Real-Time Detection of Fake News Articles using Deep Learning Techniques

Pakruddin B
Department of Computer Science & Engineering, Presidency University, Bengaluru, India
fakrubasha@gmail.com

Praveen Nandan K
Department of Computer Science & Engineering, Presidency University, Bengaluru, India
praveennandu4@gmail.com

Syed Afridi
Department of Computer Science & Engineering, Presidency University, Bengaluru, India
syedafridi644@gmail.com

Shailesh K R
Department of Computer Science & Engineering, Presidency University, Bengaluru, India
krshailesh627@gmail.com

Shubam V Patil
Department of Computer Science & Engineering, Presidency University, Bengaluru, India
sopatil.2003@gmail.com

***Abstract* — The rapid transmission of fake news that circulates on news sites constitutes a major problem for the reliability of information systems worldwide. Fake or Fake news articles can often change people's minds, destabilize the established democratic movement, and lead to confusion. This project aims to concentrate on the real-time detection of news articles to address the increasing problem of fake news propagation. To formulate this challenge, we employ a range of deep learning models– LSTM, ALBERT, FNNet, and the CNN+RNN model. The accuracy of these models is then compared to see the best architecture for identifying fake news. In textual data, temporal dependencies were captured by the LSTM model with a recorded accuracy of 96.25%. Using ALBERT, which is ideal for NLP tasks got a 100% accuracy pointing to its ability to decipher complex patterns in language. As the architecture of the FNNet is designed to detect fake news, it attained nearly 93% accuracy in its operation. Last but not least, the hybrid CNN+RNN, which combines spatial and temporal feature extraction, achieved 99% accuracy. The system demonstrated high efficacy in the real-time detection of real and fake news articles.**

Keywords— Fake news detection; Deep Learning, ALBERT and FNNet; CNN+RNN, LSTM.

I. INTRODUCTION

While the new generation is characterized by the increased usage of information technology, one of the biggest problems is the fake news which circulates on the internet far faster than the true one. This misinformation can go further and shape people's opinions change the electoral results or fuel social unrest [2]. The high frequency of fake or fake news articles is dangerous to the reputation of the media houses and diminishes the community's confidence in information channels [1]. Nevertheless, there is still a challenge experienced when trying to get the best approach to detecting fake news in real time.

There are some reasons why it is very hard to identify fake news and some of those factors include the stealthiness of the problem, the vast number of articles being produced in a second, and the fact that fake articles mimic real articles. The existing techniques as far as fact-checking is concerned are often slow, labor-intensive, and unfit for handling the volumes of information available on the internet; this is why automation and intelligent methods are key.

To overcome these drawbacks, our work aims at identifying fake news articles in real time through the use of sophisticated algorithms in machine learning and deep learning [3]. We use a set of models; each of them is aimed at solving different tasks in the field of natural language processing and classification. Long Short-Term Memory (LSTM) networks incorporate temporal properties in textual data and are thereby ideal for sequence data. A Lite BERT or ALBERT improves language comprehension with even fewer parameters compared to BERT and FNNet for the detection of fake news. Lastly, the proposed Convolutional Neural Network-CNN which is complemented with a Re- current Neural Network RNN is suitable to perform both spatial and temporal feature extraction leading to a reliable classification. Through these techniques, we believe that we can create a solution that can filter real from fake news articles with a high degree of accuracy as part of the continuous fight against fake news [4].

