International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhancing BERT for Fake News Classification Considering News Body: Segment-based Feature Extraction with Custom Task-Specific Layer

Deepti Nikumbh[1], Anuradha Thakare[2]

Submitted: 29/01/2024 **Revised**: 07/03/2024 **Accepted**: 15/03/2024

Abstract: Fake news dissemination on social media is a major concern today. Social media fake news is mostly multimodal in nature, where text forms the major component. People tend to believe whatever they read. Such news is a deception, impressively written with dishonest intents, and creates an impact on its readers, compelling them to like and further share it. Existing works mainly use the title of the news to extract features such as linguistic cues and psychological features, text semantics. These features can be extracted via different machine learning (ML) and deep learning models (DL). The most popular ones are language models in NLP. This models are trained on a large corpus like Wikipedia, which makes them capable of learning the complexities of language and capturing rich context information. For news titles, such models provide good performance for a given dataset but fail to generalize, as news titles are short providing less information to learn. Hence, processing news bodies is equally important to get state of art results. However, because of structural limitations, this architecture cannot be applied to lengthy news bodies. For example, a Language model like BERT is capable of processing limited text i.e. 512 sub-words.

In Natural Language Processing (NLP) using language models for feature extraction is a commonly adopted procedure. In this work, we try to address the problem of long document processing with respect to news bodies. We first divide the lengthy news body into overlapping segments with a maximum size of 200 words. A pre-trained BERT is used to learn local semantic contextual embeddings of the segment. All the segments of a news body are then passed to a custom-designed task specific layer to capture global contextual embeddings of the news body. We validate the efficiency of the proposed architecture on two real-world datasets. A comparative analysis in terms of various performance metrics is presented.

Keywords: *BERT, Transformer, Long document processing, fake news , machine learning(ML), deep learning(DL), fake news body , natural language processing(NLP) ,LSTM.*

1. Introduction

With the worldwide consumption of the internet and social media, the dissemination of news has become faster than ever before. Very easily, with basic authentication people create social media accounts and use it to share information. Day by day it has become difficult to verify the credibility of such information circulated on social media. Fake news is a very broad term. It is categorized as[14]: Rumour: An unverified claim about an event or a person transmitted by one person to another. False/deceptive: deliberately fabricated news with little or no truth. Misleading: stories with less factual and completely out of context. Slant/biased: stories where selective facts are hidden. Manipulated: Alerted and forged contents like images.

Fake news has a tremendous adverse effect on the society. It has become one of the greatest threats to democracy. Gruesome cases of lynching in Assam state of India[18], misleading post and information during COVID-19, and the 2016 US Presidential Election has witnessed massive spread of fake news. Reports indicate only 54% of humans have the fake deception detection ability without any help [8]. Fake news detection has become increasingly important because it reaches massive readers quickly and has a widespread malicious influence. Hence, and automated system is needed that can detect fake or misleading information quickly, preventing it from further becoming viral.

Many tech giants have now started investigating in fake news detection [13].AI techniques such as ML, DL, and NLP are used to design automated fake news detection systems. Advanced RNN architectures such as LSTM's and GRU are used to capture semantic contextual embedding. Despite advancements, this architecture still provides less performance when text with longer sequences is involved as a single state vector is used to handle the entire context of a long sequence [1]. Breakthrough transformer architecture introduced in 2017 addressed the shortcomings of recurrent networks. Transformer architectures are known for their attention mechanism and parallel processing. They are trained on large corpus and can be fine-tuned for specific tasks. BERT is one such transformer architecture that provides segment-level contextual embeddings. However, the problem with BERT is that it can only process text

[1] *Pimpri Chinchwad College of Engineering Pune-411044, India*
[2] *Pimpri Chinchwad College of Engineering Pune-411044, India*
ORCID ID :https://orcid.org/0000-0001-6281-8711
** Corresponding Author Email: deepti.nikumbh@gmail.com*

sequences of small length. When BERT is applied to long text, it will cut the input to the maximum length of 512 subwords. This is not desirable as model performance will be reduced as it will fail to capture long dependencies and global information from the text[1].

