A Hybrid Model for Effective Fake News Detection with a Novel COVID-19 Dataset

Rohit Kumar Kaliyar[1], Anurag Goswami[1] and Pratik Narang[2]
[1]Department of Computer Science Engineering, Bennett University, Greater Noida, India
[2]Department of CSIS, BITS-Pilani, Rajasthan, India

Keywords: Fake News, Social Media, Machine Learning, Word Embedding, Neural Network.

Abstract: Due to the increasing number of users in social media, news articles can be quickly published or share among users without knowing its credibility and authenticity. Fast spreading of fake news articles using different social media platforms can create inestimable harm to society. These actions could seriously jeopardize the reliability of news media platforms. So it is imperative to prevent such fraudulent activities to foster the credibility of such social media platforms. An efficient automated tool is a primary necessity to detect such misleading articles. Considering the issues mentioned earlier, in this paper, we propose a hybrid model using multiple branches of the convolutional neural network (CNN) with Long Short Term Memory (LSTM) layers with different kernel sizes and filters. To make our model deep, which consists of three dense layers to extract more powerful features automatically. In this research, we have created a dataset (FN-COV) collecting 69976 fake and real news articles during the pandemic of COVID-19 with tags like social-distancing, covid19, and quarantine. We have validated the performance of our proposed model with one more real-time fake news dataset: PHEME. The capability of combined kernels and layers of our C-LSTM network is lucrative towards both the datasets. With our proposed model, we achieved an accuracy of 91.88% with PHEME, which is higher as compared to existing models and 98.62% with FN-COV dataset.

SCIENCE AND TECHNOLOGY PUBLICATIONS

1 INTRODUCTION

In the era of social media platforms and a rapid rate of enhancement in technology (Shu et al., 2017; Kumar and Shah, 2018), fake news has become one of the major problem in both industries as well as academia (Kumar and Shah, 2018) as it has the potential to influence the decisions and opinions of the common peoples of the society. Nowadays, fake news is intentionally written by fakesters (Kumar and Shah, 2018; Miller and Leon, 2017) to mislead readers and generating faith in biased news (Miller and Leon, 2017). In traditional media, some news articles often published unintentionally due to human neglect or incorrect data extraction (Miller and Leon, 2017). Few examples of fake news are shown with the help of Figure 1. The evolution of Social Networking platforms like Twitter, Facebook, and Weibo (Ghosh and Shah, 2018) etc. is one of the remarkable breakthroughs in human life and is available to publish news articles by the public. Fake news is usually created with fabricated data in the form of text, image, video, and audio. Due to the availability of online platforms, it is indeed a challenging task to detect fake news articles without a prior fact check. Hence, there arises the need for an automated model (Ghosh and Shah, 2018; Fazil and Abulaish, 2018) for fake news detection, which effectively detects the correctness and accuracy of the news articles.

For the detection of fake news, neural networks are quite popular in the area of Artificial Intelligence (AI) for their remarkable performance. Convolutional Neural Networks (CNN's) and Recurrent Neural Networks (RNN's) (Zhong et al., 2019; Wang et al., 2018) have emerged as the two powerful architectures for fake news classification. Through convolutional filters, CNN's are capable of learning with different relations through pooling operations. RNNs are capable of handling a different sequence of any length of word embedding vectors. Long Short Term Memory networks (LSTMs) (Ruchansky et al., 2017; Dahl et al., 2011) were designed to handle the problem of gradient exploding and memory accesses.

In this paper, we propose a hybrid model for fake news detection using a combination of convolutional layers having different kernel sizes with LSTM lay-

Kaliyar, R., Goswami, A. and Narang, P.
A Hybrid Model for Effective Fake News Detection with a Novel COVID-19 Dataset.
DOI: 10.5220/0010316010661072
In *Proceedings of the 13th International Conference on Agents and Artificial Intelligence (ICAART 2021) - Volume 2*, pages 1066-1072
ISBN: 978-989-758-484-8
Copyright © 2021 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

