IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 6, Issue 4, January 2026

Impact Factor: 7.67

Survey on Fake News Detection

Soham Sathe[*], Amaan Shaikh[†], Sahil Shrotri[‡], Vaishnavi Thakur[§], A. A. Chandorkar[¶]

Dept. of Computer Engineering,

Pune Institute of Computer Technology Pune, India

sohamsathe2608@gmail.com[*], amaanmazhar211@gmail.com[†], sahil.shrotri07@gmail.com[‡], vaishnavithakur77777@gmail.com[§], aachandorkar@pict.edu[¶]

Abstract: *Detecting fake news is an important aspect of natural language processing (NLP) with implications for information integrity, public opinion, and societal trust. In this paper, we explore and compare multiple approaches for fake news detection using a common dataset. We analyse the performance of (1) baseline machine learning models, (2) deep learning models, and (3) transformer-based pre-trained models on shared evaluation metrics. Classical classifiers are first implemented as baselines. Deep learning methods are then employed to capture sequential dependencies in text. Finally, we evaluate transformer architectures, focusing on BERT and RoBERTa, which leverage large-scale pretraining and contextual embeddings to achieve state-of-the-art accuracy.*

Keywords: Fake News Detection, Natural Language Processing, Machine Learning, Deep Learning, Transformers, BERT, RoBERTa

I. INTRODUCTION

Communication and information sharing are crucial in today's digital era, where most of the communication is carried out through digital means throughout the world. Due to this digitalization, communication is faster than ever. This is advantageous as it means news from one corner of the Earth can reach the other in no time through different social media platforms. This being said, speed also has its own disadvantages as it means the faster spread of fake news, stories, rumours etc.

Fake news can be defined as any piece of information that is fabricated and does not depict the truth. This information can be fully made up or partially fabricated, which might be contrary to the factual truth. It could be misinformation (unintentional) or disinformation (intentional). Fake news has become a critical issue in today's time because of the large usage of social media platforms like X(formerly Twitter), Facebook, Instagram, etc. Rapidly spreading news has the power to change public opinion on sensitive topics like religion, politics, national and international conflicts etc. It can also affect social stability within a country. Manual fact checking can be time-consuming and resource intensive and thus there is a need for an automated detection system for fake news.

There a couple of challenges for fake news detection. The writing styles of fake news continue to evolve, which makes it difficult to detect. Sometimes the language itself is ambiguous where the difference between fake and satire is subtle making it difficult to identify whether the intention is harmful or humorous criticism of society, politics or individuals. For high precision in detection, models should understand context and have current world knowledge. Also, training models on imbalanced datasets may induce biasing in them and they would not perform well in a more generalised environment as fake news articles for training are generally less available.

Initially we study the classic ML models like Naive Bayes, Random Forest, Logistic Regression, etc. and move on to deep learning based models. Finally, we move ahead to the transformer-based architecture models like BERT and RoBERTa. Thus through this paper, we try to categorise existing methods, provide a timeline for progress, and serve as a reference for future research and practical applications.

II. RELATED WORK

Fake news detection has evolved from traditional feature-based and sequential models to recent transformer-based networks. Early deep learning approaches treated fake news detection as a text classification problem, employing CNN

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 4, January 2026 Impact Factor: 7.67

or LSTM encoders over tokenized news articles. Drif et al. [25] proposed a hybrid CNN–LSTM model, where CNN layers captured local word patterns and an LSTM modelled long-range dependencies, outperforming SVM and standalone CNN baselines. These traditional or hybrid approaches relied primarily on textual features without leveraging large-scale pre-trained models.

Graph-based architectures extended this line of research by incorporating relational context. Hu et al. [28] proposed CompareNet, a heterogeneous graph-attention model that integrates sentence, topic, and entity representations, while comparing entities against external knowledge bases to verify consistency. This GNN-based hybrid outperformed earlier text-only models on two English-language benchmarks.

Recent advances have been dominated by transformer-based pre-trained language models. In this paradigm, a model such as BERT or RoBERTa is fine-tuned on labeled news datasets. Kaliyar et al. [18] introduced FakeBERT, combining BERT with parallel CNN blocks to capture n-gram features, achieving high accuracy on social-media news. Similarly, Farokhian et al. [17] proposed MWPBert, a dual-stream model using two BERT encoders—one for headlines and one for body text—where selected article segments are concatenated for classification, outperforming prior baselines.

