Textual analysis of Fake News Detection on Social Media Platforms

Allen Jacob George[1][0009−0001−0929−7758], Ashiq Sheriff A[1][0009−0008−5290−9255], Navaneet N[1][0009−0009−7857−5051], Nukala Venkata Durga Sandeep[1][0009−0000−8916−9555], and Bagavathi C[1][0000−0002−6257−0004]

1 Department of Computer Science and Engineering, Amrita School of Computing, Coimbatore, Amrita University, India c_bagavathi@cb.amrita.edu

Abstract. : The spread of fake news on social media threatens public opinion and social stability. Large volumes, high velocities, and changing natures of online content make it difficult for traditional methods to detect fake news to remain effective. To overcome such challenges, this paper introduces a cutting-edge system to integrate deep learning models with evolution techniques that improves the detection and classification of fake news. The system put together consists of the following steps: text preprocessing, reliable feature extraction and optimization through GANs. Lastly, the improved BiLSTM network with genetic algorithms produces a classification that performs better and is adaptive. This will allow the model to adapt to the dynamic nature of fake news content and improve accuracy in detection as well as generalization across different datasets. The paper also goes on to describe the architecture in detail along with implementation methodology and its possible real-time adaptability to combat fake news, which contributes to a more reliable and secure information ecosystem.

Keywords: Fake News Detection · BERT · BiLSTM · GAN

1 Introduction

In the current digital era, social media platforms have become essential for information exchange; yet, this also presents problems with the unchecked spread of false information [7]. False information can spread swiftly, thanks to social media's reach and immediacy, altering public opinion and societal effects.

Because online content is so dynamic and so vast, traditional methods of identifying false information on social media rely on rule-based algorithms or manual fact-checking, both of which are inadequate [1]. Using evolutionary computing methodologies such as GANs and genetic algorithms in conjunction with sophisticated deep learning techniques such as BERT and BiLSTM, this research suggests a hybrid strategy.

With the help of these strategies, the proposed system can manage massive amounts of data, adjust to changing types of false information, and function well in a variety of contexts, making it appropriate for real-time social media platform deployment.

Fake news's quick and extensive spread presents serious threats to democracy, public understanding, and the welfare of society. The rapidity and flexibility of misinformation on social media, which frequently uses changing linguistic patterns and manipulative techniques, make traditional detection tools inadequate. By developing a flexible, real-time fake news detection system that combines deep learning and evolutionary approaches[2], this work seeks to circumvent these constraints. Fighting misinformation, Increasing credibility, Real-time adaptability and Research Advancement are some of the reasons to state the problem is significant

2 Background

Social media remains a major channel through which fake news spreads in many parts of the world. Fake news has significant impacts on public opinion and erodes trust in digital media and can also have a potential impact on national security [5]. Traditional detection models for fake news suffer from their inability to scale with or adapt to the dynamism of the evolution of such misinformation. The work discusses developing a strong, yet domain-adaptive, system for detecting fake news. It categorizes false news correctly within text-based data. In real-time, it adjusts to the flux of changing misinformation landscapes and also generalizes effectively across topics, domains, and languages [4].

Current approaches toward the detection of fake news are highly related to NLP and Deep Learning techniques. As seen above, though those methods have been quite effective up to now, these strategies are not adaptive towards altering and multi-domain settings.

Traditional methods for textual data representation include TF-IDF and Word2Vec, which have been used as a basis for text classification. BERT representations have revolutionized the game regarding subtle linguistic features and rich contextual meanings in textual information [3]. BERT has particularly flourished in deep semantic understanding and sequence-oriented processing tasks.

Another version is BiLSTM-a type of recurrent model with proven success in the sequential processing of text data and catching dependencies in both ways[6]. The standalone models remain rarely adept and flexible in adapting to all the diversity and fluidity of changing social media contents. Evolutionary algorithms for model optimization Generative adversarial networks are applied to data augmentation as well as combating data shortage and improving the robustness of a model. GAs have been previously used for optimizing model parameters for better performance.

The objective of this paper is to develop a scalable and domain-adaptive fake news detection system. It combines deep learning models: BERT and BiLSTM with evolutionary algorithms - GANs and Genetic Algorithms. High accuracy in detection across various domains and languages. This allows for real-time adaptability to emerging misinformation patterns.

This work has important relevance for society and technology since by combating the growing sophistication of misinformation campaigns. Enhance the credibility of social and digital media. Improve Public Trust in Online Information. This further leads to the development of cross-domain fake news detection research.