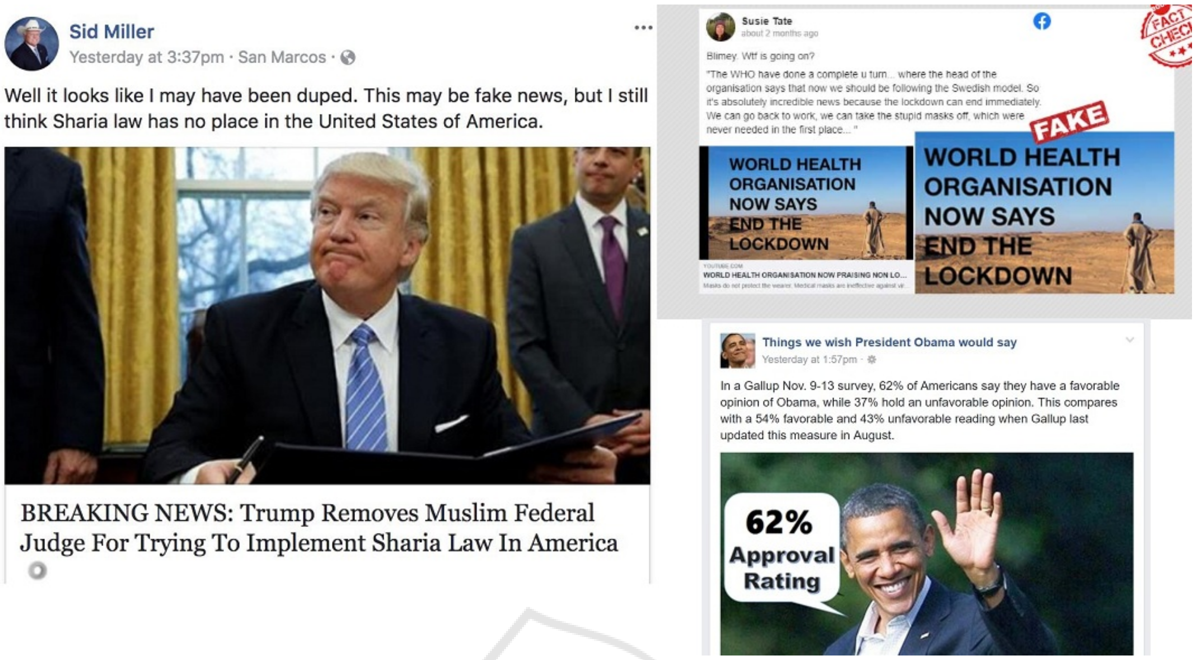

Figure 1: Examples of Fake News on social media (Source: Twitter and Facebook).

ers followed by three dense layers. We have designed our neural network using two convolutional layers with different kernel sizes for learning the model with different word size vectors. The feature maps of CNN are constructed as sequential features as the input of LSTM. In this architecture, we have organized each sentence into successive input features to help unravel factors of variations within the same sentences. Experimental results demonstrate the effectiveness of our proposed hybrid model compared to other existing CNN and RNN networks. Our proposed model utilizes the power of feature extraction using advanced pre-trained word embedding model. The embedding layer is a matrix of trainable weights which produces the vectors for each word index and improves the embedding of each word during training. The novelty of our proposed model lies in having combined different sized kernels and filters in each convolutional layer which are provided as an input to LSTM before concatenated to a fully connected layer. To make our C-LSTM model deeper, we have taken three dense layers to enable the composition of features from lower layers, potentially modelling the data, to approach the end goal quickly, and a higher-order decision boundary. Our model has performed very well on both PHEME as well as the FN-COV dataset with an accuracy of 91.88% and 98.62% respectively. Our model achieved state-of-the-art results as compared to existing methods for fake news detection.