Other studies have evaluated lighter-weight or multilingual transformers. Raza et al. [8] reported that fine-tuned RoBERTa achieved higher accuracy, outperforming BERT-base and DistilBERT on datasets that were annotated using LLMs. Alqadi et al. [27] similarly observed RoBERTa surpassing BERT in English fake news classification on FakeNewsNet dataset [29] that contains articles taken from Politifact [31] and GossipCop (now discontinued) that are fact-checking websites.

In summary, earlier research emphasized CNN/LSTM-based models or graph-attention networks, while the state oftheart now primarily consists of transformer-based architectures and classifiers. Fine-tuned transformers consistently outperform traditional baselines, confirming their suitability for modern fake news detection pipelines.

III. DATASET

TABLE I: STATISTICS OF THE ISOT FAKE NEWS DATASET

News	Size (No.)	Type	Articles Size
Real-News	21417	World-News	10145
		Politics-News	11272
Fake-News	23481	Government-News	1570
		Middle-east	778
		US-News	783
		Left-News	4459
		Politics	6841
		News	9050

In our paper we use the ISOT dataset [6] [7] created by the University of Victoria's Information Security and Object Technology (ISOT) Research Lab to train and evaluate the models. As we are working on fake news detection in English language only, the dataset also consists of data in the form of full English news articles only. It has the title, text, type of news and date of publication features rather than posts or tweets. It mostly covers topics on politics and world news. This dataset was chosen as it consists of approximately 21,000 true articles and 23,000 fake articles and is thus a large dataset and also a balanced one as there is not a very large difference between the real and fake counts. The true articles are collected from Reuters.com (highly credible international news agency) which follows strict editorial fact-checking. The fake articles are collected from unreliable websites flagged as unreliable by PolitiFact (a fact-checking organization in the USA). The distribution of the topics of content in the dataset is given in the Table I.

Limitations of this dataset is that this dataset has source-based labelling as all articles that are taken from Reuters.com are taken as real and others as fake which might induce bias. It does not contain social media type content and thus would require more training to understand more articles of those types. There is no granular fact checking for articles considered fake taken from unreliable websites that may be true.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, January 2026
Impact Factor: 7.67

IV. METHODOLOGY

In this work, we explore and compare classic ML baseline models, deep learning based and transformer architecture-based models. Different preprocessing of data is performed based on the requirements of the models but are evaluated against the same metrics.

A. ML baseline models

For all baseline models, the same steps were performed for data pre-processing. The dataset was available as two different subsets of fake and real articles. A new feature was added to assign the label 'FAKE' to the fake articles and 'REAL' to the true ones. During inspection, we observed that almost all the real articles had a Reuters tag associated with them. This posed a risk of spurious correlations, where the model might classify news as really simply by detecting the source tag instead of learning deeper semantic patterns. To mitigate this, we pre-processed the dataset to remove these tags for these baseline models. Then both subsets were combined and the articles were shuffled to avoid ordering biasing. Null values, duplicate articles were checked and removed. Title and text features were combined under a single feature. Label encoding to assign fake articles to 0 and real to 1. Applied TF-IDF Vectorization to convert textual data into numerical feature vectors. Configured parameters such as maximum features and n-grams. The dataset was split into training and test sets with an 80/20 ratio for train to test dataset. Stratification was applied on labels to maintain equal class distribution across splits.

Logistic Regression: Logistic regression [1] is a linear statistical model used for binary or multi-class classification. It estimates the probability of a sample belonging to a class using the logistic (sigmoid) function. Despite its simplicity, it performs well on linearly separable text classification tasks as depicted in Fig. 1 and we obtained an accuracy of 99.1%.