Thus, in this work, we have proposed an architecture to handle long text sequences that are encountered in fake news bodies and capture relevant information which will help in classification. The summarized research contributions of the work are as follows:

- To perform news classification by capturing the local and global semantic contextual information from the news body.
- To design hierarchical model combining BERT and a task-specific layer. The task-specific layer has two versions one with recurrent network and other with a transformer.
- To evaluate the effectiveness of the models with regards to various performance parameters like accuracy and loss, confusion matrix, precision,recall,F1-Score and AUC-ROC.
- To evaluate the generalizing ability of the model by applying it across two real-world datasets.

The rest of the paper is organized as follows: Section II discusses the review of the literature in the area of artificial intelligence for fake news detection. Section III. explains the proposed architecture and in detail explains the different modules involved in it. Section IV. provides the implementation details with respect to the experimental setup, information about the datasets used in the study, performance analysis, and comparative performance analysis. Section V provides the conclusion and future work.

2. Literature Review

Most of the existing work mainly extracts linguistic clues and semantic features from the news text.

R. Dzisevič et al. [11] demonstrated text feature extraction for smaller sentences by using TF-IDF, TF-IDF with LSA and TF-IDF with LDA.The approach with TF-IDF outperformed other approaches.

Bharadwaj et al. [7] used feature extraction techniques such as TF-IDF, N-gram including unigram and bigram models and recurrent neural network (RNN) to extract features from news text. The extracted features were further classified using Naive Bayes and random forest(RF) algorithms. To improve the classification accuracy linguistic clues and metadata information of the post can be combined with above mentioned semantic feature.

S. Girgis et al. [8] RNN'S and CNN to detect a piece of news as fake. The architectures were trained using LAIR dataset. CNN gave the best accuracy. Future work includes combining CNN and GRU

O. Ajao et al. [9] developed a classifier using 1D CNN and LSTM to classify Twitter posts as fake or authentic. They used the PHEME Dataset. They obtained an accuracy of 82%. Authors highlighted that along with the news post , geo location and origin of fake news should be identified.

M. Dong et al. [10] considered side information of post along with textual information. They developed an attentive bidirectional GRU and a deep neural network for extracting text and side information related features. In future authors would like to include images and videos related features while making final prediction.

J. Xue et al. [1]Attention pooling is used to address the problem of long document processing. The long document text is first broken into k fractions and each fraction is fed to the BERT base model. Attention pooling captures global dependencies. Authors in their future work would like to apply other pre-trained models such as XLNet and GPT for the task at hand.

Lu et al. [2] developed a hierarchical BERT for document classification task. Two BERT encoders were used namely ROBERTa to capture token embeddings and BERT to capture sentence embeddings. To match the input size of the sentence level BERT, averaging is done on the token embedding of each sentence. In future work, authors would like to apply the proposed model on large dataset and compare the performance.

Nishant et al. [3] developed an hybrid model with BERT and LSTM for fake news classification. The hidden state information of BERT is passed to the LSTM unit. The authors further would like to extract linguistic and lexical features of fake news and train a mode and tune the hyperparameters of BERT and its following layers.

Farokhian et al.[4] developed an algorithm named MaxWorth which selects a continuous range of news text to a maximum of 512 length which is more valuable for fact-checking.ClaimBuster API is used to evaluate and rank the chosen text. Other modalities such as context, comments, feedback, and non-text modalities should be used to detect news fakeness.

Pappagari et al.[5] mentioned that inability to process long documents is a major issue of transformers .They proposed two hierarchical transformer-based approaches:RoBERT, ToBERT. ToBERT is best in terms of accuracy and complexity. The proposed models can be tested on another long document-processing task.

[12] discussed techniques which can be implemented to used BERT effectively for long documents. This includes trimming long text into smaller segments and segment output votes averaging.

S Singh et al.[6] detailed summary of advanced language models used for various NLP task. They highlighted that, NLP has advanced exponentially through the introduction of attention and transformer architectures.These language models which give performance boost to NLP applications require huge computing power. Reinforcement learning can be incorporated in various NLP task like Machine Translation and text summarization.

3. Methodology

Figure 2 provides a description of the proposed architecture. The architecture has three modules namely: input generation module, the feature learning module, and the classification module. The input generation module processes the long news body into segments and generates a vector representation of it. The feature learning module is responsible for obtaining the contextual representation of the provided text. The classification module further classifies the learned features to predefined classes.