2 RELATED WORK

In this section, work done in the field of fake news detection is summarized using deep learning techniques. Wang et al. (Wang et al., 2018) have proposed a framework called Event Adversarial Neural Network (EANN), which can learn transferable features for unseen events. Experiments were conducted on two large scale datasets collected from the popular social media platforms: Twitter and Weibo. They have also investigated with many deep learning methods using a Kaggle fake news dataset and authenticated news articles from Signal Media News. The authors observed that LSTM (Long Short Term Memory), GRU (Gated Recurrent Unit) and Bi-LSTM (Bi-directional Long Short Term Memory) classifiers gave better results than CNN (Convolution Neural Network). Roy et al. (Roy et al., 2018) have explored a combination of Convolutional Neural Network (CNN) and Bi-directional Long Short Term Memory (Bi-LSTM) model. Collobert et al. (Collobert et al., 2011) have deployed neural networks in their research with different convolutional filters to extract global features by max-pooling. Kim et al. (Kim, 2014) have proposed a CNN model with multiple filters and different window size. Kalchbrenner et al. (Kalchbrenner et al., 2014) have proposed a dynamic k-pooling method for text classification. Hochreiter et al. (Hochreiter and Schmidhuber, 1997) have discussed different versions of RNN's to store and access

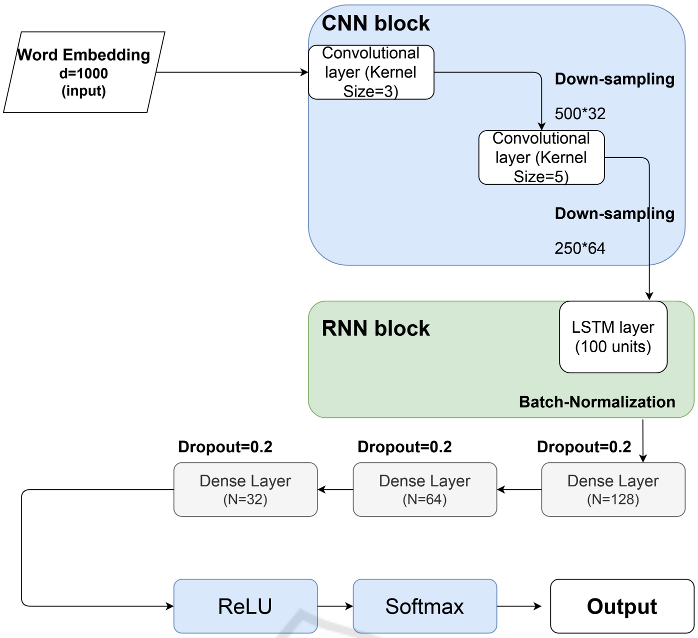

Figure 2: Proposed Model.

Table 1: Optimal Hyper-parameters for our proposed Hybrid Model.

Hyper-parameter	Description or Value
No. of Convolutional layer	2
No. of Max-pooling layer	2
No. of Kernel-sizes	3 and 5
No. of Dense layer	3
No. of filters in conv-layers	128,64,32
No. of filters in dense-layers	128,64,32,2
Loss function	binary_crossentropy
Activation function	Relu
Optimizer	Adam
Metrics	Accuracy
Batch-size	128, 32
Batch-Normalization	Yes
Number of Epochs	20
Dropout	0.2

memories.

Zubiaga et al. (Zubiaga et al., 2016) have proposed different machine learning models with different word embedding models using PHEME dataset. With their best approach, they were able to achieve accuracy with 81%. Abedalla et al. (Abedalla et al., 2019) have proposed a novel deep learning framework for the detection of fake news. They have developed four different models to validate the performance having different convolutional layers, LSTM layers, filters, dropout, and batch-normalization process. Ma et al. (Ma et al., 2018; Ma et al., 2016) have proposed a new architecture for fake news detection. With their model, they were able to achieve accuracy with 86.12%. Ajao et al. (Ajao et al., 2018) have proposed a framework to detects and classifies fake news messages from Twitter posts using a hybrid neural network model. Ruchansky et al. (Ruchansky et al., 2017) have explored a detection approach-CSI, which consists of three modules: Capture, Score, and Integrate. The model tested on Twitter and Weibo datasets.