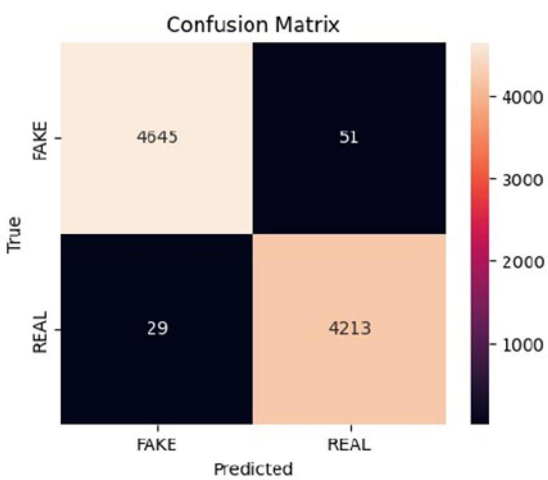

Fig. 1. Confusion Matrix for Logistic Regression

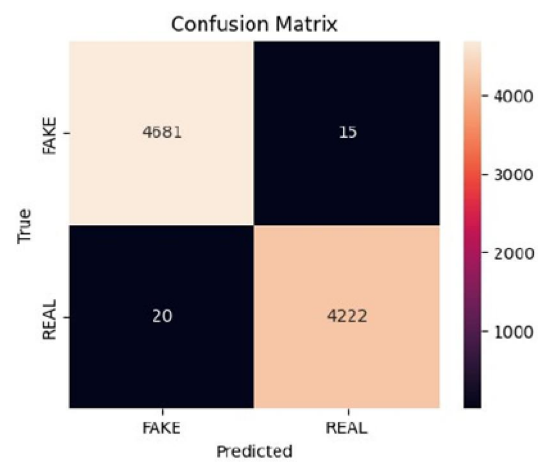

Fig. 2. Confusion Matrix for Decision Tree

Decision Tree classifier: A non-parametric model that splits the feature space recursively into regions based on decision rules, forming a tree-like structure [4]. It is interpretable but prone to overfitting when used alone. Observed a prediction accuracy of 99.6% with confusion matrix depicted in Fig. 2.

Support Vector Machines (SVM): A discriminative classifier that finds the hyperplane which maximizes the margin between different classes [5]. In text classification, Linear SVMs are particularly effective due to high-dimensional sparse features like TF-IDF. Predictions shown in Fig. 3 with an observed accuracy of 99.6%.

Naive Bayes classifier: A probabilistic classifier based on Bayes' Theorem with the assumption of feature independence [3]. The Multinomial Naive Bayes variant is particularly effective for text classification with word frequency features. Observed an accuracy of 95.6% as given in Fig. 4.

DOI: 10.48175/IJARSCT-31032

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, January 2026
Impact Factor: 7.67

Random Forest Classifier: A learning method that builds multiple decision trees during training and outputs the majority vote (classification) or mean prediction (regression) [2]. It reduces overfitting and improves generalization as shown in Fig. 5.