3.1. The input generation module:

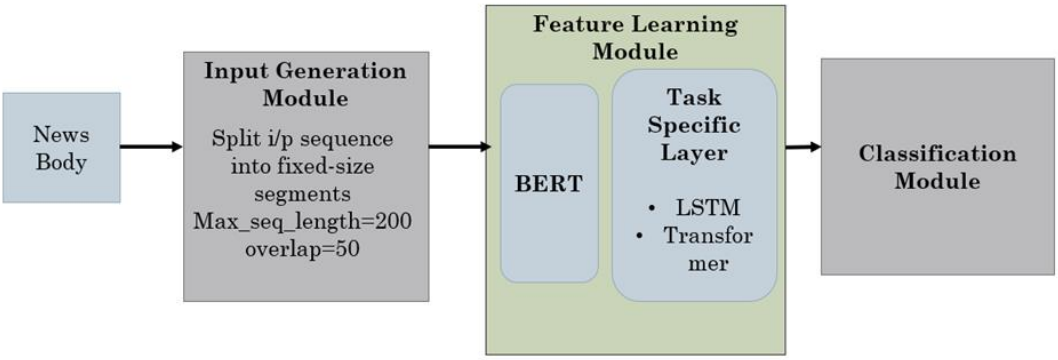
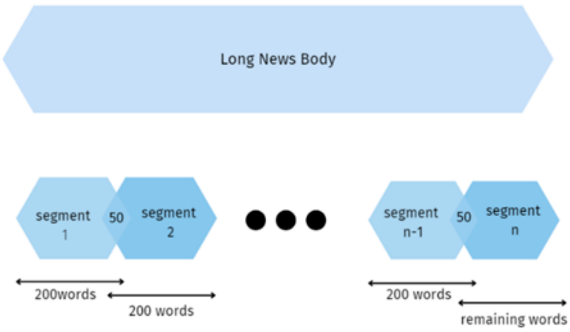


Fig.2. Proposed Methodology

One limitation of BERT is that it cannot process text which is longer than 512 tokens. Hence, in this module, the long text of the news body is split into smaller segments of the size of less than or equal to 200 tokens. An overlap of 50 words is kept between the segments [5]. Figure 1 describes the process of splitting the long text document. Each word in the segment is first converted into token and then subjected to segment and position embedding. Then a final embedding vector is generated per segment which can be processed by BERT.

3.2. Feature Learning Module:



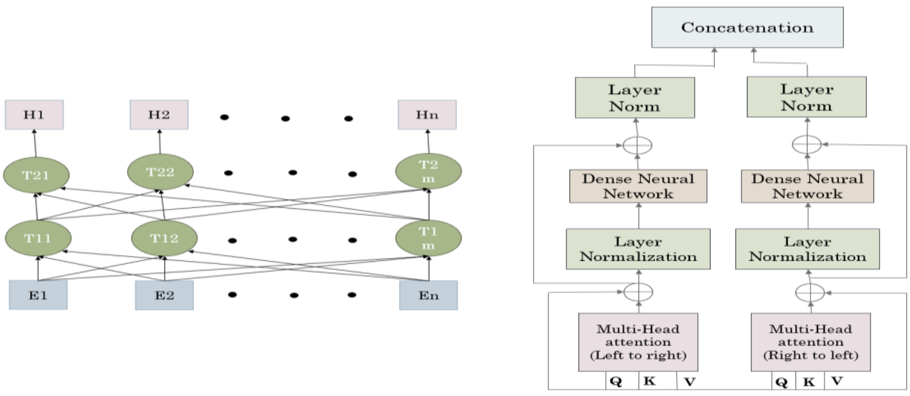


Fig. 3. BERT architecture on the left , E indicates the vector embedding of each word after incorporating token, segment and position embedding, T indicates one transformer block ,H indicates the output representation of BERT. The diagram on right is a detailed transformer block[19].

Fig 1: Long news body broken down into chunks of 200 tokens, the last segment contains the remaining words which will be less than 200 words.

The numerical vectors obtained from the previous module are then fed to BERT. BERT is an advanced pre-trained language model based on transformer architecture. It can capture rich contextual representations of the sentence. BERT is so popular because it performs bi-directional training, i.e., it processes text from right hand to left and from left hand to right, provides multi-headed attention, residual connections, and layer normalization. BERT performs well on NLP tasks as it has a complex structure and non-linear learning potential. Google has released two versions of BERT whose configuration is shown in Table 1.