In this research, we stack CNN and LSTM layers in a unified structure. This combination is effective because of multiple-branch convolution having different filters and kernel sizes for effective feature mapping with LSTM network having batch-normalization process. LSTM layer is capable of extracting meaningful information from convolutional layer for the effective detection of fake news articles. This combination of layers made our model more effective as compared to other existing models. In our model, we have considered different filter sizes across each dense layer having the capability of varying length feature mapping. We have trained our model with mini-batches 32 and 128 for better learning. The accuracy of our proposed model is better than the previously published models.

3 METHODOLOGY

In this section, the methodology and architecture of our proposed model have discussed.

3.1 Dataset

In this subsection, the datasets used in this research have discussed.

3.1.1 FN-COV

For creating the dataset, we have collected around 69,976 news articles with 44.84% of fake in total form the GDELT project [1] supported by Google. It has recently released brief snippets of worldwide English language news coverage mentioning COVID-19. The collection has included several topics that have been trending during COVID-19. For our experiment, we have selected COVID19, quarantine, and social distancing tag related news articles. This dataset consists of five attributes: ‘Date’, ‘URL’, ’Title’, ‘Text’, and ‘Label’. The date corresponds to the published date of the news, ‘URL’ represents the web address of the published news, ‘Title’ represents the headline of the news, ‘Text’ represents the content of the news, and ‘Label’ indicates whether the news is fake or not.

3.1.2 PHEME

PHEME dataset [2] is a collection of tweets scraped from Twitter, posted at the time of breaking news having five events. The events included are: Charlie Hebdo shooting from which we have 38,290 instances of news with 22% rumor content. Ferguson event dataset consists of 24,177 instances with 24.8% of rumor content. German wings plane crash, which comprises of 4489 instances of tweet level text data in the set with 50.7% rumor. Ottawa shooting which took place in Ottawa Parliament Hills in 2014 comprises 12,284 instances with 52.8% rumor and Sydney Siege where the gunmen took hostages at a cafe in Sydney in December 2014, consists of 24,001 instances tweet level text streams with 42.8% rumoured tweets.

3.2 Pre-processing

Text preprocessing is the practice of cleaning and preparing text data. In short, preprocessing refers to all the transformations on the raw data before it fed to the machine learning or deep learning algorithm. NLTK and re are standard Python libraries used to handle text preprocessing tasks. Such transformations are: Remove HTML tags, remove extra white-spaces, remove special characters, lowercase all texts, convert number words to numeric form, and remove numbers.

[1] https://blog.gdeltproject.org

[2] http://www.zubiaga.org/datasets/

3.3 Architecture of Our Proposed Hybrid C-LSTM Model

In this research, experiments have been conducted with our proposed C-LSTM network using both real-world fake news dataset. In Figure 2, the layered architecture of our deep neural network is shown. In our architecture, first layer is an embedding layer which accepts the input as a vector of 1000 word indices of length 32 following by a convolutional layer which performs matrix multiplications-based operations (Vasudevan et al., 2019; Sainath et al., 2015; Yang et al., 2018; Collobert and Weston, 2008). The first convolutional layer consists of kernel size=3 and 32 filters, followed by max-pooling. The second convolutional layer consists of kernel size=5 and 64 filters, followed by max-pooling. In our network, we have taken two pooling layer (Vasudevan et al., 2019; Yang et al., 2018; Zhong et al., 2019) which effectively down-samples the output of the prior layer, and reduce the number of operations required for all the following layers present in the network. Next layer in the architecture is an LSTM layer which is used to handle the nature of sequential data (Roy et al., 2018). This layer takes convoluted word combinations as input and the length of several units as output. Next, we have a flatten layer as a function that converts the features taken from the pooling layer and map it to a single column that is further passed to the fully connected layer. Then we have considered three dense layers in our neural network. The functionality of a dense layer as a linear operation (Vasudevan et al., 2019; Yang et al., 2018; Zhang et al., 2020) in which every input is connected to every output by some weight. First dense layer has 128 nodes and a dropout of 0.2. With small value of dropout, the accuracy will gradually increase and loss will gradually decrease. We have selected the value of dropout because it is helpful to reduce the complexity of the classification model and prevent over-fitting (Vasudevan et al., 2019; Yang et al., 2018; Zhang et al., 2020). The second dense layer also has 64 hidden nodes with a dropout of 0.2. The third dense layer has 32 hidden nodes and a dropout of 0.2. We have taken ReLU (Rectified Linear Unit) as the activation function. It is capable enough to remove negative values from an activation map by setting them to zero in a given network and increases the non-linear properties of the decision-making function without affecting any other fields of the convolution layer. We can define the equation of ReLU as:

$$\sigma = max(0, z) \tag{1}$$

We have used binary cross-entropy as loss function which measures the performance of a classification

Table 2: Comparison with Existing classification results using publicly available dataset-PHEME.