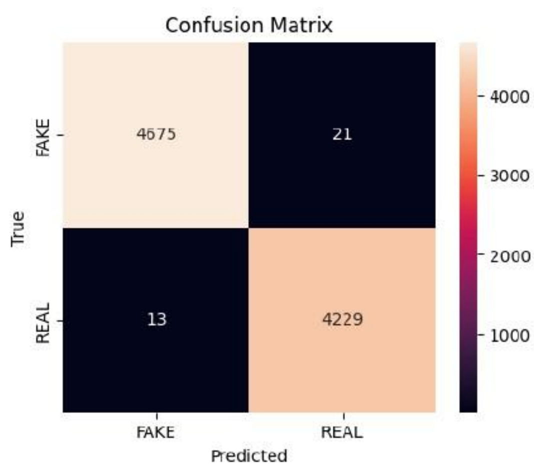

Fig. 3. Confusion Matrix for SVM

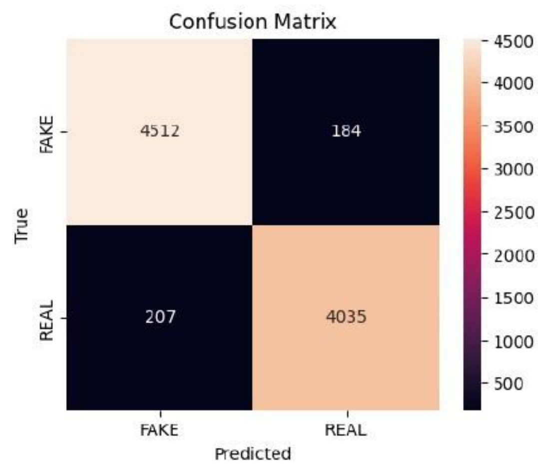

Fig. 4. Confusion Matrix for Naive Bayes

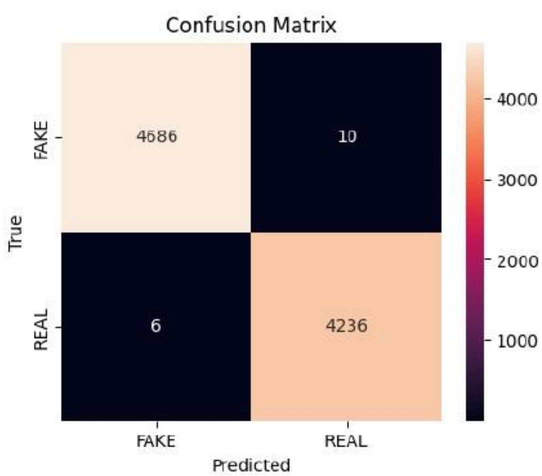

Fig. 5. Confusion Matrix for Random Forest

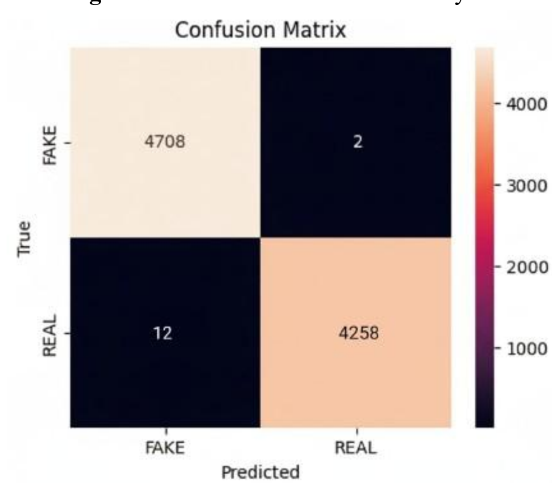

Fig. 6. Confusion Matrix for Bi-LSTM

B. Deep learning based approaches

For deep learning based approaches we make use of LSTM(Long Short-Term Memory) [32] model that takes the concatenated title and text and predict if the news is true or fake. Word2Vec embeddings [33] along with the LSTM model is used to implement uni-directional and bidirectional LSTM models to evaluate the accuracy of prediction [34] [35].

In this approach the same data preprocessing steps were followed for both the unidirectional and bidirectional LSTM. The true and fake subsets were firstly labeled, with 0 for fake and 1 for true and were then combined. Articles were shuffled to remove order biasing, missing values were dropped and the article text was cleaned by removing punctuations, links etc. Keras tokenizer is then used on the clean corpus with a vocabulary cap and text is converted to integer sequences and padded/truncated to a fixed length so batches have uniform length for LSTM processing. Word2Vec captures distributional semantics so similar words have similar vectors. An embedding matrix is then produced that maps the tokenizer indices to Word2Vec vectors. In both cases, the dataset is then split into training and test dataset in the ratio 80:20. The models are compiled with the Adam optimizer, binary cross-entropy loss, and accuracy as the evaluation metric. The models are trained for 10 epochs with a reasonable batch size of 32. Validation is performed against the held-out test set to monitor generalization.

Uni-LSTM: The uni-LSTM model is made of the Sequential layer that consists of layers shown in Table II.

TABLE II: MODEL SUMMARY: UNIDIRECTIONAL LSTM CLASSIFIER

Layer (type)	Output Shape	Param #
Embedding (Word2Vec)	(None, 500, 100)	13,536,200
LSTM (128 units)	(None, 128)	117,248
Dropout (0.2)	(None, 128)	0
Dense (64, ReLU)	(None, 64)	8,256
Dropout (0.2)	(None, 64)	0
Dense (1, Sigmoid)	(None, 1)	65
Total Parameters	13,661,769	
Trainable Params	125,569	
Non-trainable	13,536,200	

Bi-LSTM: The Bi-LSTM model also makes use of the Sequential layer that consists of the layers shown in Table III. The predictions of this model are shown in Fig. 6.

The BiLSTM extends the unidirectional LSTM by reading text in both forward and backward directions, enabling it to capture richer contextual dependencies in sentences. This often improves classification accuracy for tasks like fake news detection, where meaning can depend on both past and future words in a sequence. The Bi-LSTM model achieved nearperfect accuracy though the divergence between training and validation loss curves indicates mild overfitting suggesting the model may be memorizing patterns beyond what is necessary for generalization.

TABLE III: MODEL SUMMARY: BIDIRECTIONAL LSTM CLASSIFIER

Layer (type)	Output Shape	Param #
Embedding (Word2Vec)	(None, 500, 100)	13,536,200
Bi-LSTM (128 units)	(None, 256)	234,496
Dropout (0.2)	(None, 256)	0
Dense (64, ReLU)	(None, 64)	16,448
Dropout (0.2)	(None, 64)	0
Dense (1, Sigmoid)	(None, 1)	65
Total Parameters	14,289,229	
Trainable Params	251,009	
Non-trainable	13,536,200	

C. Transformer models

Transformer models are a class of deep learning architectures introduced by Vaswani et al. in the paper “Attention is All You Need” [36]. Unlike recurrent neural networks (RNNs), transformers dispense with recurrence and instead rely entirely on a self-attention mechanism, which allows the model to capture long-range dependencies in text more efficiently. Transformers have since become the foundation of most modern NLP models.

As in the earlier approaches, the fake and real subsets were combined, labelled with numerical labels (fake = 0, real = 1) and shuffled. Duplicate articles were removed and dataset was divided into train, test and validation set in the ratio 70:20:10.

Google’s BERT: BERT (Bidirectional Encoder Representations from Transformers) was proposed by Devlin et al. [9]. It extends the transformer encoder by training bidirectionally, meaning it learns contextual representations of words by

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 4, January 2026 Impact Factor: 7.67

jointly attending to tokens both to the left and right of a target token. Pre-trained on large corpora using two unsupervised tasks: Masked Language Modelling (MLM), where random words are masked and predicted, and Next Sentence Prediction (NSP), where the model learns sentence-level coherence.

The 'bert-base-uncased' pre-trained BERT tokenizer is loaded and used to tokenize and encode the datasets. ParameterEfficient Fine-Tuning (PEFT) using LoRA is applied to the model to reduce the number of trainable parameters. Training arguments are defined as in the Table IV.

The tokenized sequences are dynamically padded and the trainer is initialized with the model, arguments, datasets, and tokenizer and the model is trained on the training data. Predictions are made on the test set and a classification report is generated, showing precision, recall, and F1-score for both classes. A confusion matrix Fig. 7 is plotted to visualize the performance of the model in classifying fake and real news.

TABLE IV: BERT TRAINING PARAMETERS

Parameter	**Value**
Number of Epochs	5
Learning Rate	2e-5
Batch Size	8
Weight Decay	0.01

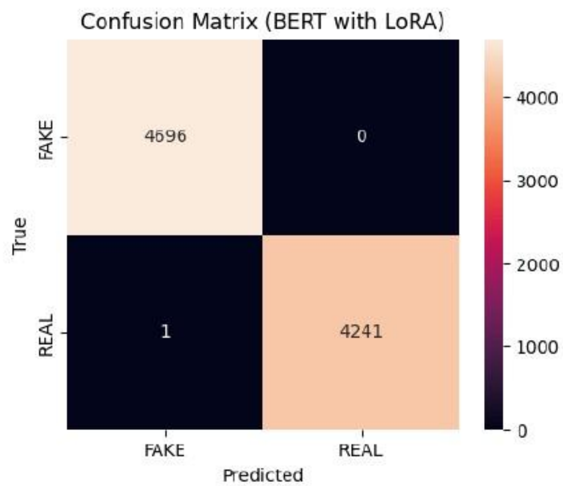

Fig. 7. Confusion Matrix for BERT with LoRA

Meta's RoBERTa: RoBERTa (Robustly Optimized BERT pretraining Approach) was introduced by Liu et al. [10] as an improved variant of BERT. RoBERTa modifies the pretraining process by (i) training with much larger datasets (ii) removing the NSP objective, (iii) using larger minibatches and longer sequences, and (iv) training for more steps. These optimizations allow RoBERTa to achieve better performance across many benchmarks compared to vanilla BERT, while retaining the same underlying transformer encoder architecture.

The model that was chosen was 'roberta-base', a pretrained language model. The text data was tokenized using the roberta-base tokenizer, with padding and truncation applied to ensure uniform input length of 256 and was formatted as PyTorch tensors. The model was initialized with two output labels for fake and real. The model was trained using the following training arguments as shown in Table V. The model was then evaluated on evaluation metrics such as accuracy, precision, recall, and F1-score and it's confusion matrix is just the same as that of BERT in 7.

TABLE V: ROBERTA TRAINING HYPERPARAMETERS

Parameter	**Value**
Number of Epochs	3
Learning Rate	2e-5

Batch Size	4
Weight Decay	0.01
Mixed Precision (FP16)	Enabled

V. EVALUATION METRICS

To assess model performance, we use four standard metrics derived from the confusion matrix: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

Accuracy measures the overall correctness:

$$Accuracy = \frac{TP + TN}{TP + TN + FP +} \quad (1)$$

Precision indicates how many predicted positive samples are actually positive:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall measures the proportion of actual positives correctly identified:

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F1-score is the harmonic mean of Precision and Recall:

$$F1\text{-}score = 2 \times \frac{Precision \times R}{Precision + R} \quad (4)$$

VI. RESULTS

Table VI summarizes the accuracies obtained by each model in all the different approaches with the highest highlighted in bold in each category**.**

TABLE VI: MODEL ACCURACY COMPARISON

Approaches	**Models**	**Accuracy (%)**
ML based models	Logistic Regression Decision Tree Linear SVC Naive Bayes Random Forest	99.10 99.60 99.61 95.62 99.82
DL based approaches	Uni-LSTM Bi-LSTM	99.93 99.84
Transformer models	BERT RoBERTa	99.98 99.98

VII. CONCLUSION

In this paper, we conducted a comparative study on fake news detection using three categories of approaches: classical machine learning baselines, deep learning (Uni-LSTM and Bi-LSTM), and modern transformer-based architectures (BERT and RoBERTa). The results demonstrate that while baseline models such as Logistic Regression and Random Forest achieve competitive accuracy, deep learning models further enhance performance by capturing sequential dependencies in text. However, the transformer-based models, particularly BERT and RoBERTa, outperform both categories and achieve the highest accuracy, showcasing the effectiveness of pretrained contextual embeddings in handling nuanced and complex news content. Our findings highlight the evolution of fake news detection techniques and reaffirm the superiority of transformer models in natural language processing tasks.

LIMITATIONS AND FUTURE WORK

Despite the high performance observed, there are several limitations to this content-based classification of fake news.

Dataset dependency: The ISOT dataset, while widely used, may not fully represent the dynamic and evolving nature of fake news across diverse platforms (e.g., Twitter, Reddit, or multimedia sources). This may limit real-world generalizability.

Overfitting risk: High accuracies might suggest potential overfitting to the dataset. More rigorous cross-dataset evaluation is required.

Contextual limitations: The models primarily focus on textual content and do not incorporate external signals such as user credibility, propagation patterns as in Monti et al. [30], or multimodal data(images/videos) like Bansal et al. [37].

For future work, expanding to multi-source and multilingual datasets, exploring multimodal fake news detection, and integrating social context and graph-based models can make detection systems more robust. Additionally, deploying lighter transformer variants (e.g., DistilBERT, TinyBERT) may enable real-time fake news detection on resource-constrained environments such as mobile and edge devices. Finally, future work could explore explainable AI (XAI) methods [37] to increase interpretability, helping end-users and policymakers understand why a particular article is flagged as fake.

ACKNOWLEDGMENT

We thank our mentor Prof. A. Chandorkar, from the Department of Computer Engineering, PICT for her guidance and help that she has provided us with throughout the writing of this research paper. She has continuously motivated us throughout this semester project and thus was of great help.

REFERENCES

[1]. D. R. Cox (1958, Dec), "The Regression Analysis of Binary Sequences", Journal of the Royal Statistical Society: Series B (Methodological), Wiley, UK, Available: https://doi.org/10.1111/j.2517-6161.1958.tb00292.x

[2]. L. Breiman (2001, Oct), "Random Forests", Machine Learning, vol. 45, no. 1, Springer, USA, Available: https://doi.org/10.1023/A:1010933404324

[3]. T. M. Mitchell (1997), "Machine Learning", McGraw Hill, USA, ISBN: 978-0070428072

[4]. J. R. Quinlan (1986, Mar), "Induction of Decision Trees", Machine Learning, vol. 1, pp. 81–106, Springer, USA, Available: https://doi.org/10.1007/BF 00116251

[5]. C. Cortes, V. Vapnik (1995, Sep), "Support-Vector Networks", Machine Learning, vol. 20, pp. 273–297, Springer, USA, Available: https://doi.org/10.1 007/BF00994018

[6]. Ahmed H, Traore I, Saad S., (2018), "Detecting opinion spams and fake news using text classification", Journal of Security and Privacy, Volume 1, Issue 1, Wiley.

[7]. Ahmed H, Traore I, Saad S., (2017) "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques. In: Traore I., Woungang I., Awad A. (eds) Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments. ISDDC 2017. Lecture Notes in Computer Science, vol 10618. Springer, Cham (pp. 127- 138).

[8]. S. Raza, D. Patterson, C. Ding, (2024, Dec), "Fake News Detection: Comparative Evaluation of BERT-like Models and Large Language Models with Generative AI-Annotated Data", arXiv preprint, Available: https://arxiv.org/abs/2412.14276

[9]. J. Devlin, M. Chang, K. Lee, K. Toutanova, (2019, May), "BERT: Pretraining of Deep Bidirectional Transformers for Language Understanding", arXiv preprint, Available: https://arxiv.org/abs/1810.04805

[10]. Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, (2019, Jul), "RoBERTa: A Robustly Optimized BERT Pretraining Approach", arXiv preprint, Available: https://arxiv.org/abs/1907.11692

[11]. F. Monti, F. Frasca, D. Eynard, D. Mannion, M. Bronstein, (2019, Feb), "Fake news detection on social media using geometric deep learning", arXiv preprint, Available: https://arxiv.org/abs/1902.06673

[12]. A Sadick, B. Smith, F. Nasirzadeh; S. Sadeghi, S. Ayal, M. Bouadjenek.(2025) "Comparative Analysis of BERT and RoBERTa for Construction Site Incident Report Classification," CIB Conferences: Vol. 1 Article 248. DOI: https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1839&context=cibco nferences

[13]. E. Tacchini, G. Ballarin, M. Vedova, S. Moret, L. Alfaro, (2017, Apr), "Some Like it Hoax: Automated Fake News Detection in Social Networks", arXiv preprint, Available: https://arxiv.org/abs/1704.07506

[14]. W. Wang, (2017, May), "Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection", arXiv preprint, Available: https://arxiv.org/ab s/1705.00648

[15]. R. Kaliyar, A. Goswami, P. Narang, S. Sinha, (2020, Jun), "FNDNet – A deep convolutional neural network for fake news detection", Vol. 61, pg.32-44, Cognitive Systems Research, Available: https://www.sciencedirect.com/science/article/pii/S1389041720300085

[16]. W. Chen, Y. Zhang, C. Yeo, C. Liau, B. Lee, (2018, Apr), "Unsupervised rumor detection based on users' behaviors using neural networks", Vol. 105, pg. 226-233, Pattern Recognition Letters, Available: https://www.sciencedirect.com/science/article/pii/S0167865517303823

[17]. M. Farokhian, V. Rafe, H. Veisi, (2023, Oct), "Fake news detection using dual BERT deep neural networks", Multimed Tools Appl 83, 43831–43848. Available: https://doi.org/10.1007/s11042-023-17115-w

[18]. R. Kaliyar, A. Goswami, P. Narang, (2021, Jan), "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach", Multimed Tools Appl 80, 11765–11788. Available: https://doi.org/10.1007/s11042-020-10183-2

[19]. H. Alcott, M. Gentzkow, (2017, May), "Social Media and Fake News in the 2016 Election", Vol. 31, pg. 211–236, Journal of Economic Perspectives, Available: https://www.aeaweb.org/articles?id=10.1257/jep.31.2.211

[20]. S. Murugan, R. Arthi, E. Boomika, S. Jeyanth, S. Kaviyarasu, (2024), "Challenges and Insights in Identifying Hate Speech and Fake News on Social Media", pg. 16-21, Proceedings of the 21st International Conference on Natural Language Processing (ICON): Shared Task on Decoding Fake Narratives in Spreading Hateful Stories (Faux-Hate), ACL anthology, Available: https://aclanthology.org/2024.icon-fauxhate.4.pdf