Table1: BERT configuration

BERT model	*Layers*	*Output dimension*	*Multi-head attention*	*Total Parameters*
BERTbase	12L	768D	12H	110M
BERTLarge	24L	1024D	16H	340M

The structure of the BERT is shown in Figure 3. As mentioned, BERT has a hierarchical structure. Here the transformer blocks are placed in a hierarchical fashion.

In the work, BERTbase architecture is used. Thus 12 encoders consisting of 12 multi-headed attention is applied on each segment embedding vector. BERT provides output in two forms namely pooled output and sequence output. Pooled output i.e. CLS token embedding which captures the entire context of the segment is taken as the final output of BERT. For a news body D, all its pooled output is stacked together and if given to a task specific layer to learn long-term context information and generate the global embeddings of the entire news body. The representation of the output of BERT for a news body is shown as follows:

$$D^i = \{s_1^i, s_2^i \dots \dots ., s_n^i\} \qquad (1)$$

Where i is the news body and n is the total number of segments per news body. S is the CLS token embedding vector of 768 dimension.

3.2.1. Task Specific Layer:

Task-specific layer is responsible for learning global contextual embeddings and long-term dependencies from the complete news body. This layer is made up of LSTM or Light weight transformer.

1) Light weight Transformer layer:

For each news body, all its segment-level embeddings obtained from the BERT are stacked, and then fed to a small lightweight custom-designed transformer. The architecture of the custom transformer is given in Figure 4. Dropouts and residual connections are encouraged in the transformer block. Dropouts prevent the networks from overfitting and residual connections provide a smooth gradient flow by avoiding significant loss of information while passing to the next layer. Layer normalization has an advantage over batch normalization. It performs feature-wise normalization and will not affect performance even if the

batch size is 1. The output of the transformer servers as a final news body embedding.

2) Long Short Term Memory(LSTM's):

Since the news body is divided into segments and these segments are connected, Recurrent network such as LSTM can also be a good choice for task-specific layer. LSTM's unfold in time. Their architecture can preserve long-term dependencies of the text. The only issue with the LSTM is its complex architecture and sequential processing ability.

In the propose work, two hierarchical models are created: BERT+transformer and BERT+LSTM.

3.3. Final classification module:

On the top of the feature learning module, a fully connected layer is added to perform the classification task. During the training phase, the BERT and task-specific components like transform, LSTM, and fully connected layers are learned via an objective function. The final layer has a single neuron with a sigmoid activation function which does the final classification.

3.4. Task-specific training:

The pre-trained BERT, task-specific layer is utilized during training. Algorithm 1 provides the detailed working of proposed model. In the feature learning phase, BERT is initialized with the pre-trained weights and task-specific components are randomly initialized. Then during training, the BERT weights are used and the task-specific components are optimized by reducing the cross-entropy loss.

Equation (1) shows the entropy loss function where ti is the true label and zi is the probability for the ith class and n is the number of classes.

$$loss = -\sum_{n=0,1} t_i \log (z_i) \quad (2)$$

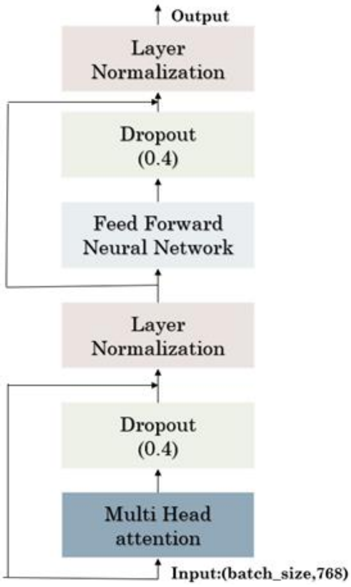


Fig. 4. Architecture of lightweight transformer

3.4.1. Algorithm of proposed work:

News_feature_extractor (news_body):

Input: body of news article.

Output: Fake or authentic news.

1. Max_seq_length=200
2. Batch_size=32
3. For each news in news_body:

   token_seg=get split(news) // Divide the news body in tokens, each token can have a maximum 200 words with 50 words overlap.

   seg_embedding=BERT_base(token_seg).// Generate the context embeddings for each segment of dimension (32,768).

4. body_embeddings=combine_seg_embeddings(seg_embeddings) // Combines all the segment embeddings of a specific news body.
5. global_body_emdeddings=task_specific_layer(body_emdeddings)// Contains a light weight transformer or LSTM blocks
6. news_feature= global_body_embedding
7. prediction=dense_layer(news_feature)
8. return(prediction)

4. Implementation:

4.1. Experimental Setup:

Google Colab pro and Google Drive service are used for designing the model and accessing the dataset. Python, Keras library over TensorFlow framework and BeautifulSoup are used. The hyperparameter setup used for model training is shown in Table 2.

