

An Enhanced Deep Learning Framework for Fake News Detection Using Sequential and Transformer-Based Models

*Sheetal Adhikari
**Sushma Kushwaha

ABSTRACT

The rapid growth of online news and social-media communication has made the timely identification of fabricated information an important natural language processing problem. This study develops and evaluates a deep-learning-based framework for classifying news as genuine or fake. The investigation considers three sequential architectures—Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and a CNN-BiLSTM hybrid—and compares them with a fine-tuned Bidirectional Encoder Representations from Transformers (BERT) model. Experiments are performed using the ISOT Fake News dataset after a common text-cleaning and title-based preprocessing procedure. The combined data are divided into training, validation, and test subsets using a 64:16:20 ratio. In addition to the original class distribution, an intentionally imbalanced version is constructed to examine model behavior under a more challenging class distribution. Across both settings, the transformer-based BERT classifier achieves the strongest reported performance. On the balanced test set, BERT obtains an accuracy of 98.74%, while the corresponding accuracy on the imbalanced setting is 99.03%. The findings indicate that contextual representations learned by a transformer can provide a substantial advantage over conventional recurrent architectures for title-level fake-news classification. The study also examines incorrectly classified samples to identify directions for improving robustness and practical reliability.

Keywords:- Fake News Detection, Deep Learning, LSTM, BiLSTM, CNN-BiLSTM, BERT, Text Classification.

*Sheetal Adhikari, CSE Dept., REC, Bhopal, India, sheetaladhikari253@gmail.com.

**Sushma Kushwaha, Professor, CSE Dept., REC, Bhopal, India, sushma.svcst@gmail.com.

I. INTRODUCTION

False or misleading information has existed for centuries, but the scale and speed of its circulation have changed substantially with the emergence of digital communication platforms. In contemporary information ecosystems, fabricated stories can be distributed almost instantly through social networks, websites, messaging services, and other online channels. Such content may mislead readers, influence public opinion, create unnecessary social concern, and reduce confidence in reliable information sources [1], [2]. The problem therefore extends beyond simple text classification: an effective detection mechanism must distinguish misleading content from legitimate reporting while operating on large volumes of rapidly changing information.

The expansion of social media has further increased the accessibility of news. A large proportion of users obtain at least some news through social platforms, making online environments important channels for information consumption and redistribution [3]. At the same time, the low cost of publishing and sharing digital content provides favorable conditions for the rapid propagation of unverified or fabricated claims [4]. Manual verification alone is consequently difficult to scale. Automated fake-news detection systems can assist this process by learning linguistic and semantic patterns associated with different classes of news.

Research in this area has explored several families of methods. Traditional machine-learning systems commonly rely on manually engineered lexical, stylistic, semantic, or source-related features. Deep neural networks, including convolutional and recurrent models, attempt to learn useful representations directly from text. More recently, pretrained transformer language models such as BERT have demonstrated strong contextual modelling capabilities and have reduced the dependence on task-specific feature engineering [5]. Other pretrained language models, including GPT-based and ELMo-based approaches, have also influenced the development of modern NLP systems [6], [7].

Despite these developments, fake-news detection remains challenging because the linguistic characteristics of fabricated and genuine news can overlap. Headlines may contain incomplete context, persuasive language, named entities, emotionally charged expressions, or deliberately constructed wording. Models trained on a single dataset can also perform differently when the class distribution changes. These observations motivate a systematic comparison of recurrent, hybrid, and transformer-based approaches under both balanced and imbalanced conditions.

Accordingly, this work investigates LSTM, BiLSTM, CNN-BiLSTM, and BERT for title-based fake-news classification. The proposed experimental framework combines standardized preprocessing, model-specific tokenization and embedding procedures, fine-tuning of a pretrained BERT classifier, and evaluation using accuracy, precision, recall, and F1-score. The main contribution is an empirical comparison of these architectures on the ISOT dataset, together with an examination of performance under an intentionally imbalanced data configuration.

The primary objective of this study is to develop and experimentally evaluate a deep learning-based framework for automated fake-news detection using textual information from news headlines. The specific objectives of the research are as follows:

- **To develop an effective pre-processing pipeline** for cleaning and normalizing news headlines so that textual information can be used efficiently by deep learning models.
- **To investigate sequential deep learning approaches** by implementing Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (BiLSTM) models for fake-news classification.
- **To develop a hybrid CNN-BiLSTM model** that combines convolution-based local feature extraction with bidirectional sequence modelling for improved representation of news headlines.
- **To employ a pretrained BERT model** and fine-tune it for binary classification of genuine and fake news, thereby exploiting contextual representations learned through transformer-based self-attention.
- **To comparatively evaluate the performance** of LSTM, BiLSTM, CNN-BiLSTM, and BERT using accuracy, precision, recall, and F1-score on the ISOT Fake News dataset.
- **To examine model performance under different class distributions** by comparing the original dataset with an intentionally imbalanced configuration that more closely represents a challenging classification setting.
- **To analyse classification errors and identify limitations** of the best-performing model, particularly with respect to the difficulty of distinguishing fabricated and genuine headlines using title-level textual information alone.