Author	Model	Accuracy (%)
(Zubiaga et al., 2016)	C-Random Forest	63.00%
(Zubiaga et al., 2016)	BOW+NB	68.15%
(Yu et al., 2017)	1-layer CNN	79.74%
(Zubiaga et al., 2016)	TF-IDF+KNN	80.94%
(Zubiaga et al., 2016)	BOW+DT	81.00%
(Ajao et al., 2018)	1-layer LSTM	82.76%
(Ajao et al., 2018)	LSTM-CNN	83.53%
(Ajao et al., 2018)	BILSTM-CNN	84.66%
(Ma et al., 2018; Ma et al., 2016)	RNN	86.12%
Our proposed model	**C-LSTM**	**91.88%**

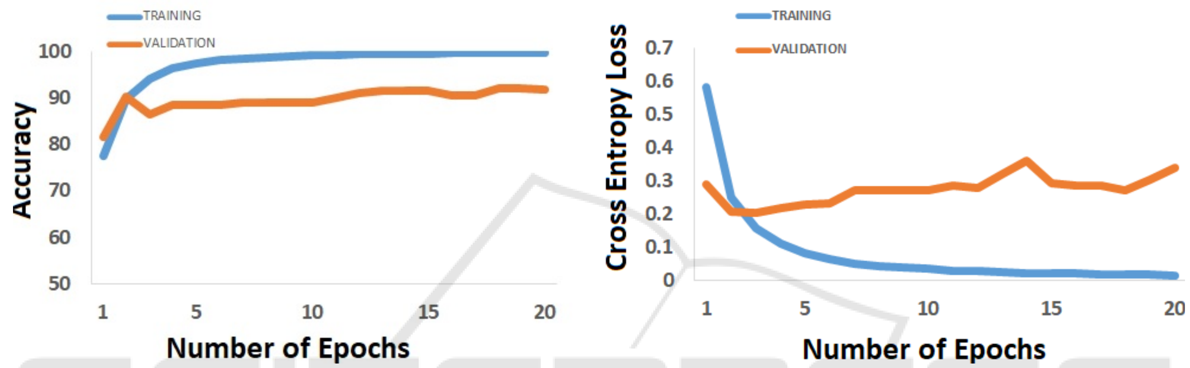

Figure 3: Accuracy and Cross Entropy Loss with C-LSTM using PHEME.

Table 3: Performance of our C-LSTM model with PHEME and FN-COV.

Dataset	Model	Precision	Recall	F1-Score
PHEME	C-LSTM	0.902	0.904	0.903
FN-COV	C-LSTM	0.992	0.989	0.994

Table 4: Accuracy of our C-LSTM model with PHEME and FN-COV.

Dataset	Model	Accuracy (%)
PHEME	C-LSTM	91.88
FN-COV	C-LSTM	98.62

model with a probability value between 0 and 1. It also increases as the predicted probability diverges from the actual label. In binary classification (number of classes M equals 2), cross-entropy can be calculated as:

$$L = -(ylog(p) + (1-y)log(1-p)) \quad (2)$$

If M>2 (i.e. multi-class classification), we calculate a separate loss for each class label per observation and sum the result:

$$-\sum_{c=1}^{M} y_{o,c} log(p_{o,c}) \quad (3)$$

Here, M - number of classes, log - the natural log, y - binary indicator (0 or 1) if class label c is the correct classification for observation o, p - predicted probability observation o is of class c. In our network, we have taken Adam as an optimizer. We have considered optimal hyper-parameters (see Table 1 for more details) for our proposed hybrid model. Hyperparameter optimization finds a tuple of hyperparameters that yields an optimal model which minimizes a predefined loss function.