II. LITERATURE REVIEW

Deep learning is a sub-field of machine learning that uses artificial neural networks to try and learn from raw data without having to have specific features engineered for it, unlike most other machine learning algorithms. In this context, the deep learning architecture is called a Deep Neural Network (DNN) which is made up of layers like the Convolutional layer, pooling layer, and fully connected layer. These layers allow deep learning models to learn features in an automatic and hierarchical method, which is particularly important for tasks like fake news detection. The use of deep learning for deception detection in the text is a field problem that is close to fake news detection as explored in the paper by Karimi et. al., 2019[11]. As the study shows CNN has the Convolutional layers for performing the ability of feature extraction from the input text to map the patterns out for detecting the deceptive content. In as much as the Convolutional Neural Network (CNN), the authors describe it as a data architecture for mapping features from the data by passing the data through different filter sizes. The pooling layer then enhances the convolutional layer by lowering the computational complexity, which increases the model's efficiency without deleting significant information, including misinformation prevention on social media; Pennycook & Rand, 2020 [13]. Even though their work does not directly relate to the utilization of deep learning in the recognition of fake news it is nevertheless relevant to the discussion on the subject stating that input from humans could help supplement the algorithmic systems of detection. Thus this work poses several hypotheses and to test these hypotheses three forms of deep learning are used these include CNN, LSTM, and FNNet with a significant emphasis being placed on the CNN owing to its feature mapping properties.

III. Proposed Methodology

The proposed methodology in this study has two main phases. The training phase is the first stage as shown in Fig. 1 pulls out the training data from its database. This trained data will then be used in cleaning of somewhat inconsistent (for the lack of raw input)data from multiple sources.

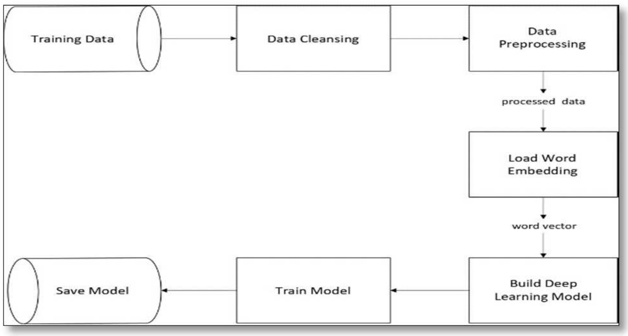

Fig. 1: Training Phase

After that, the data is preprocessed and converted into work vectors according to word embeddings. Fine-tuning deep learning model using pretrained word embeddings: We evaluated the major algorithms including ALBERT, CNN+RNN, Bidirectional LSTM and FNNet. Store the trained model to database, so we could use it in testing phase later on Fig. 2.

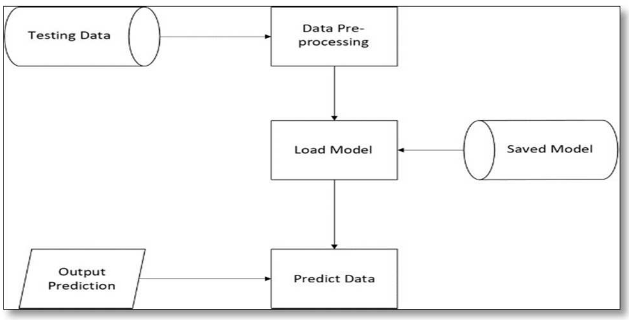

Fig. 2: Testing Phase

A. Dataset Used

In the past few years, it has become up to date which is what made us a global internet resource and everyone started using that. In turn, this necessitates substantial datasets of high quality examples from the intersection problem (that is real and fake news articles) to be able train models to detect it in real-time. ISOT Fake News Dataset [5]: Contains source articles and fake news from real news and fake sites. For the enlargement of content, The Fake News Dataset [6] adds a variety of news articles in different domains to further expand the scope for model generalization. There is sparse information avail- able in context of Binary Classification for news When comes to work on Bi- nary type of classification such as distinguish between real and fake news the Fack or News Dataset[7].The Fake News Detection Dataset [8] as well reinforces the model to provide either a suitable structured and context enhanced content dynamically with better or lesser agility depending upon where it detects which real news ends, fake starts over its samples on varied sections of articles — entity datasets true-fake orienteering for upto the-minute detection. The ISOT Fake News Dataset [5] provides the real world source articles as well as labeled fake news articles from some of the popular web pages. Improving the diversity of news articles in different domains from this content, that would help in a better generalization capability for The Fake News Dataset [6]. The Fake or Real News Dataset [7] is a key resource on the binary classification of news stories as true or false, parsing that information quite clearly established. There's also the Fake News Detection Dataset [8] making the model more resilient to put up coherent and targeted content by distinguishing as much between real news and fake. Those details are listed in the special tables of this paper. The dataset contains information on all three prominent videogame systems released over the years. This table we also found useful for understanding the number of each dataset to use these numbers in training our machine learning models, and the LSTM, ALBERT, FNNet, and CNN+RNN models allowed for high accuracy in real-time scenarios[9].