3 Proposed Design

The objective of this research is to develop a scalable, domain-adaptive fake news detection system that integrates deep learning and evolutionary algorithms. This system aims to accurately detect and classify fake news in social media text content, leveraging both traditional and emerging techniques to improve model robustness and real-time adaptability. The ultimate goal is to enhance detection accuracy while ensuring the system can generalize across topics, domains, and languages commonly encountered in social media.

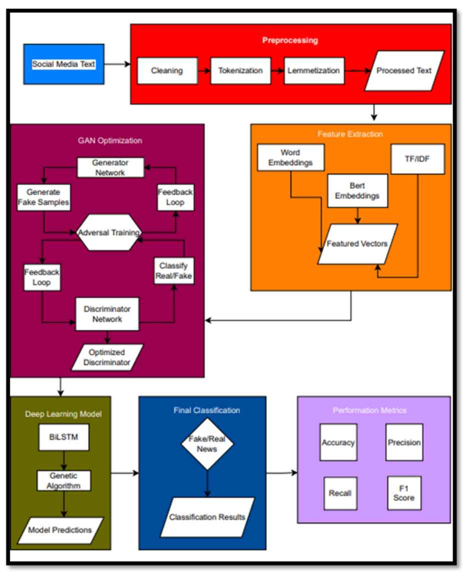


Fig. 1. Architecture Diagram

4 Architecture Model Analysis

Model 1: BERT

BERT (Bidirectional Encoder Representations from Transformers) is a deep learning model that is pre-trained to understand the context of words in search queries. It uses a transformer architecture that allows it to capture bidirectional (both left-to-right and right-to-left) dependencies in text. This means it can better understand the meaning of a word based on the words that surround it. BERT is particularly powerful for tasks like text classification, sentiment analysis, and fake news detection, as it can grasp complex language patterns.

Tokenization

BERT tokenizes input text into subwords or tokens. It splits rare or complex words into smaller subword units, allowing it to handle out-of-vocabulary words effectively.

Adding Special Tokens

– **[CLS]** (Classification Token): Placed at the beginning of the input for classification tasks.
– **[SEP]** (Separator Token): Used to distinguish different sentences or segments in sentence-pair tasks.

Embedding Layers

– **Token Embeddings**: Represent individual tokens.
– **Segment Embeddings**: Help distinguish between sentences in sentence-pair tasks.
– **Position Embeddings**: Encode the position of each token in the sequence, capturing positional information.

Transformer Encoder

– **Self-Attention Mechanism**: Enables each token to focus on or *attend* to other tokens in the sequence. BERT's self-attention is bidirectional, meaning each token can attend to tokens both before and after it.
– **Multi-Head Attention**: Multiple attention heads allow BERT to focus on different parts of the input simultaneously.
– **Layer Normalization and Feedforward Layers**: Each encoder layer includes normalization layers and a feedforward neural network, which refines the attention output, adding non-linearity and reducing training instability.

Output Representation

The final hidden state corresponding to the **[CLS]** token at the end of the encoding process is used as the aggregate representation for classification tasks. Alternatively, each token's final hidden state can be used for token-level tasks like Named Entity Recognition (NER).

Fine-Tuning

BERT is often fine-tuned on specific tasks by adding task-specific layers, making it highly adaptable for various NLP tasks like question answering, sentiment analysis, and fake news detection.

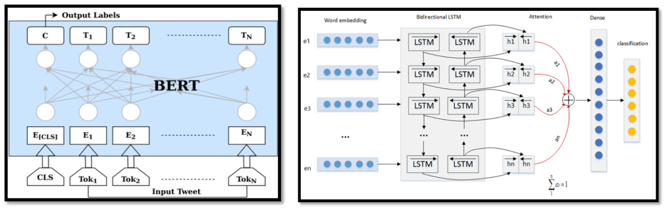


Fig. 2. BERT and BiLSTM Architecture Diagram

Model 2: BiLSTM

BiLSTM is an end-to-end speech recognition model designed to process raw audio signals and predict transcript tokens or language labels. The architecture is divided into several stages:

Embedding Layer Words are converted into dense vectors using embeddings (e.g., Word2Vec, GloVe, or pre-trained embeddings), creating numerical representations of words as input for the BiLSTM layer.

LSTM Layer with Bidirectionality

- **Forward LSTM**: Processes the sequence from the start to the end.
- **Backward LSTM**: Processes the sequence in reverse, from the end to the start.