In this research, the work was carried using the NVIDIA DGX-1 V100 machine. The machine is equipped with 40600 CUDA cores, 5120 tensor cores, 128 GB RAM and 1000 TFLOPS speed.

4 RESULTS AND DISCUSSION

Experimental and evaluation results have been tabulated in Table 2,3, and 4 using real-world fake news dataset: PHEME and our designed fake news dataset (FN-COV). Selection of optimal hyperparameters is shown in Table 1. Experimental results demonstrate that our proposed model performs state-of-the-art results compared to other existing detection models for fake news.

From Table 4, we can observe that the shallow machine learning models have resulted in a maximum

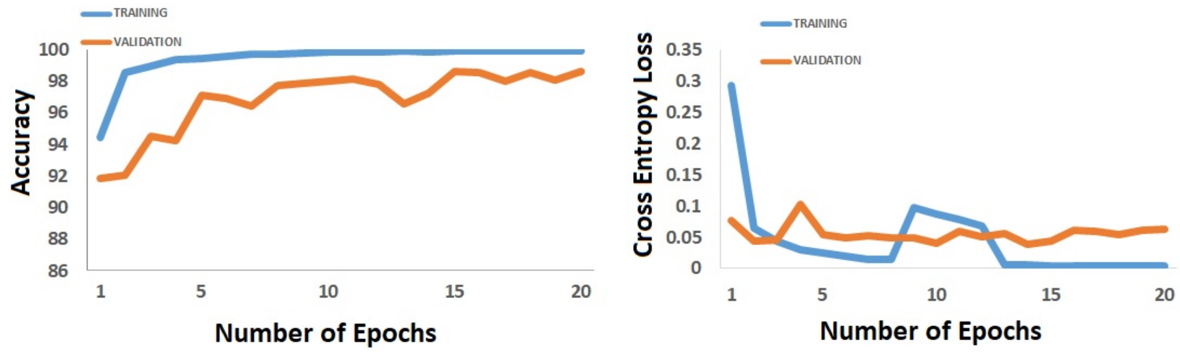

Figure 4: Accuracy and Cross Entropy Loss with C-LSTM using FN-COV.

of 81% accuracy using PHEME dataset. With deep learning models (CNNs, RNNs, etc.), the recurrent neural network architecture with glove pre-trained word embedding has achieved with 86.12% accuracy. Our proposed models, which are deep and hybrid (a combination of both CNN and LSTM layers) in nature, have performed exceptionally well and have resulted in more than 90% accuracy using PHEME dataset. It has also achieved an accuracy of 98.62% with our designed fake news dataset: FN-COV.

Figure 3 and 4 show the accuracy and cross-entropy loss using PHEME and FN-COV dataset. It also traces the learning ability and generalizing power of our proposed model. We can observe the performance of our proposed model over 20 epochs itself is quite remarkable on diverse and new dataset-FN-COV, respectively.

Cross-entropy loss is minimal in case of FN-COV. In Table 2, we have considered various performance parameters like precision, recall, and F1-Score, to validate the results. We have achieved the F1-score with 99.40% with FN-COV and 90.30% with PHEME dataset. In Table 4, a comparison with existing classification results using publicly available dataset (PHEME) is shown. We have achieved a 5% higher accuracy than the state-of-the-art methods with our proposed hybrid model. Our proposed model performed well with an accuracy of 98.62% using FN-COV. Classification results have shown a significant improvement in fake news detection using social media data with our proposed model. We have validated our proposed model with other real-world fake news datasets. We have achieved motivated results with other datasets also.

5 CONCLUSION AND FUTURE WORK

With our proposed C-LSTM model, we have achieved exemplary results, as it could capture both, the temporal semantics as well as phrase-level representations and achieved with optimized accuracies with minimal loss. In addition, we have created a novel dataset of fake news propagating during COVID-19. The experimental results empirically showed the effectiveness of the proposed model for fake news detection problem using both PHEME as well as the FN-COV dataset.