B. Objective

With the help of the WWW Growth as this entity to society, internet has tried to transform society into the digital one. Another area where the Internet is uniquely a risk initiator is with the coming of Fake News. Fake news is anything that leads people astray in the worst way possible. At the beginning of the 21st century, fake news become the major issue that can hardly be solved. It looks like a real news sight and making it hard to decide whether the delivered news or the source sight is attractive or manipulated. This is because the effects of fake news are that it erodes the public's confidence in Media institutions [18, 19], and health. For example, regarding treatments and prevention strategies, people got confused and engaged in dangerous practices in COVID-19. Fake news have been spread without much challenge because of negligence on the side of the public. Fake news can be naturally spread due to inaccurate sources and it can be created by an individual with a specific intention.

C. Key Indicators

Using advanced deep learning models we combined CNN and RNN architectures to improve the accuracy and speed of detecting fake news. This approach helps us identify misinformation more reliably, while also optimizing performance for greater efficiency. Continuous Improvement of Iteratively refining models guarantees precise and timely identification of fake news within dig- ital environments. The key indicator we are discussing in detail is Advanced Algorithms, which play a critical role in enhancing the accuracy and efficiency of real-time fake news detection.

D. Deep Learning Models used

The paper focuses on comparing several machine learning classifiers for fake news detection on ISOT Fake News, Fake or Real News, and Final News Datasets [10].LSTM (Long Short- Term Memory) delivered high performance as it focused on learning temporal dependencies in text and on capturing contextual flow, which has a great impact on fake news detection. It can find disparities in writing mode or the provision of information for instance, an article heading could be different from the article content. The proposed ALBERT (A Lite BERT), which applies a factorized embedding matrix and cross-layer parameter sharing, obtained a result of 100% for accuracy, precision, recall, and F1-score. ALBERT showed outstanding performance because the output is optimized on the Fake News detection data-set that aids in learning fake news from the anomalous feature in language patterns of fake news, from the perspective of an ordering of sentences and its context analysis. When using a combined CNN+RNN model which extracts local features in CNN and processes sequences in RNN the results were good and stable. It is more useful for identifying regular misspellings, and slang expressions and is often used by fakers to populate fake news [14]. FNNet (Feed-forward Neural Network) was also used in the study as the last classifier that takes features from other models including BERT and LSTM. When contextual embed-dings, sequence patterns, and sub-word information are taken into account, FNNet increases its ability to accurately determine between the two categories of real and fake news [15].

In this work, we present a deep learning architecture consisting of a series of interconnected layers intended for synchronization of the data input. As represented in the model illustrated in Fig. 3, we start with an Input Layer with a shape of (16, 64) (16, 64) (16, 64). Afterward, the data goes through the AlbertEmbeddings layer to get an embedded representation of size (16,128) (16,128) (16,128). After this, the Albert Transformer layer further processes the item-embedded data outputting an item summary that is a comprehensive summary of shape (16,768)(16, 768)(16,768). To increase the model's ability to generalize and also overcome the issue of over- fitting, a Dropout layer is used which retains the dimension of the output as (16,768) (16, 768) (16,768). Lastly, the shape of (16, 2) (16, 2) (16, 2) specifies the two targets for classification. This architecture makes it possible to well extract the features and classify them which are rele- vant activities addressed in this study.