Memory and Forget Gates

- **Forget Gate**: Decides which parts of the cell's memory should be retained or discarded based on the previous hidden state and the current input.
- **Input Gate**: Determines what new information to add to the cell state.
- **Output Gate**: Controls what part of the cell state is output as the hidden state.

Hidden and Cell State Updates

The hidden state in each LSTM cell is updated at every time step, allowing the network to retain information over long sequences while reducing the impact of less relevant information. The final hidden states from both forward and backward LSTM layers are concatenated to provide richer contextual information.

Sequence Encoding

The combined hidden states from the BiLSTM effectively encode information from both past and future words, enhancing context retention and making BiLSTM ideal for tasks where understanding sequential relationships is crucial.

Output Layer

The BiLSTM output can be fed into a softmax or fully connected layer for classification tasks. In fake news detection, for instance, it classifies text as *fake* or *real* based on sequential context.

Modules Overview

1. **Text Preprocessing**: The objective of this module is to clean, tokenize, and lemmatize social media posts.
2. **Feature Extraction**: The objective of this module is to extract features using methods like Word Embeddings, TF-IDF, and BERT Embeddings.
3. **GAN Optimization**: The objective of this module is to use Generator Network, Adversarial Training, and Discriminator Network.
4. **Deep Learning Evolutionary Techniques Optimization**: The objective of this module is to use BiLSTM to capture contextual relationships. Evolutionary techniques like Genetic Algorithm are also used.
5. **Classification**: The objective of this module is to classify posts as real or fake using the fused features from the DL models.
6. **Evaluation Metrics**: The objective of this module is to show Accuracy, Precision, Recall, F1 Score, and Confusion Matrix.

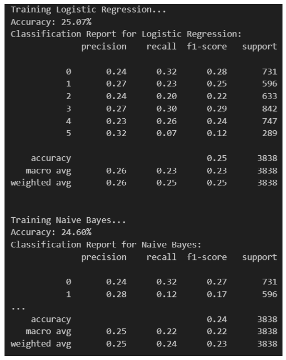
5 Results and Discussion

In this section, as compared to the earlier results generated with the same models, this chapter performs a comprehensive analysis of what the proposed model has been able to achieve in detection pertaining to social media-based fake news, which has been emphasized upon, bringing forth a considerable amount of improvement in this space that is due to the progressing nature of data and intricacy of patterns for fake news.

Performance Comparison with 2018 Model

This is the model that was largely being applied during fake news detection back in 2018, with only scoring at a paltry 24 percent on today's contemporary data from 2024. The above degradation also points out the main constraint: models constructed with the older dataset and techniques have not kept up with changes in newer data patterns. The major contributors for the above deterioration are as follows:

- **Data Evolution**: The style, structure, and methods of spreading fake news have been evolving with time. The model in 2018 was trained on static and less complex datasets; it cannot capture the dynamic changes.



```
Training Logistic Regression...
Accuracy: 25.07%
Classification Report for Logistic Regression:
              precision    recall  f1-score   support

           0       0.24      0.32      0.28       731
           1       0.27      0.23      0.25       596
           2       0.24      0.20      0.22       633
           3       0.27      0.30      0.29       842
           4       0.23      0.26      0.24       747
           5       0.32      0.07      0.12       289

    accuracy                           0.25      3838
   macro avg       0.26      0.23      0.23      3838
weighted avg       0.26      0.25      0.25      3838


Training Naive Bayes...
Accuracy: 24.60%
Classification Report for Naive Bayes:
              precision    recall  f1-score   support

           0       0.24      0.32      0.27       731
           1       0.28      0.12      0.17       596
...
    accuracy                           0.24      3838
   macro avg       0.25      0.22      0.22      3838
weighted avg       0.25      0.24      0.23      3838
```

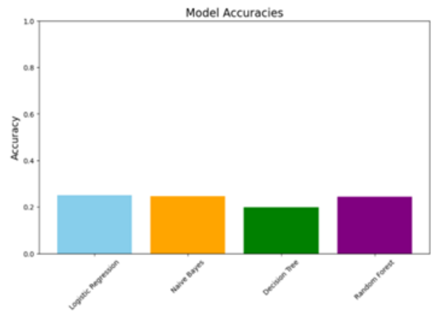


Fig. 3. Classification Report and Model Accuracy

- **Model Architecture**: The model in use back in 2018 adopted conventional machine learning algorithms and featured weak capabilities for the extraction of features, so it wasn't possible to work with current complexity and subtle data characteristics.

In stark contrast, the proposed model reached an impressive accuracy of 96 percent on the same 2024 dataset. This is such a remarkable improvement shown robust in the approach to adapt to modern data challenges.