For future work, we will incorporate multiple metadata for more accurate classification and graph-based analysis to find the exact propagation path of fake news articles.

REFERENCES

Abedalla, A., Al-Sadi, A., and Abdullah, M. (2019). A closer look at fake news detection: A deep learning perspective. In *Proceedings of the 2019 3rd International Conference on Advances in Artificial Intelligence*, pages 24–28.

Ajao, O., Bhowmik, D., and Zargari, S. (2018). Fake news identification on twitter with hybrid cnn and rnn models. In *Proceedings of the 9th international conference on social media and society*, pages 226–230.

Collobert, R. and Weston, J. (2008). A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning*, pages 160–167. ACM.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., and Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of machine learning research*, 12(ARTICLE):2493–2537.

Dahl, G. E., Yu, D., Deng, L., and Acero, A. (2011). Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Transactions on audio, speech, and language processing*, 20(1):30–42.

Fazil, M. and Abulaish, M. (2018). A hybrid approach for detecting automated spammers in twitter. *IEEE Transactions on Information Forensics and Security*, 13(11):2707–2719.

Ghosh, S. and Shah, C. (2018). Towards automatic fake news classification. *Proceedings of the Association for Information Science and Technology*, 55(1):805–807.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

Kalchbrenner, N., Grefenstette, E., and Blunsom, P. (2014). A convolutional neural network for modelling sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 655–665.

Kim, Y. (2014). Convolutional neural networks for sentence classification proceedings of the 2014 conference on empirical methods in natural language processing, emnlp 2014, october 25-29, 2014, doha, qatar, a meeting of sigdat, a special interest group of the acl. *Association for Computational Linguistics, Doha, Qatar*.

Kumar, S. and Shah, N. (2018). False information on web and social media: A survey. *arXiv*, pages arXiv–1804.

Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B. J., Wong, K.-F., and Cha, M. (2016). Detecting rumors from microblogs with recurrent neural networks. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pages 3818–3824.

Ma, J., Gao, W., and Wong, K.-F. (2018). Rumor detection on twitter with tree-structured recursive neural networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1980–1989.

Miller, T. P. and Leon, A. (2017). Introduction to special issue on literacy, democracy, and fake news: Making it right in the era of fast and slow literacies. *Literacy in Composition Studies*, 5(2):10–23.

Roy, A., Basak, K., Ekbal, A., and Bhattacharyya, P. (2018). A deep ensemble framework for fake news detection and classification. *arXiv*, pages arXiv–1811.

Ruchansky, N., Seo, S., and 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*, pages 797–806.

Sainath, T. N., Vinyals, O., Senior, A., and Sak, H. (2015). Convolutional, long short-term memory, fully connected deep neural networks. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4580–4584. IEEE.

Shu, K., Sliva, A., Wang, S., Tang, J., and Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 19(1):22–36.

Vasudevan, V., Zoph, B., Shlens, J., and Le, Q. V. (2019). Neural architecture search for convolutional neural networks. US Patent 10,521,729.

Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., Su, L., and Gao, J. (2018). Eann: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 849–857. ACM.

Yang, Y., Zheng, L., Zhang, J., Cui, Q., Li, Z., and Yu, P. S. (2018). Ti-cnn: Convolutional neural networks for fake news detection. *arXiv*, pages arXiv–1806.

Yu, F., Liu, Q., Wu, S., Wang, L., and Tan, T. (2017). A convolutional approach for misinformation identification. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 3901–3907.

Zhang, J., Dong, B., and Philip, S. Y. (2020). Fakedetector: Effective fake news detection with deep diffusive neural network. In *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, pages 1826–1829. IEEE.

Zhong, B., Xing, X., Love, P., Wang, X., and Luo, H. (2019). Convolutional neural network: Deep learning-based classification of building quality problems. *Advanced Engineering Informatics*, 40:46–57.

Zubiaga, A., Liakata, M., and Procter, R. (2016). Learning reporting dynamics during breaking news for rumour detection in social media. *arXiv*, pages arXiv–1610.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS