depicts the relationship between the number of articles and the word count in each article for the three datasets.

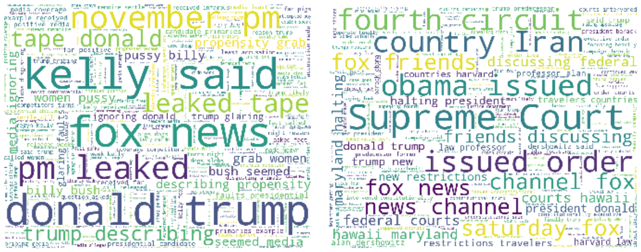


Figure 3. Word cloud for (a) Fake and (b) real news of the dataset

4. EXPERIMENTS AND EVALUATION

4.1 Evaluation Metrics

A common performance metric used when evaluating the performance of a classifier on a balanced dataset is accuracy (the proportion of correctly classified instances). However, accuracy is not a suitable metric for evaluating a classifier in a domain with an imbalance in class distribution since it tells little about the classifier's performance on minority classes when the classifier achieves high accuracy being biased towards majority classes. Hence, in an imbalanced domain, it is more common to analyze the confusion matrix, Table-3, reporting four different classes.

Table 3. Confusion matrix for a binary-class classification problem.

	Positive Prediction	**Negative Prediction**
Positive Class	True Positive (TP)	False Negative (FN)
Negative Class	False Positive (FP)	True Negative (TN)

TP: the number of positive instances that a classifier correctly predicted as positives, (When predicted fake news pieces are actually annotated as fake news)

FN: the number of positive instances that a classifier incorrectly predicted as negatives, (When predicted true news pieces are actually annotated as fake news)

TN: the number of negative instances that a classifier correctly predicted as negatives and (When predicted true news pieces are actually annotated as true news)

FP: the number of negative instances that a classifier incorrectly predicted as positives. (When predicted fake news pieces are actually annotated as true news)

By formulating this as a classification problem, we can define the following metrics:

$$accuracy = \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|} \quad (8)$$

$$precision = \frac{|TP|}{|TP| + |FP|} \quad (9)$$

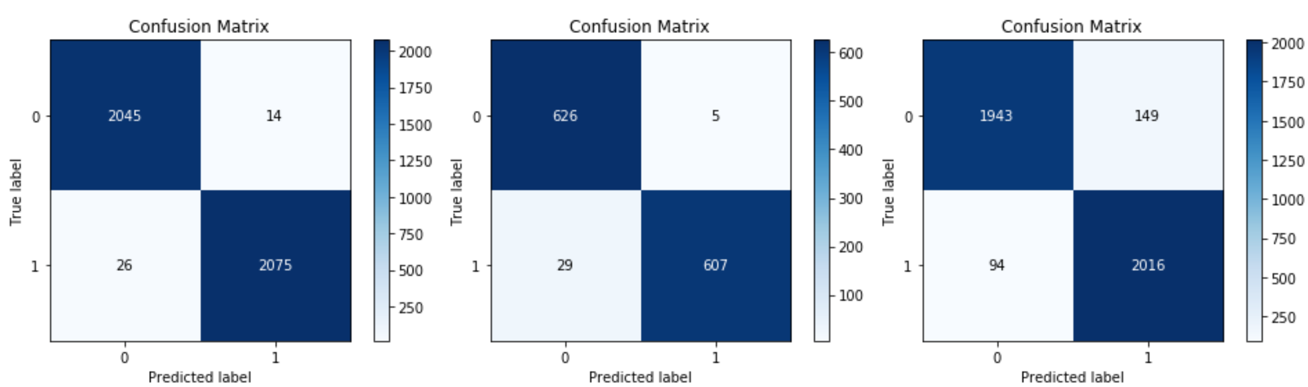
$$recall = \frac{TP}{|TP| + |FN|} \quad (10)$$

$$F1 = 2 . \frac{precision \,.\, recall}{precision + recall} \quad (11)$$

Precision measures the fraction of all detected fake news that are annotated as fake news, addressing the important problem of identifying which news is fake. However, because fake news datasets are often skewed, a high precision can be easily achieved by making fewer positive predictions. Thus, recall is used to measure the sensitivity, or the fraction of annotated fake news articles that are predicted to be fake news. F1 is used to combine precision and recall, which can provide an overall prediction performance for fake news detection.

4.2 Experimental Analysis

The code was written in Python3.7.0, using TensorFlow r1.10, NumPy and Keras library. As there were some unlabeled samples in both Kagle and George Mclntire, all unlabeled data were dropped. Very short articles of length less than ten were ignored. Stop words were removed. From Tokenizer class of Keras, python deep learning library, the method *texts_to_sequences( )* was used to split the text data with space, filter out punctuation marks, and change all text to lowercase. Sequences were generated, padded and truncated using *pad_sequence( )* method with *maxlen* parameter set to 5000 for convolutional networks and 500 for LSTM. After all preprocessing, each dataset was divided as training and validation set. We used 80% of the data as training set and the remaining 20% was set aside as validation set. Separate experiments were made on each dataset and results were documented properly.



(a) **(b)** **(c)**

Figure 4. Confusion matrix of Deep Convolutional Neural Networks using: (a) Kaggle, (b) Geaorge Mcnltire and (c) KaggleMclntire datasets.

Naïve Baye's, LSTM, and Deep Convolutional networks with max-pooling were trained. Naïve Baye's was selected to be used as a traditional baseline classifier. The performance of each model on the validation set is recorded in Table-4. All scores are recorded based on the evaluation metrics we have defined in Sectiom-4.1.

For convolutional networks, we used an embedding dimension of 200, with 10,000 unique words and 5000 sequence length post padded with zeros. The output of the embedding layer is fed into a dense network of 128 neurons with *ReLU* (Rectified Linear Unit) as activation function whose output is then passed into a one dimensional GlobalMaxPooling layer. The output of this layer is again fed into a dense network of 128 neurons with *ReLU* as activation function whose output is finally passed into a one dimensional dense network with *sigmoid* as activation function. The model is compiled using *rmsprop* as optimization technique and *binary-crossentropy* as the loss function. The model is then fitted with batch_size set to 128 and 10 epochs.

For the LSTM network, we set the maximum input length to be 500 with post padding by zeros. Each input sequence is embedded into 64-dimensional vectors. Each of the embedded inputs are then fed into an LSTM network having 100 neural units. The output of the LSTM network is then passed into a one dimensional dense network with *sigmoid* as activation function. The model is then compiled with *adam* optimizer and *binary_crossentropy* used as a loss function. The model is then fitted with batch_size set to 64 and 100 epochs.

Table 4. Model Performance on the Validation Set

Dataset	**Model**	**Naïve Baye's**	**LSTM**	**Deep ConvNets**
Kaggle	Accuracy	.8964	.9567	.9904
	Recall	-	.9482	.9876
	Precision	-	.9657	.9933
	F1 Score	-	.9569	.9905
George Mclntire	Accuracy	.8982	.9089	.9732
	Recall	-	.9007	.9544
	Precision	-	.9174	.9918
	F1 Score	-	.9090	.9728
KagleMclntire	Accuracy	.8964	.9041	.9422
	Recall	-	.9140	.9555
	Precision	-	.8955	.9312
	F1 Score	-	.9046	.9432

The Deep Convolutional Network model has shown its best performance after running 10 epochs and show no difference upon increasing the number of epochs. The LSTM model was found to be good after running 10 epochs but far better with 100 epochs. Detail result for each model and dataset is provided in Table 4. The performance of Deep Convolutional Networks in text classification makes them the best in the area in addition to their common use in image processing. As per the results, we can say that content based methods of fake news detection are suitable for early detection of fake news and are much promising. The CSI model [19] has scored 95.3% maximum accuracy, while 3HAN [26] and LSTM-1[15] scored accuracy of 96.24% and 89.6% respectively. George Mclntire using his dataset has developed a fake news classification model which score a maximum accuracy of 88%. Our model maximum score using convolutional networks scored accuracy of 99.04% which is better than the others top models. This is just to mention some of the top models, otherwise most of them use other auxiliary features in addition to the content of the news. Our model uses news content only.

5. CONCLUSION

In this paper, we have studied early detection of fake news on social media using deep learning techniques. We have argued, for early detection of fake news, content of the news article has to be given more priority and a generic model can be better trained using a wide range of dataset in terms of source, topic, domain etc. The results have demonstrated that this approach can be very effective. Deep convolutional networks have shown the highest performance witnessing that they are best tools not only for image processing, in which they are well known, but also for natural language processing. Our model using convolutional neural networks has the best performance, *above 99% accuracy*. Further work in organizing more suitable dataset for text and other multimedia contents has to be done. Inclusion of other auxiliary features with more emphasis on content is the other dimension worth exploring.

6. REFERENCES

[1] Aditi Gupta and Ponnurangam Kumaraguru. 2012. Credibility Ranking of Tweets During High Impact Events. In Proceedings of the 1st Workshop on Privacy and Security in Online Social Media (PSOSM '12). ACM, New York, NY, USA, Article 2, 7 pages.

[2] Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, and Anupam Joshi. 2013. Faking Sandy: Characterizing and Identifying Fake Images on Twitter During Hurricane Sandy. In Proceedings of the 22nd International Conference on World Wide Web (WWW '13 Companion). ACM, New York, NY, USA, 729–736.

[3] Aditi Gupta, Ponnurangam Kumaraguru, Carlos Castillo, and Patrick Meier. 2014. Tweetcred: Real-time credibility assessment of content on twitter. In International Conference on Social Informatics. Springer, 228–243.

[4] Allcott, H., and Gentzkow, M. Social Media and Fake News in the 2016 Election. https://web.stanford.edu/~gentzkow/research/fakenews.pdf, January, 2017.

[5] Anupama Aggarwal, Ashwin Rajadesingan, and Ponnurangam Kumaraguru. 2012. Phishari: Automatic Realtime Phishing Detection on Twitter. In eCrime Researchers Summit (eCrime). IEEE, 1–12. https://doi.org/10.1109/eCrime.2012. 6489521.

[6] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In Proceedings of the 20th international conference on World wide web. ACM, 675–684.

[7] Conneau et al. Very Deep Convolutional Networks for Text Classification. arXiv:1606.01781v2 [cs.CL] 27 Jan 2017.

[8] David M. Markowitz1 and Jeffrey T. Hancock, Linguistic Traces of a Scientific Fraud: The Case of Diederik Stapel. PLOS ONE, 1 August 2014 | Volume 9 | Issue 8 | e105937.

[9] David Mandell Freeman. 2017. Can You Spot the Fakes?: On the Limitations of User Feedback in Online Social Networks. In WWW. 1093–1102.

[10] Fabricio Benevenuto, Gabriel Magno, Tiago Rodrigues, and Virgilio Almeida. 2010. Detecting Spammers on Twitter. In Collaboration, Electronic Messaging, Anti-Abuse and Spam conference (CEAS), Vol. 6. 12.

[11] Facebook. 2016. News Feed FYI: Addressing Hoaxes and Fake News. https://newsroom.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-andfake-news/. (December 2016).

[12] Germany closely examining reports of Russian fake news campaigns. (January 2017). https://www.reuters.com/article/us-germany-russia-cyber-idUSKBN14T1SZ.

[13] Germany investigating unprecedented spread of fake news online. (January 2017). www.theguardian.com/world/2017/jan/09/germany-investigating-spread-fake-news-online-russia-election

[14] Hochreiter, S. and Schmidhuber, J. 1997. Long Short-Term Memory, Neural Computation.

[15] Jing Ma et al. Detecting Rumors from Microblogs with Recurrent Neural Networks. Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16).