Through these objectives, the study aims to establish a systematic comparison of recurrent, hybrid, and transformer-based approaches and determine the effectiveness of contextual transformer representations for fake-news detection.

The remainder of the paper is organized as follows. Section 2 reviews relevant research on fake-news detection. Section 3 describes the dataset, preprocessing strategy, representation methods, and proposed workflow. Section 4 discusses the evaluated deep-learning and transformer architectures. Section 5 presents the experimental design and results, including the imbalanced-data study and analysis of classification errors. Section 6 concludes the paper and identifies directions for further investigation.

II. RELATED WORK

Fake-news detection has developed into a multidisciplinary research area involving natural language processing, machine learning, social-media analysis, information retrieval, and human-centered methods. Existing studies differ considerably in the type of evidence used for classification. Some methods concentrate on textual content, whereas others incorporate social interactions, source information, multimodal signals, graph structures, or human judgments.

Saleh et al. [9] investigated OPCNN-FAKE, a convolutional neural-network-based approach designed specifically for fake-news identification. Such methods demonstrate the value of neural representation learning for discovering discriminative patterns in textual data. Gupta et al. [10] examined the broader socio-technical problem of misinformation and discussed interventions involving multiple stakeholders. Their work highlights that technical classification is only one component of a larger misinformation-mitigation ecosystem.

Kemp et al. [11] studied the cognitive consequences of recalling misinformation when authentic-news corrections are encountered. This line of work is important because it demonstrates that misinformation detection and correction cannot be considered solely as computational problems; the way information is remembered and updated by users also affects its impact. Fang et al. [12] proposed NSEP for early fake-news detection using semantic-environment perception, emphasizing the importance of contextual information when only limited evidence is available.

Alghamdi et al. [13] explored transformer-based techniques for COVID-19-related fake-news detection. Their work illustrates the applicability of contextual language representations to domain-specific misinformation. Rai et al. [14] combined transformer-based methods with enhanced LSTM and BERT architectures, providing evidence that recurrent and transformer representations can be integrated for classification. Uppada et al. [15] considered fake news together with fake-account detection and incorporated social engagement and visual-content information, demonstrating the value of moving beyond text-only signals.

Other studies have investigated complementary directions. Wang et al. [16] proposed blockchain-based mechanisms for traceability and verification of news, emphasizing provenance and authentication rather than text classification alone. Hu et al. [17] examined causal inference for exploiting image-text matching bias in multimodal fake-news detection. Soga et al. [18] combined stance similarity with graph neural networks, indicating that relationships among news items or users can provide useful contextual evidence. Allein et al. [19] addressed ethical considerations and the risk of profiling in fake-news detection systems.

Research has also examined the human and behavioral dimensions of misinformation. Omar et al. [20] investigated intrinsic and extrinsic factors associated with fake-news sharing and considered the moderating role of awareness. Bonet-Jover et al. [21] used a Human-in-the-Loop strategy to construct a dataset for content reliability assessment, illustrating how human judgments can complement automated methods. Collectively, these studies show that fake-news detection can benefit from linguistic, contextual, social, multimodal, provenance-based, and human-centered evidence.

A limitation of many existing approaches is that the computational architecture and the information source are often tightly coupled. Text-only models may not have access to social or source-level evidence, while multimodal or graph-based methods can require additional metadata that is unavailable in many practical settings. The present study therefore focuses on a reproducible text-only setting and compares increasingly expressive representation mechanisms: recurrent sequence modelling, convolutional-recurrent feature extraction, and transformer-based contextual encoding. This provides a controlled basis for assessing the contribution of architectural complexity to title-level fake-news classification.

III. METHODOLOGY

The experimental workflow consists of four principal stages: dataset preparation, textual preprocessing, representation learning, and model training and evaluation. The same underlying data partitions are used for all architectures so that the comparison is based primarily on differences in model representation and learning capability.

3.1 Dataset

The dataset we use is the ISOT Fake News dataset introduced by ISOT Research Lab at University of Victoria in Canada [21]. This dataset is a compilation of several thousand fake news and truthful articles, obtained from different legitimate news sites and sites flagged as unreliable by Politifact.com.

To get insight into this dataset, we visualized it with word clouds for real and fake news respectively. **Figure 1(a)** shows the word cloud of the real news in the dataset, and **Figure 1(b)** shows the one of the fake news in the dataset.