Table 2: Hyperparameter setup for the proposed model.

Hyperparameters	***Value***
Optimizer	Stochastic gradient descent
Loss function	Binary cross entropy(log loss)
Batch size	32
Learning rate	0.0001
Early stopping	Yes with patience of 13.
Number of epochs	100

To prevent overfitting of the proposed model, dropout and regularizes are incorporated in the network. Relu activation is used for the intermediate layers and Sigmoid is used for the final output layer.

4.2. Datasets:

Most of the fake news datasets are created by collecting social media tweets , news headlines , claims, and statements given by political figures etc. But this text is short in length and does not provide much information about the news. The proposed work uses full-length news articles. Thus, the following datasets are used for evaluating the proposed model.

4.2.1. ISOT Fake News Dataset:

It is a dataset containing true and fake news along with news body. The true news articles are collected from Reuters.com and the fake news articles are collected from unreliable websites flagged by Wikipedia and PolitiFact. The dataset is balanced with 12600 true and 12600 fake news articles[17].

4.2.2. Fake News Dataset:

This dataset provides news body along with news titles. It is a huge dataset available at Kaggle with 40,000 training samples and 4000 test samples which is sufficient to train and validate the proposed model[16].

In the work, only news body greater than 200 words is used. The overview of samples with news body greater than 200 words in the above-mentioned dataset is shown in table 3.

Table3: Distribution of samples in the mentioned datasets.

Dataset	***Fake***	***Real***	***Total***
Fake News Dataset	9000	8500	1750
ISOT Fake News Dataset	12600	12600	25200

4.3. Performance analysis of the proposed model:

Two models namely BERT+LSTM and BERT+Transformer are designed and evaluated on two real world datasets. Table 4 provides the accuracy and loss measures for the same.

The BERT+LSTM provides good accuracy on the mentioned datasets. The performance plots for the model are shown in figure 56,9,10. Plot's depicts that model has properties of good fit as both training and validation accuracy go hand in hand and the loss of the model decreases as the training progresses. Thus, the model approaches low bias and low variance. The BERT+transformer provides comparatively less accuracy but learns more quickly than BERT+LSTM. Plots depict a slight overfitting as there is a difference between training and validation accuracy resulting in a model with low bias and high variance. The performance plots for BERT+transformer are shown in figure 7,8,11,12.

Table 4. Performance Analysis of proposed models.

Model Architecture	***Dataset***	***Training Accuracy***	***Validation Accuracy***	***Test Accuracy***	***Test Loss***
BERT+LSTM	Fake News Dataset	96.82	96.22	96.93	0.086
BERT+TRANSFORMER	Fake News Dataset	95.94	83.66	83.95	0.566
BERT+LSTM	ISOT	97.22	96.85	96.44	0.108
BERT+TRANSFORMER	ISOT	95.32	81.94	81.76	0.552

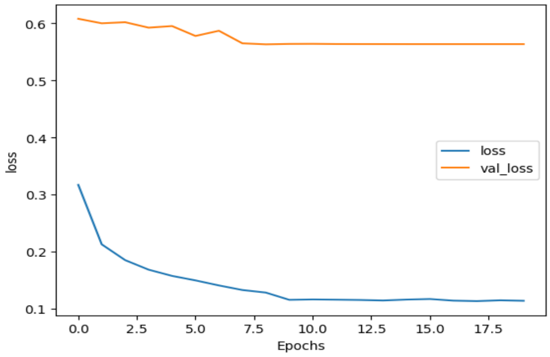


Fig. 5. Loss plots(Fake News dataset) for BERT+LSTM

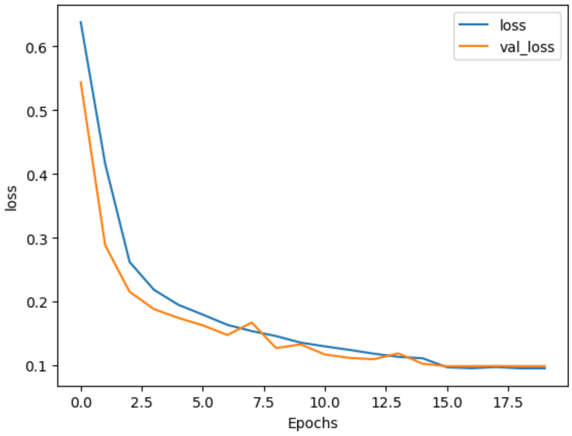


Fig. 6. Accuracy plot(Fake News Net) for BERT+LSTM

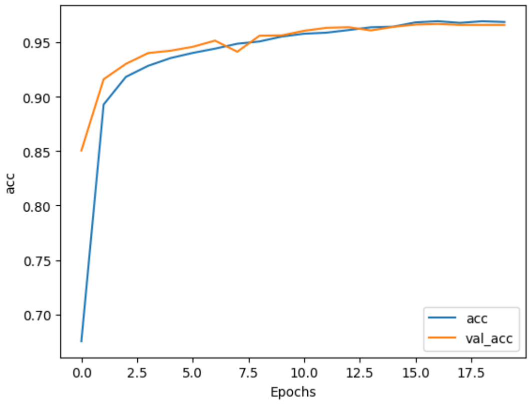


Fig. 7. Loss plot(Fake News Net) for BERT+Transformer

A detailed performance analysis of the proposed models is presented in table 5The confusion matrix for the same on fake news dataset is shown in figure13. BERT+LSTM exhibits high True Positive and True Negative values, as majority of samples are correctly predicted by the model, and very low False positives and

False negatives, resulting in high precision, recall and F1 score.

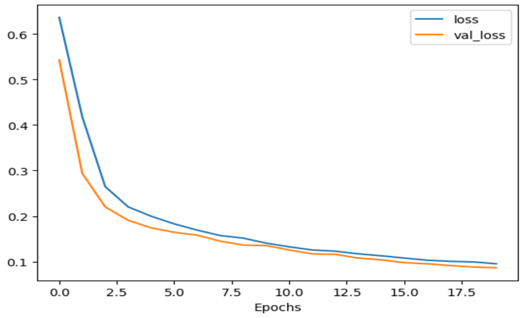


Fig. 9. Loss plots (ISOT Fake News dataset) for BERT+LSTM

Table 5. Detailed Performance Analysis of proposed models.

Model Architecture	*Dataset*	*Precision*	*Recall*	*F1-Score*	*False Positive Rate*	*False Negative Rate*	*AUC-ROC*
BERT+LSTM	ISOT	96.14	98.49	97.30	0.015	0.04	0.9675
BERT+TRANSFORMER	ISOT	74.44	**97.56**	84.44	0.009	0. 402	0.7944
BERT+LSTM	Fake News Dataset	96.86	96.39	96.63	0.036	0.038	0.9629
BERT+TRANSFORMER	Fake News Dataset	75.56	**99.09**	85.74	0.009	0.391	0.8048

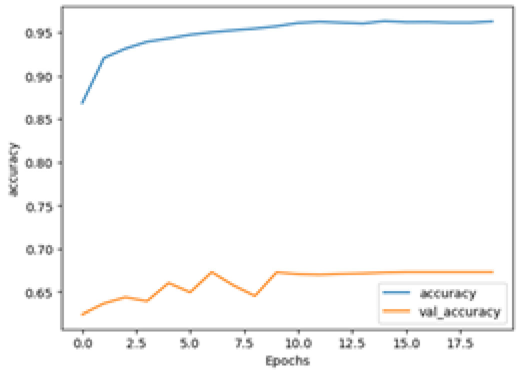
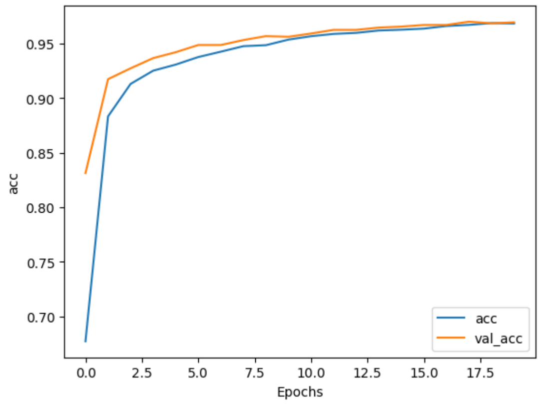


Fig. 8. Accuracy plot(Fake News Net) for BERT+Transformer

Fig. 10. Accuracy plots (ISOT Fake News dataset) for BERT+LSTM



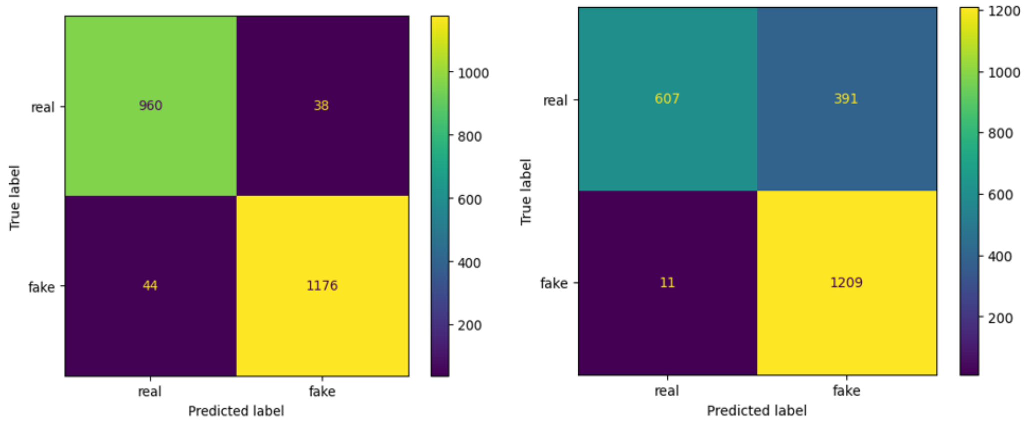


Fig 13: Confusion matrix left figure for the model BERT+LSTM and right figure for BERT+Transformer.

BERT+Transformer has comparatively high False Positives i.e. real news articles classified as fake. This drastically reduces the precision of the model. However, the model has comparatively fewer False Negatives i.e. fake articles classified as real resulting in a high recall value and low False Positive rate.

4.4. Discussions:

BERT is used for feature extraction. It provides a contextual embedding of the segment. The idea is to capture the embeddings of the [CLS] token, as it is the first token in the sequence and captures the entire context which is sufficient for a downstream task such as classification. A layer on top of BERT is used to learn global embeddings and enhance the performance of the model. Recurrent networks or a lightweight transformer could be a good choice for the top layer.

LSTM's process the data sequentially. The 768-dimensional BERT output embeddings of each segment are passed sequentially to LSTM units. These LSTM units unfold in time. They remember the information from previous timestamps thus capturing the long-term dependencies in the segments.

Transformers process the data in a parallel manner. Compared to LSTM transformers learn faster. The reason for this could be the multi-headed attention mechanism which provides alignment scores of each input with all the other ones, followed by residual connections. Such complex architecture elevates the interconnection rate resulting in more parameters to learn. Hence transformers learn quickly and they tend to overfit. But overfitting in down-streaming task is useful. Models can continue to learn even after they overfit[15].

Thus, if the size of the dataset is small, LSTMs are good option, else transformers are the best choice as they learn quickly and will be beneficial over a long run.

5. Conclusion And Future Work:

In this work, we have addressed a long document processing problem with respect to the news body. In the existing work of fake news detection, news titles are mostly processed to extract useful information. But these news titles are very small in length, hence

cannot provide complete information about the news, hence the processing of news body is important. Existing language models like BERT can handle text with limited words making them ineffective for long text.

We designed a hierarchical approach where a task-specific layer of LSTM or a lightweight transform is stacked on top of pre-trained BERT. All the segment-level BERT embeddings of a long document are passed to the task-specific layer to learn global embeddings. The performance plots of both models are shown in table 2. BERT+LSTM provides promising accuracy with a test set accuracy of 96.93% , whereas BERT+transformer provides an test accuracy of 83.95%. We presented a comparative study and highlighted the fact that the transformer layer is a good choice for task-specific layer provided sufficient data is available for training and validation. With such promising results, the news body processing can also be added in the fake news detection systems.

Future authors would like to experiment with the designed BERT+Transformer model on larger datasets to achieve greater performance than the current state.

References

[1] J. Xue, X. Tang and L. Zheng, "A Hierarchical BERT-Based Transfer Learning Approach for Multi-Dimensional Essay Scoring," in IEEE Access, vol. 9, pp. 125403-125415, 2021, doi: 10.1109/ACCESS.2021.3110683.

[2] Lu, J., Henchion, M., Bacher, I., Namee, B.M. (2021). A Sentence-Level Hierarchical BERT Model for Document Classification with Limited Labelled Data. In:

Soares, C., Torgo, L. (eds) Discovery Science. DS 2021. Lecture Notes in Computer Science(), vol 12986. Springer, Cham. https://doi.org/10.1007/978-3-030-88942-5_18

[3] Nishant Rai, Deepika Kumar, Naman Kaushik, Chandan Raj, Ahad Ali, "Fake News Classification using transformer based enhanced LSTM and BERT", International Journal of Cognitive Computing in Engineering, Volume 3, 2022, Pages 98-105, ISSN 2666-3074, https://doi.org/10.1016/j.ijcce.2022.03.003.

[4] Farokhian, M., Rafe, V., & Veisi, H. (2022). Fake news detection using parallel BERT deep neural networks. ArXiv, abs/2204.04793.

[5] Pappagari, R., Żelasko, P., Villalba, J., Carmiel, Y., & Dehak, N. (2019). Hierarchical Transformers for Long Document Classification. 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 838-844.

[6] S. Singh and A. Mahmood, "The NLP Cookbook: Modern Recipes for Transformer Based Deep Learning Architectures," in IEEE Access, vol. 9, pp. 68675-68702, 2021, doi: 10.1109/ACCESS.2021.3077350.

[7] Bharadwaj, Pranav Shao, Zongru Darren, Seth. (2019). FAKE NEWS DETECTION WITH SEMANTIC FEATURES AND TEXT MINING. Journal of Natural Language Processing. 8. 17-22.

[8] S. Girgis, E. Amer and M. Gadallah, "Deep Learning Algorithms for Detecting Fake News in Online Text," 2018 13th International Conference on Computer Engineering and Systems (ICCES), 2018, pp. 93-97, doi: 10.1109/IC17 CES.2018.8639198.

[9] O. Ajao, D. Bhowmik, and S. Zargari. 2018. Fake News Identification on Twitter with Hybrid CNN and RNN Models. In Proceedings of the 9th International Conference on Social Media and Society (SMSociety '18). Association for Computing Machinery, New York, NY, USA, 226–230. DOI:https://doi.org/10.1145/3217804.3217917

[10] M. Dong, L. Yao, X. Wang, B. Benatallah, X. Zhang and Q. Z. Sheng, "Dual-stream Self-Attentive Random Forest for False Information Detection," 2019 International Joint Conference on Neural Networks (IJCNN), 2019, pp. 1-8, doi: 10.1109/IJCNN.2019.8851765

[11] R. Dzisevi˘c and D. Se˘sok, "Text Classification using Different Feature Extraction Approaches," 2019 Open Conference of Electrical, Electronic and Information Sciences (eStream), 2019, pp. 1-4, doi: 10.1109/eStream.2019.8732167.

[12] Pinjarkar V., Jain A., "Analysis of Classification Rules Optimization Using Genetic Algorithm", Journal of Design Engineering, volume 2021, issue 9, pp: 4096-4107, 2021.

[13] BERT:How to Handle Long Documents [Online]. Available: https://www.saltdatalabs.com/blog/bert-how-to-handle-long-documents Accessed on: September 15,2023.

[14] Romano, A. (2018). Mark zuckerberg lays out facebooks 3-pronged approach to fake news.

[15] Fake News: Categories. [Online]. Available: https://library-nd.libguides.com/fakenews/categories Accessed on: September 29,2023.

[16] Learning After Overfitting, 06 April, 2022. [Online]. Available: https://www.deeplearning.ai/the-batch/learning-after-overfitting/ Accessed on: October 10,2023.

[17] Fake News Dataset, [Online]. Available: https://www.kaggle.com/search?q=fake+news+dataset Accessed on: August 25,2023.

[18] ISOT Fake News Dataset [Online]. Available: https://www.kaggle.com/datasets/emineyetm/fake-news-detection-datasets Accessed on: August 30, 2023.

[19] Assam Lynching case, 18 June,2018 [Online]. Available: https://timesofindia.indiatimes.com/india/assam-lynching-case-16-arrested-search-on-for-remaining-accused/articleshow/64533326.cms Accessed on 21 July ,2023.

[20] Dr. Niraj Kumar, Deep Learning, [Online]. Available: https://ai-leader.com/deep-learning/ Accessed on: 10, July,2023.