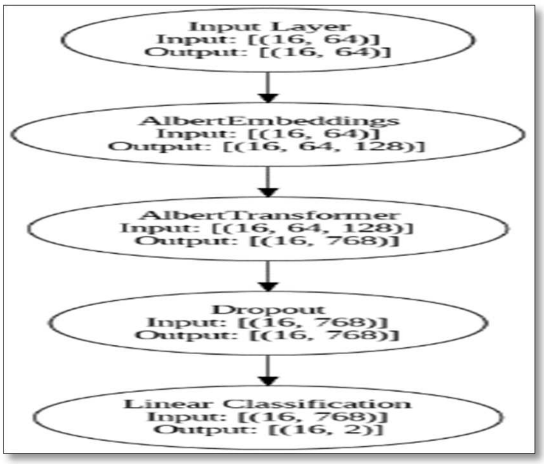

Fig. 3: ALBERT Architecture

IV. EXPERIMENTAL RESULTS

TABLE I. Performance analysis of four models

Algorithms	Dataset Used	F1-Score	Precision	Recall	Accuracy
	Results from proposed work				
ALBERT	ISOT	1.00	1.00	1.00	100%
	Fake News	1.00	1.00	1.00	100%
	Real or Fake	1.00	1.00	1.00	100%
	Fake News Detection	1.00	1.00	1.00	100%
LSTM	ISOT	0.99	0.99	0.99	100%
	Fake News	0.97	0.97	0.97	100%
	Real or Fake	0.77	0.77	0.77	100%
	Fake News	0.85	0.85	0.85	84.79%
CNN + RNN	ISOT	1.00	1.00	1.00	99%
	Fake News	0.98	0.98	0.98	99%
	Real or Fake	1.00	1.00	1.00	100%
	Fake News	1.00	1.00	1.00	100%
FNNet	ISOT	0.99	0.99	0.99	100%
	Fake News	0.99	0.99	0.99	100%
	Real or Fake	0.79	0.79	0.79	79%
	Fake News	0.93	0.93	0.93	93%

Table I. compares the performance metrics of albert, lstm, cnn+rnn, and fnnet on the all four discussed dataset, highlighting albert's consistent 100% accuracy across all tasks. In Table II, while ALBERT, CNN+RNN, and FNNet correctly identified the news as true, while LSTM mistakenly classified it as fake. In Table III. All models—LSTM, ALBERT, CNN+RNN, and FNNet—correctly identified the news as fake, demonstrating their accuracy in this example. Fig. 4. Showcases the bar chart illustrating the proportion of correctly learnt data patterns by different models. Fig. 5. Highlights the performance of ALBERT model for the fake news detection: Confusion matrices of four datasets have been shown below. Fig. 6. Showcases the confusion matrices of the same databases employing the CNN+RNN model, which puts the second best model presented in this research.

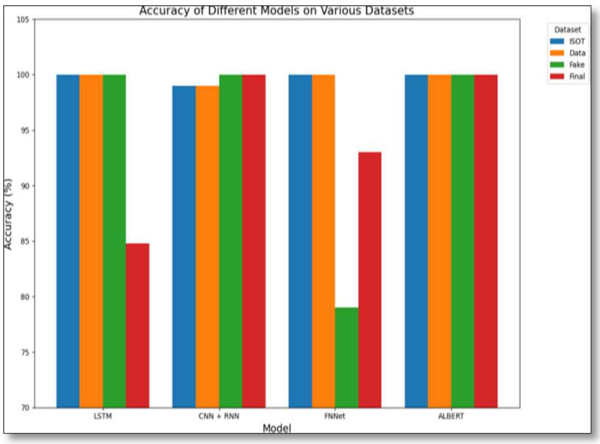

Fig. 4: Accuracy of the models across the Datasets

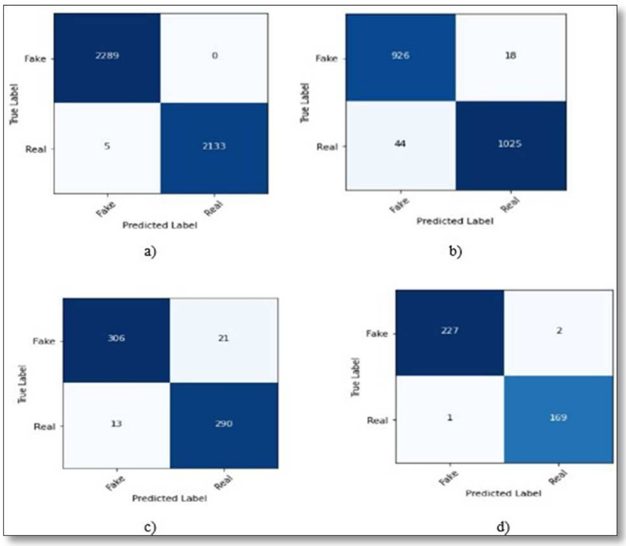

Fig. 5. Confusion matrices for the four datasets using the ALBERT model

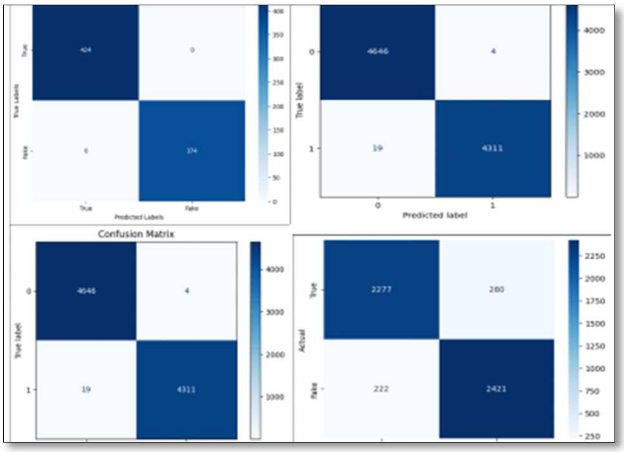

Fig. 6. Confusion Matrices for the four datasets using the CNN+RNN model

TABLE II. Prediction of true news

Sample of true news	Model	Prediction
The state government is nearing the final stages of selecting a site for Bengaluru's second international airport, with the location likely to be near Nelamangala. After a thorough evaluation, the area between Nelamangala and Kunigal, along National Highway 75, has been identified as the most suitable for the new airport, according to sources.	LSTM	Fake
	ALBERT	True
	CNN + RNN	True
	FNNet	True

TABLE III. Example of fake news

Sample of fake news	Model	Prediction
Fred Rogers served as a sniper during the Vietnam War and had a large number of confirmed kills. Fred Rogers wore his iconic sweaters to conceal the extensive tattoos on his arms that were acquired while serving in the military.	LSTM	Fake
	ALBERT	Fake
	CNN + RNN	Fake
	FNNet	Fake

V. Discussion and Future Challenges

The Act of Finding the difference between Fake and Real News has the following challenges; the existing real fake news mimics real news in style and; the availability of high-quality datasets is limited [16]. The more often people type words it is important to involve NLP and data science as well as journalism experts in using deep learning to identify fake news. Cooperation with media literacy organizations provides an ex- change of information and ideas. However, one of the major challenges is the unavailability of a large dataset. Uninterrupted collection, labeling, and dissemination of data regarding fake news to the related scientific community are crucial.

Therefore, our future work will consist of getting datasets on various news around the world which can be helpful for researchers working in the field of Print or Digital Media to build applications based on modern techniques [20]. We have many pretrained models and do not have an optimized model but when an optimized model is implemented for a particular task like Real-Time Detection of the authenticity of given News at any given time[17], then we can proceed as follows having well-defined objectives: data-collection, data-cleaning, data preprocessing, model-selection, transfer learning, hyper parameter tuning, regularization, loss, validation, testing, ensemble, Model optimization, regular up- dates, and Model deployment.

VI. Conclusion

As we continue marching toward the future, a new threat emerges: the possibility of falling victim to fake news. The society itself uses real-time detection techniques to solve everything This study was also concluded with the Evaluation of four different datasets using Four special deep learning models LSTM, ALBERT, CNN +RNN combined and FNNet. Word-level pre-trained word embeddings like Word2Vec, GloVe and fast Text were used along with each deep learning model. From the given test results, ALBERT has got in hand to out-perform all other three models since it guarantees 100% accuracy rate for each of the 4 datasets.

REFERENCES

[1] Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of Economic Perspectives, 31(2), 211–236. https://doi.org/10.1257/jep.31.2.211.

[2] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22–36. https://doi.org/10.1145/3137597.3137600.

[3] Ruchansky, N., Seo, S., & Liu, Y. (2017). CSI: A hybrid deep model for fake news detection. In Proceedings of the 2017 ACM Conference on Information and Knowledge Management (pp. 797–806). https://doi.org/10.1145/3132847.3132877.

[4] Zhou, X., & Zafarani, R. (2018). Fake news: A survey of research, detection methods, and opportunities. arXiv preprint arXiv:1812.00315. https://arxiv.org/abs/1812.00315.

[5] Rani, P., Shokeen, J. FNNet: a secure ensemble-based approach for fake news detection using blockchain. J Supercomput 80, 20042–20079 (2024). https://doi.org/10.1007/s11227-024-06216-4.

[6] Jiang, M., Jing, C., Chen, L. et al. An application study on multimodal fake news detection based on Albert-ResNet50 Model. Multimed Tools Appl 83, 8689–8706 (2024). https://doi.org/10.1007/s11042-023-15741-y.

[7] M. Takayasu, K. Sato, Y. Sano, K. Yamada, W. Miura, H. Takayasu, Rumor diffusion and convergence during the 3.11 earthquake: A Twitter case study, PLoS One 10 (2015) e0121443, https://dx.doi.org/ 10.1371/journal.pone.0121443.

[8] S.R. Sahoo, B.B. Gupta, Multiple features based approach for automatic fake news detec-tion on social networks using deep learning, Appl. Soft Comput. 100 (2021) 106983, https://dx.doi.org/10.1016/j. asoc.2020.106983.

[9] Zhang, X., & Ghorbani, A. A. (2020). An overview of online fake news: Characterization, detection, and discussion. Information Processing & Management, 57(2), 102025. ttps://doi.org/10.1016/j.ipm.2019.03.004.

[10] Conroy, N. K., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology, 52(1), 1–4. https://doi.org/10.1002/pra2.2015.145052010082.

[11] Karimi, H., Tang, J., Li, F., & Shah, S. (2019). Toward end-to-end deception detection in text. In Proceedings of the AAAI Conference on Artificial Intelligence, 33(01), 9713–9714. https://doi.org/10.1609/aaai.v33i01.33019713.

[12] Yin, W., Kann, K., Yu, M., & Schütze, H. (2017). Comparative study of CNN and RNN for natural language processing. arXiv preprint arXiv:1702.01923. https://doi.org/10.48550/arXiv.1702.01923.

[13] Pennycook, G., & Rand, D. G. (2020). Fighting misinformation on social media using crowdsourced judgments of news source quality. Proceedings of the National Academy of Sciences, 117(5), 2322–2328. https://doi.org/10.1073/pnas.1912444117.

[14] S. Wen, M.S. Haghighi, C. Chen, Y. Xiang, W. Zhou, W. Jia, A sword with two edges: Prop-agation studies on both positive and negative information in online social networks, IEEE Trans. Comput. 64 (2015) 640–653, https://dx.doi.org/10.1109/TC.2013.2295802.

[15] Brennen, J. S., Simon, F. M., Howard, P. N., & Nielsen, R. K. (2020). Types, sources, and claims of COVID-19 misinformation. Reuters Institutehttps://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation.

[16] Ferreira, W., & Vlachos, A. (2016). Emergent: A novel dataset for stance classification. Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL), 1163–1168. https://doi.org/10.18653/v1/N16-1138.

[17] A.B. Soliman, K. Eissa, S.R. El-Beltagy, Aravec: A set of arabic word embedding models for use in arabic NLP, Procedia Comput. Sci. 117 (2017) 256–265, https://dx.doi.org/10.1016/j.procs.2017.10.117.

[18] S.R. Sahoo, B.B. Gupta, Multiple features based approach for automatic fake news detec-tion on social networks using deep learning, Appl. Soft Comput. 100 (2021) 106983, https://dx.doi.org/10.1016/j. asoc.2020.106983.

[19] Khattar, D., et al. (2019). Mvae: Multimodal variational autoencoder for fake news detec-tion. In The World Wide Web Conference (pp. 2915-2921), .https://dl.acm.org/doi/10.1145/3308558.3313552.

[20] Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146–1151. https://doi.org/10.1126/science.aap9559.