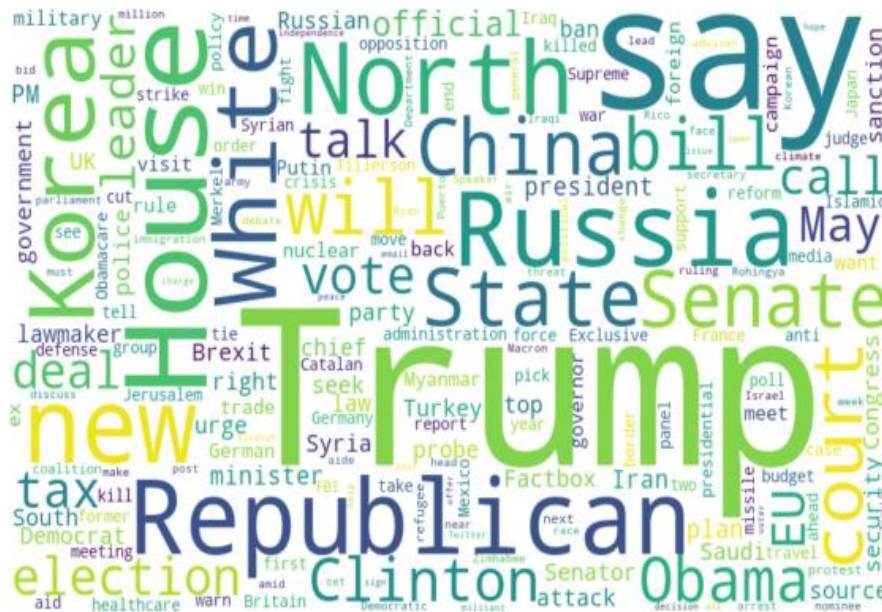


Figure 1(a)

We can see that, in the real news word cloud, ‘Trump’, ‘say’, ‘Russia’, ‘House’, ‘North’, and Korea’ appeared frequently; while in fake news one, ‘VIDEO’, ‘Trump’, ‘Obama’, ‘WATCH’, and ‘Hillary’ appeared the most frequently. ‘Say’ appears frequently in real news but does not in fake news. ‘VIDEO’, and ‘WATCH’ appear frequently in fake news but do not in real news. From these two word clouds, we can get some important information to differentiate the two classes of data.

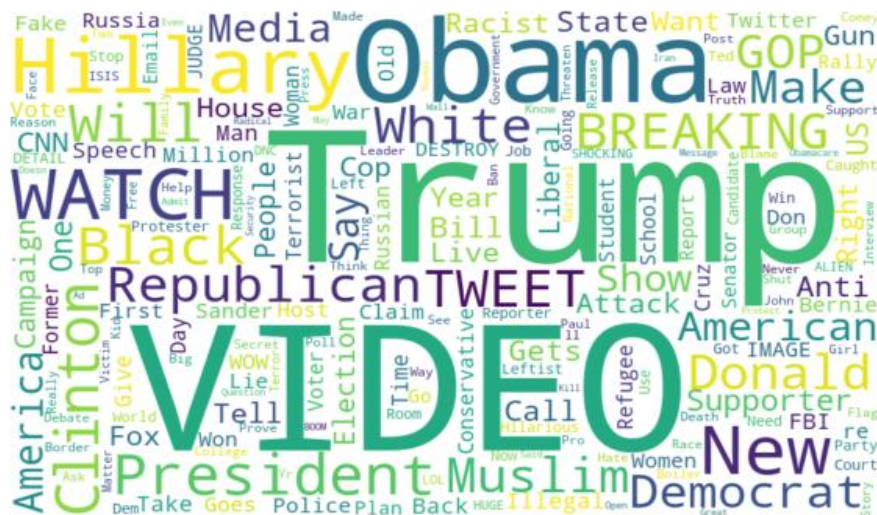


Figure 1(b).

Figure 1 (a) & (b): Word Cloud of Real&fake news

The original form of the dataset is two CSV files containing fake and real news respectively. We combined the dataset and split it into training, validation, and test sets with shuffling at the ratio of 64%:16%:20%. The original combined dataset contains 44,898 pieces of data, and **Table 1** shows the distribution of data in the training, validation, and test sets.

Table1: Distribution of Dataset

Subset	Percentage	Number of samples
Training	64%	28,734
Validation	16%	7,184
Test	20%	8,980
Total	100%	44,898

3.2 Data Preprocessing

The dataset contains both headlines and full article text. The average headline length is approximately 12.45 words, whereas the average article length is approximately 405.28 words. Because processing the complete articles substantially increases the computational burden and introduces large amounts of additional information, the present experiments use news titles as the principal textual input. This also provides a consistent input representation across the evaluated architectures.