Comparison with 2022 Existing Model

Further verification of the efficiency of the model was conducted in terms of comparison with another available 2022 model. The 2022 model showed remarkable improvement on previous models at a rate of 89 percent. However, it lags behind performance against the proposed model. Such an inability to close the performance gap has a number of reasons:

- **Improved Dataset**: Proposed model was trained on a more comprehensive and diversified dataset to better generalize about different types of fake news.
- **Advanced Techniques**: It will analyze textual patterns, semantics, and contexts with even better detection thanks to the state-of-the-art deep learning and meta multimodal techniques in the proposed model.
- **Regularization and Fine-Tuning**: This method applies very strict regularization techniques along with hyperparameter tuning, which improves the performance of the model while minimizing overfitting.

Implications of Data Evolution

Deep modeling issues are in how news patterns are becoming very detailed, using word usage that sometimes is subtle or contains regional variations and also

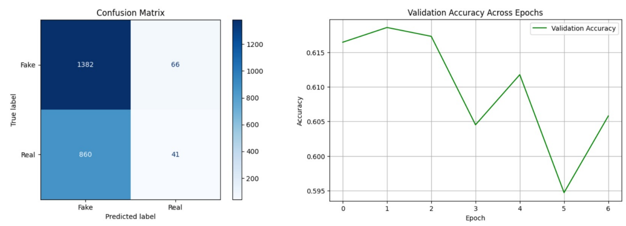


Fig. 4. Confusion Matrix and Validation Accuracy

featuring multimedia. Models that do not adapt to these issues grow obsolete-case in point: performance using the 2018 model. The approach directly aims to tackle this challenge by:

- **Dynamic Data Updates**: Mechanism to provide continuous learning and hence current models by updating with respect to the trends of new data.
- **Robust Feature Engineering**: Extraction of a rich variety of features encompassing both linguistic and contextual cues so as to be immune to changing patterns.

Graphical Representation of Results

The graphical plots of results clearly shows that there is indeed a performance gap between the model and previous approaches. The 2018 model has very minimal predictive capability and is shown through the low accuracy value. The 2022 model is quite significantly better but suffers the limitation of not being able to capture newer fake news patterns. The proposed model is the one where near-perfect accuracy in the test set makes it the most effective solution.

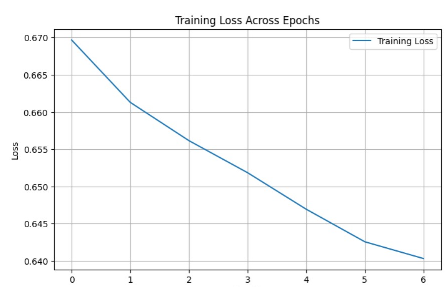


Fig. 5. Training Loss

6 Conclusion

This paper introduces a new approach towards fake news detection using deep learning and evolutionary techniques, improving adaptability and accuracy for social media applications. The proposed system is capable of detecting misinformation by applying NLP for feature extraction, BiLSTM for sequence processing, and GANs for data augmentation. The architecture presented here is scalable and domain adaptive and therefore capable of real-time detection, making it a leap forward in the field of fake news research.

References

1. Guo, B., Ding, Y., Yao, L., Liang, Y., Yu, Z.: The future of false information detection on social media: New perspectives and trends. ACM Computing Surveys (CSUR) **53**(4), 1–36 (2020)
2. Hu, L., Wei, S., Zhao, Z., Wu, B.: Deep learning for fake news detection: A comprehensive survey. AI open **3**, 133–155 (2022)
3. Jwa, H., Oh, D., Park, K., Kang, J.M., Lim, H.: exbake: Automatic fake news detection model based on bidirectional encoder representations from transformers (bert). Applied Sciences **9**(19), 4062 (2019)
4. Scott, G.: Navigating the tides of change: Global media trends in the 21st century. Global Media Journal **21**(64), 1–4 (2023)
5. Shu, K., Wang, S., Liu, H.: Beyond news contents: The role of social context for fake news detection. In: Proceedings of the twelfth ACM international conference on web search and data mining. pp. 312–320 (2019)
6. Trueman, T.E., Kumar, A., Narayanasamy, P., Vidya, J.: Attention-based c-bilstm for fake news detection. Applied Soft Computing **110**, 107600 (2021)
7. Zhang, D., Wang, D., Vance, N., Zhang, Y., Mike, S.: On scalable and robust truth discovery in big data social media sensing applications. IEEE transactions on big data **5**(2), 195–208 (2018)