[16] Kai Shu, Amy Slivaz, Suhang Wangy, Jiliang Tang, and Huan Liuy, Fake News Detection on Social Media: A Data Mining Perspective. arXiv:1708.01967v3 [cs.SI] 3 Sep 2017.

[17] Kumaran Gunasekaran, Gowtham Ganesan, Sudarshan Srinivasa Ramanujan, Balasubramaniam Srinivasan, Fake News Detection in Social Media. https://cseweb.ucsd.edu/classes/wi17/cse258-a/reports/a085.pdf

[18] Minyoung Huh, Andrew Liu, Andrew Owens and Alexei A. Efros. Fighting Fake News: Image Splice Detection via Learned Self-Consistency, arXiv:1805.04096v1 [cs.CV] 10 May 2018.

[19] Natali Ruchansky, Sungyong Seo and Yan Liu. CSI: A Hybrid Deep Model for Fake News Detection. arXiv:1703.06959v4 [cs.LG] 3 Sep 2017.

[20] Saif M Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. Stance and sentiment in tweets. ACM Transactions on Internet Technology (TOIT), 17(3):26, 2017.

[21] Samir Bajaj. “The Pope Has a New Baby!” Fake News Detection Using Deep Learning. Stanford University, 2017.

[22] Savvas Zannettou, Michael Sirivianos, Jeremy Blackburn and Nicolas Kourtellis. The Web of False Information: Rumors, Fake News, Hoaxes, Clickbait, and Various Other Shenanigans. arXiv:1804.03461 [cs.SI]

[23] Sebastian Tschiatschek, Adish Singla, Manuel Gomez Rodriguez, Arpit Merchant, and Andreas Krause. 2018. Fake News Detection in Social Networks via Crowd Signals. In WWW ’18 Companion: The 2018 Web Conference Companion, April 23–27, 2018, Lyon, France.

[24] Sejeong Kwon, Meeyoung Cha, and Kyomin Jung. 2017. Rumor Detection over Varying Time Windows. PLOS ONE 12, 1 (2017), e0168344.

[25] Shaheen Karodia and Rondebosch, Real Time Credibility Assessment of Twitter Content Related to the South African Political Context. http://projects.cs.uct.ac.za/honsproj/cgi-bin/view/2017/karodia_kinmont_lu.zip/resources/KRDSHA003_LitReviewFinal.pdf.

[26] Sneha Singhania, Nigel Fernandez, and Shrisha Rao. 3HAN: A Deep Neural Network for Fake News Detection, Springer International Publishing AG 2017.

[27] Svitlana Volkova, Kyle Shaffer, Jin Yea Jang and Nathan Hodas. Separating Facts from Fiction: Linguistic Models to Classify Suspicious and Trusted News Posts on Twitter. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Short Papers), pages 647–653 Vancouver, Canada, July 30 - August 4, 2017.

[28] Vanessa Wei Feng and Graeme Hirst. Detecting deceptive opinions with profile compatibility. International Joint Conference on Natural Language Processing, pages 338–346, Nagoya, Japan, 14-18 October 2013.

[29] Victoria L Rubin, Yimin Chen, and Niall J Conroy. 2015. Deception detection for news: three types of fakes. Proceedings of the Association for Information Science and Technology 52, 1 (2015), 1–4.

[30] William Ferreira and Andreas Vlachos. 2016. Emergent: a novel data-set for stance classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. ACL.

[31] Yimin Chen, Niall J. Conroy, and Victoria L. Rubin. ASIST 2015. November 6-10, 2015, St. Louis, MO, USA.

[32] Zhe Zhao, Paul Resnick, and Qiaozhu Mei. 2015. Enquiring minds: Early detection of rumors in social media from enquiry posts. In Proceedings of the 24th International Conference on World Wide Web. ACM, 1395–1405.

[33] ZHOU, et al. Fake News: A Survey of Research, Detection Methods, and Opportunities. arXiv:1812.00315v1 [cs.CL] 2 Dec 2018.