A three-stage normalization procedure is applied. First, characters outside the English alphabetic range are replaced with whitespace. Second, all alphabetic characters are converted to lowercase. Third, common inflectional endings such as “ed,” “est,” “s,” and “ing” are removed during stemming. After normalization, titles are limited to a maximum of 42 words. The length restriction produces fixed-size model inputs after padding and reduces the effect of unusually long headlines.

It should be noted that title-only classification is a deliberate experimental constraint. Although it improves computational efficiency and provides a controlled comparison, it also removes article-level evidence that could be useful for distinguishing subtle cases. This limitation is considered when interpreting the reported results.

3.3 Text Representation and Word Embedding

Different architectures require different input representations. For LSTM, BiLSTM, and CNN-BiLSTM, the normalized titles are tokenized into integer sequences. Each sequence is then zero-padded to a length of 42 tokens. A trainable embedding layer maps the token indices into dense vector representations. The embedding parameters are initialized and subsequently learned during model training, allowing the networks to adapt their representations to the classification task [22].

BERT uses a different tokenization strategy. The BERT tokenizer applies WordPiece segmentation, which permits words outside the basic vocabulary to be represented through smaller subword units [23]. Special tokens are inserted to identify sequence boundaries. The tokenized input is represented through token IDs, segment information, and attention masks. The attention mask distinguishes meaningful tokens from padding, allowing the transformer to focus its computation on the relevant input positions.

The use of model-specific tokenization is important because recurrent architectures operate on learned token embeddings, whereas BERT relies on the contextual representations produced by its pretrained transformer encoder. Thus, the comparison evaluates not only different classifiers but also different mechanisms for representing linguistic context.

3.4 Proposed Experimental Algorithm

The complete experimental procedure can be summarized as follows:

- Collect the ISOT fake and real news files and merge them into a common dataset.
- Randomly partition the combined dataset into training, validation, and test subsets using a 64%:16%:20% ratio.
- Generate class-specific word clouds for exploratory examination of lexical differences.
- Normalize the titles by replacing non-alphabetic characters, converting text to lowercase, and applying stemming.
- Limit each title to a maximum of 42 words.
- For LSTM, BiLSTM, and CNN-BiLSTM, tokenize the titles, convert them to integer sequences, and pad each sequence to 42 positions.

- For BERT, apply WordPiece tokenization and construct the required token, segment, and attention-mask inputs.
- Train and evaluate LSTM, BiLSTM, CNN-BiLSTM, and fine-tuned BERT classifiers.
- Repeat the comparison using an intentionally imbalanced data configuration in which the fake-news class is reduced to approximately one-tenth of its original size.
- Compute accuracy, precision, recall, and F1-score for each model and inspect incorrectly classified BERT samples.

The overall workflow of the proposed fake-news detection framework is illustrated in Fig. 2. The framework begins with the collection and integration of fake and real news samples from the ISOT dataset, followed by systematic preprocessing of the news headlines. The preprocessed data are divided into training, validation, and testing subsets, after which four deep learning architectures—LSTM, BiLSTM, CNN-BiLSTM, and fine-tuned BERT—are trained and evaluated. Model performance is assessed using accuracy, precision, recall, and F1-score. To further examine model behavior under class imbalance, an additional experiment is performed by reducing the number of fake-news samples and retraining the models. Finally, the obtained results are comparatively analyzed, misclassified samples are examined, and the best-performing model is selected for final fake-news classification.

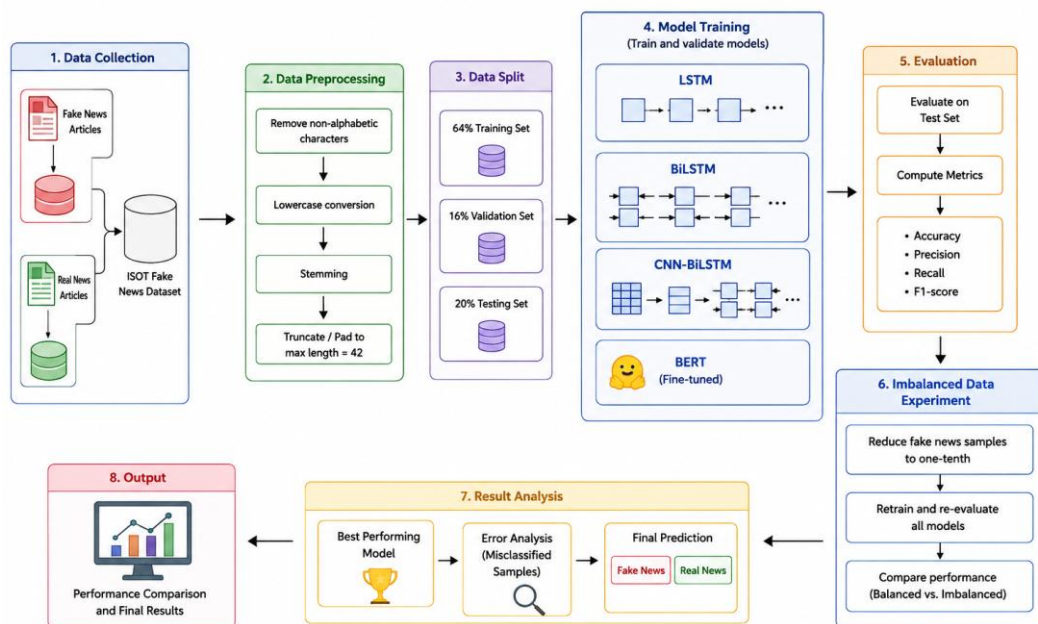


Figure 2: Flow diagram of the proposed algorithm for fake news detection.

IV. MODELS

Four architectures are evaluated to examine the effect of progressively richer sequence and contextual representations. LSTM provides a recurrent baseline, BiLSTM introduces information from both temporal directions, CNN-BiLSTM combines local feature extraction with bidirectional sequence modelling, and BERT provides pretrained transformer-based contextual representations.

4.1 LSTM

Long Short-Term Memory networks are recurrent neural architectures designed to preserve useful information over longer sequences through gated memory cells. In text classification, an LSTM processes tokens sequentially and learns dependencies between words. The architecture is therefore suitable for modelling the ordered structure of a news headline while mitigating some of the limitations associated with conventional recurrent networks.

4.2 Bidirectional LSTM

A Bidirectional LSTM contains two recurrent components that process the same sequence in opposite directions [24]. The forward component captures information from earlier to later positions, while the backward component captures dependencies in the reverse direction. Combining the two representations allows the classifier to use contextual information from both sides of a token, which can be beneficial when the meaning of a word depends on its surrounding headline context.

4.3 CNN-BiLSTM

The CNN-BiLSTM model combines convolutional feature extraction with bidirectional recurrent modelling. The convolutional stage is intended to identify local patterns and discriminative combinations of neighboring tokens. The resulting feature sequence is then supplied to the BiLSTM so that the model can integrate local patterns with information from both directions of the sequence. This hybrid design attempts to exploit complementary strengths of convolutional and recurrent architectures.

4.4 BERT

Bidirectional Encoder Representations from Transformers (BERT) is a pretrained transformer language model introduced for contextual language understanding [5]. Unlike a conventional

recurrent network that processes tokens sequentially, BERT uses self-attention to model relationships among tokens within the input sequence. Its bidirectional contextual representation allows each token to be interpreted using information from the broader sequence.

The experiments use the bert-base-uncased pretrained model. The model does not distinguish between uppercase and lowercase forms and provides a general-purpose contextual representation that can be adapted to downstream classification tasks. In this study, the pretrained representation is fine-tuned using the fake-news classification data rather than training a transformer from scratch.

4.5 BERT Fine-Tuning

The BERT model is adapted to binary sequence classification. The architecture includes the pretrained BERT encoder, pooling of the sequence representation, dropout for regularization, and a linear classification layer that produces logits for the two target classes. During fine-tuning, the pretrained parameters are updated using the task-specific training data so that the learned representations become specialized for distinguishing fake and genuine news headlines.

V. EXPERIMENTAL EVALUATION

5.1 Balanced and Imbalanced Data

Although the original dataset contains substantial numbers of both classes, real-world information environments may not exhibit the same class distribution. To investigate the effect of class imbalance, two experimental configurations are considered. The first retains the original distribution. In the second configuration, the fake-news training, validation, and test samples are reduced to approximately one-tenth of their original counts while the number of genuine-news samples is retained. This produces a more asymmetric setting for comparing model behavior. Table 2 shows the true/fake distribution of data in the original training set, imbalanced training set, original validation set, imbalanced validation set, original test set, and imbalanced test set.

Table 2: Distribution of Imbalanced Data

Class	Training Original	Training Imbalanced	Validation Original	Validation Imbalanced	Test Original	Test Imbalanced
True	13,765	13,765	3,409	3,409	4,243	4,243
Fake	14,969	1,497	3,775	378	4,737	474

5.2 Evaluation Measures

Model performance is assessed using the confusion-matrix components True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Here, TP denotes a fake-news item correctly classified as fake, TN denotes a genuine-news item correctly classified as genuine, FP denotes a genuine item incorrectly labelled as fake, and FN denotes a fake item incorrectly labelled as genuine.

Accuracy measures the proportion of all correctly classified examples. Precision measures the proportion of predicted fake-news examples that are actually fake, whereas recall measures the proportion of actual fake-news examples that are successfully identified. The F1-score provides a harmonic mean of precision and recall and is useful when both types of classification error are important. These measures together provide a more informative assessment than accuracy alone.

5.3 Results on the Balanced Dataset

Table 3: Performance of the Different Models on Balanced Datasets

Model	Accuracy	Precision	Recall	F1-Score
LSTM	0.9697	0.97	0.97	0.97
BiLSTM	0.9710	0.97	0.97	0.97
CNN-	0.9722	0.97	0.97	0.97

BiLSTM				
BERT	0.9874	0.99	0.99	0.99

The balanced-data results show a gradual improvement as the representation mechanism becomes more expressive. LSTM reaches 96.97% accuracy, while BiLSTM improves the result to 97.10%. The CNN-BiLSTM combination reaches 97.22%, suggesting that local convolutional patterns and bidirectional sequence information provide complementary benefits. BERT produces the strongest result, with 98.74% accuracy and approximately 0.99 precision, recall, and F1-score. The result supports the view that pretrained contextual representations can capture useful linguistic information that is difficult to learn from the task data alone. Table3 show the performance of the different models we used on balanced data sets with different metrics mentioned above, respectively.

5.4 Results on the Imbalanced Dataset

Table 4: Performance of the Different Models on Imbalanced Datasets

Model	Accuracy	Precision	Recall	F1-Score
LSTM	0.9780	0.98	0.98	0.98
BiLSTM	0.9799	0.98	0.98	0.98
CNN-BiLSTM	0.9791	0.98	0.98	0.98
BERT	0.9903	0.99	0.99	0.99

The imbalanced configuration produces accuracy values of 97.80%, 97.99%, 97.91%, and 99.03% for LSTM, BiLSTM, CNN-BiLSTM, and BERT, respectively. BERT again achieves the highest overall performance. Interestingly, the reported aggregate metrics are slightly higher under the imbalanced configuration. This observation should be interpreted cautiously because accuracy can be influenced by the class distribution; therefore, precision, recall, F1-score, and class-specific confusion-matrix analysis remain important when evaluating models intended for

deployment in naturally skewed environments. Table 4 shows the performance of the different models we used on imbalanced datasets with different metrics mentioned above, respectively.

The results also preserve the relative ranking observed in the balanced experiment. BiLSTM improves over the basic LSTM, and the CNN-BiLSTM hybrid remains competitive. The transformer-based classifier provides the most consistent overall performance across both experimental settings, indicating that its pretrained contextual representations are particularly effective for this title-level classification task.

5.5 Analysis of Misclassified Samples

High aggregate performance does not imply that every headline is classified correctly. An analysis of the BERT errors is therefore included to identify the remaining failure cases. The original analysis indicates that most incorrectly classified examples belong to the fake-news class, whereas a smaller proportion originates from the genuine-news class. This suggests that some fabricated headlines possess linguistic characteristics sufficiently similar to legitimate reporting to confuse the classifier. Figure 3 shows the insight of the data which is wrongly classified by our BERT-based model.

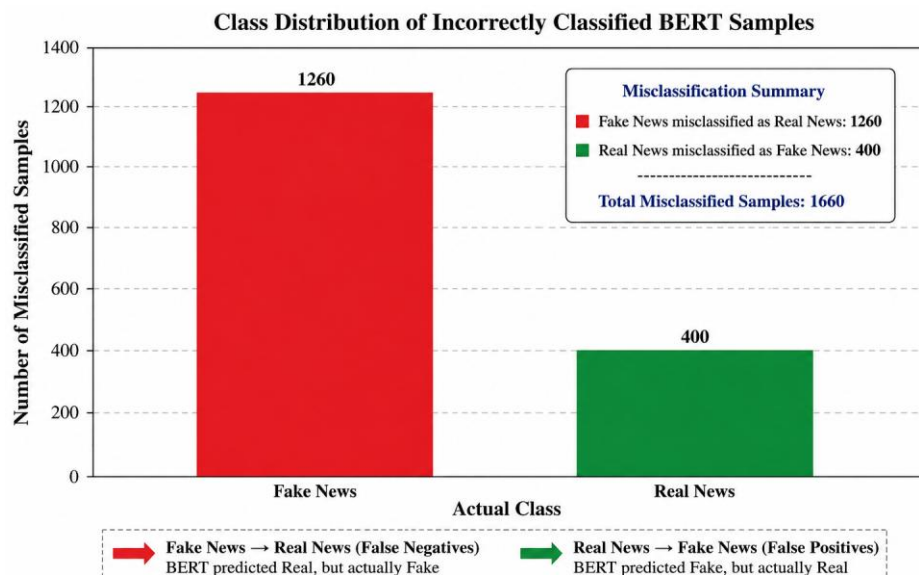


Figure 3: Class distribution of incorrectly classified BERT samples

The error pattern also indicates an important limitation of title-only classification. A headline can be ambiguous when viewed without the article body, source identity, publication history, or external evidence. A model may therefore assign a class based on linguistic regularities even when the factual status of the claim cannot be established from the headline itself. Future

experiments should examine article-level context, source credibility, temporal information, and external fact-checking evidence to address this limitation.

5.6 Comparative Discussion

The experimental comparison provides three main observations. First, the recurrent baseline already achieves high classification performance, indicating that headline-level lexical and sequential information contains substantial signal for the task. Second, bidirectional processing and convolutional feature extraction provide incremental gains over the basic LSTM architecture. Third, BERT substantially outperforms the recurrent and hybrid alternatives in both balanced and imbalanced settings.

The BERT advantage can be interpreted in terms of contextual representation. LSTM and BiLSTM learn task-specific sequence representations from the available training data, while BERT begins with representations acquired during large-scale pretraining. The self-attention mechanism also allows the model to consider relationships among tokens without relying exclusively on sequential recurrence. These properties are useful for headlines in which the meaning of a term may depend on multiple non-adjacent words.

At the same time, the reported results should not be interpreted as evidence that BERT alone solves fake-news detection in operational environments. The experiments are performed on a single benchmark dataset and use headlines rather than full articles. Dataset-specific vocabulary, source characteristics, duplicated or related content, and temporal effects can all influence performance. Consequently, cross-dataset validation and testing on current real-world news streams would be necessary before making stronger claims about generalization.

VI. CONCLUSION

This study presented an experimental framework for automated fake-news detection using LSTM, BiLSTM, CNN-BiLSTM, and BERT models. The ISOT Fake News dataset was processed through a common normalization pipeline, and news titles were used as the model input to maintain a manageable computational setting. The data were divided into training, validation, and test subsets, and both the original and an intentionally imbalanced class distribution were evaluated.

Among the evaluated architectures, BERT produced the best overall results. It achieved 98.74% accuracy on the balanced configuration and 99.03% accuracy on the imbalanced configuration,

while also obtaining approximately 0.99 precision, recall, and F1-score. The recurrent and hybrid models also performed strongly, with BiLSTM and CNN-BiLSTM showing incremental improvements over the basic LSTM model. These findings support the usefulness of pretrained transformer representations for title-based fake-news classification.

The study nevertheless has several limitations. The experiments use a single benchmark dataset, restrict the input to news titles, and do not incorporate external evidence such as source credibility, publication metadata, social propagation patterns, or factual knowledge bases. In addition, aggregate performance metrics do not fully describe the nature of individual model failures. Future work should therefore investigate cross-dataset and cross-domain evaluation, article-level modelling, multimodal evidence, explainable predictions, robustness to adversarial or deliberately manipulated headlines, and deployment-oriented efficiency. A broader evaluation on contemporary and noisy real-world news streams would provide stronger evidence regarding practical generalization.

Overall, the experimental findings indicate that transformer-based contextual modelling offers a strong foundation for fake-news classification, while also demonstrating that reliable misinformation detection requires evaluation beyond a single benchmark and beyond headline text alone.

REFERENCES

1. S. De Sarkar, F. Yang, and A. Mukherjee, "Attending sentences to detect satirical fake news," in Proc. 27th Int. Conf. Comput. Linguistics (COLING), Santa Fe, NM, USA, 2018, pp. 3371–3380.
2. H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," J. Econ. Perspect., vol. 31, no. 2, pp. 211–236, 2017, doi: 10.1257/jep.31.2.211.
3. A. Moon, "Two-thirds of American adults get news from social media: Survey," Reuters, 2017. [Online]. Available: Reuters news archive. [No DOI verified.]
4. K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," ACM SIGKDD Explor. Newsl., vol. 19, no. 1, pp. 22–36, 2017, doi: 10.1145/3137597.3137600.
5. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. 2019 Conf. North Amer. Chapter Assoc. Comput. Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2019, pp. 4171–4186, doi: 10.18653/v1/N19-1423.
6. A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," OpenAI, Tech. Rep., 2018. [No DOI verified.]
7. J. Sarzynska-Wawer, A. Wawer, A. Pawlak, J. Szymanowska, I. Stefaniak, M. Jarkiewicz, and L. Okruszek, "Detecting formal thought disorder by deep contextualized word representations," Psychiatry Res., vol. 304, Art. no. 114135, 2021, doi: 10.1016/j.psychres.2021.114135.

8. H. Saleh, A. Alharbi, and S. H. Alsamhi, "OPCNN-FAKE: Optimized convolutional neural network for fake news detection," *IEEE Access*, vol. 9, pp. 129471–129489, 2021, doi: 10.1109/ACCESS.2021.3112806.
9. A. Gupta et al., "Combating fake news: Stakeholder interventions and potential solutions," *IEEE Access*, vol. 10, pp. 78268–78289, 2022, doi: 10.1109/ACCESS.2022.3193670.
10. P. L. Kemp, T. R. Alexander, and C. N. Wahlheim, "Recalling fake news during real news corrections can impair or enhance memory updating: The role of recollection-based retrieval," *Cogn. Res.: Princ. Implic.*, vol. 7, Art. no. 85, 2022, doi: 10.1186/s41235-022-00434-1.
11. X. Fang, H. Wu, J. Jing, Y. Meng, B. Yu, H. Yu, and H. Zhang, "NSEP: Early fake news detection via news semantic environment perception," *Inf. Process. Manage.*, vol. 61, no. 2, Art. no. 103594, 2024, doi: 10.1016/j.ipm.2023.103594.
12. J. Alghamdi, Y. Lin, and S. Luo, "Towards COVID-19 fake news detection using transformer-based models," *Knowl.-Based Syst.*, vol. 274, Art. no. 110642, 2023, doi: 10.1016/j.knosys.2023.110642.
13. N. Rai, D. Kumar, N. Kaushik, C. Raj, and A. Ali, "Fake news classification using transformer based enhanced LSTM and BERT," *Int. J. Cogn. Comput. Eng.*, vol. 3, pp. 98–105, 2022, doi: 10.1016/j.ijcce.2022.03.003.
14. S. K. Uppada, K. Manasa, B. Vidhathi, R. Harini, and B. Sivaselvan, "Novel approaches to fake news and fake account detection in OSNs: User social engagement and visual content centric model," *Soc. Netw. Anal. Mining*, vol. 12, no. 1, Art. no. 52, 2022, doi: 10.1007/s13278-022-00878-9.
15. X. Wang, H. Xie, S. Ji, L. Liu, and D. Huang, "Blockchain-based fake news traceability and verification mechanism," *Heliyon*, vol. 9, no. 7, Art. no. e17084, 2023, doi: 10.1016/j.heliyon.2023.e17084.
16. L. Hu, Z. Chen, Z. Zhao, J. Yin, and L. Nie, "Causal inference for leveraging image-text matching bias in multi-modal fake news detection," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 11, pp. 11141–11152, 2023, doi: 10.1109/TKDE.2022.3231338.
17. K. Soga, S. Yoshida, and M. Muneyasu, "Exploiting stance similarity and graph neural networks for fake news detection," *Pattern Recognit. Lett.*, vol. 177, pp. 26–32, 2024, doi: 10.1016/j.patrec.2023.11.019.
18. L. Allein, M.-F. Moens, and D. Perrotta, "Preventing profiling for ethical fake news detection," *Inf. Process. Manage.*, vol. 60, no. 2, Art. no. 103206, 2023, doi: 10.1016/j.ipm.2022.103206.
19. B. Omar, O. D. Apuke, and Z. M. Nor, "The intrinsic and extrinsic factors predicting fake news sharing among social media users: The moderating role of fake news awareness," *Current Psychology*, vol. 43, no. 2, pp. 1235–1247, 2024, doi: 10.1007/s12144-023-04343-4.
20. A. Bonet-Jover, R. Sepúlveda-Torres, E. Saquete, P. Martínez-Barco, A. Piad-Morffis, and S. Estevez-Velarde, "Applying Human-in-the-Loop to construct a dataset for determining content reliability to combat fake news," *Eng. Appl. Artif. Intell.*, vol. 126, Art. no. 107152, 2023, doi: 10.1016/j.engappai.2023.107152.
21. H. Ahmed, I. Traore, and S. Saad, "Detecting opinion spams and fake news using text classification," *Security and Privacy*, vol. 1, no. 1, Art. no. e9, 2018, doi: 10.1002/spy2.9.
22. A. Akbik, T. Bergmann, D. Blythe, K. Rasul, S. Schweter, and R. Vollgraf, "FLAIR: An easy-to-use framework for state-of-the-art NLP," in *Proc. 2019 Conf. North Amer. Chapter Assoc. Comput. Linguistics: Demonstrations*, Minneapolis, MN, USA, 2019, pp. 54–59, doi: 10.18653/v1/N19-4010.
23. Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, et al., "Google's neural machine translation system: Bridging the gap between human and machine translation," *arXiv:1609.08144*, 2016. [No DOI verified for the cited arXiv version.]

24. A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Networks*, vol. 18, nos. 5–6, pp. 602–610, 2005, doi: 10.1016/j.neunet.2005.06.042.