[21]. M. Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H. Stanley, W. Quattrociocchi, (2019), "The spreading of misinformation online", Vol. 113, No. 3, pg. 554-559, Proceedings of the national academy of Sciences, Available: https://www.pnas.org/doi/pdf/10.1073/pnas.1517441113

[22]. A Ahmed, A. Aljabouh, P. Donepudi, M. Choi, (2021, Feb), "Detecting Fake News Using Machine Learning : A Systematic Literature Review", arXiv preprint, Available: https://arxiv.org/abs/2102.04458

[23]. L. Huang, (2022), "Deep learning for fake news detection: Theories and models", pg. 1322-1326, Proceedings of the 2022 6th International Conference on Electronic Information Technology and Computer Engineering, Available: https://dl.acm.org/doi/pdf/10.1145/3573428.3573663

[24]. H. Jwa, D. Oh, K. Park, J. Kang, H. Lim, (2019), "exBAKE: Automatic Fake News Detection Model Based on Bidirectional Encoder Representations from Transformers (BERT)", Vol. 9, No. 19, pg. 4062, Applied Sciences, Available: https://doi.org/10.3390/app9194062

[25]. A Drif, Z. Hamida, S. Giordano, (2019), "Fake news detection method based on text-features." France, International Academy, Research, and Industry Association (IARIA) : 27-32.

[26]. J. Zhang, B. Dong, P. Yu (2020), "FakeDetector: Effective Fake News Detection with Deep Diffusive Neural Network", pg. 1826-1829, IEEE 36th International Conference on Data Engineering (ICDE) Available: https://ieeexplore.ieee.org/document/9101457/

[27]. B. Alqadi, S. Alsuhibany, S. Yousafzai, S. Alzu'bi, D. Alsekait, D. AbdElminaam, (2025), "Transfer learning driven fake news detection and classification using large language models", Vol. 15, No. 1, pg. 28490, Scientific Reports, Available: https://www.nature.com/articles/s41598-025-10670-2

[28]. L. Hu, T. Yang, L. Zhang, W. Zhong, D. Tang, C. Shi , N. Duan, M. Zhou, (2021, Aug), "Compare to The Knowledge: Graph Neural Fake News Detection with External Knowledge", Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on

Natural Language Processing (Volume 1: Long Papers), ACL, Available: https://aclanthology.org/2021.acl-long.62.pdf

[29]. K. Shu, D. Mahudeswaran, S. Wang, D. Lee, H. Liu, (2019, Mar), "FakeNewsNet: A Data Repository with News Content, Social Context and Spatialtemporal Information for Studying Fake News on Social Media", arXiv preprint, Available: https://arxiv.org/abs/1809.01286

[30]. F. Monti, F. Frasca, D. Eynard, D. Mannion, M. Bronstein, (2019, Feb), "Fake News Detection on Social Media using Geometric Deep Learning" , arXiv

[31]. preprint, Available: https://arxiv.org/abs/1902.06673PolitiFact Available: https://www.politifact.com/

[32]. S. Hochreiter , J. Schmidhuber, (1997), "Long Short-Term Memory", Vol. 9, No. 8, pg. 1735-1780, Neural Computation, Available: https://doi.org/10.1 162/neco.1997.9.8.1735

[33]. T. Mikolov, K. Chen, G. Corrado, J. Dean, (2013, Sep), "Efficient Estimation of Word Representations in Vector Space", arXiv preprint, Available: https://arxiv.org/abs/1301.3781

[34]. J. Wang, T. Liu, X. Luo, L. Wang, (2018), "An LSTM Approach to Short Text Sentiment Classification,", pg. 214-223, Proceedings of the 30th conference on computational linguistics and speech processing (ROCLING 2018), Available: https://aclanthology.org/O18-1021.pdf

[35]. H. Padalko, V. Chomko, D. Chumachenko (2024), 'A novel approach to fake news classification using LSTMbased deep learning models", Vol. 6,

[36]. Frontiers in big Data, Available: https://www.frontiersin.org/journals/bigdata/articles/10.3389/fdata.2023.1320800/full

[37]. A Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, Ł. Kaiser, and I. Polosukhin, (2017, Jun), "Attention is All You Need", Advances in Neural Information Processing Systems (NeurIPS), Available: https://arxiv.org/abs/1706.03762

[38]. M Ngueajio, S. Aryal, M. Atemkeng, G. Washington, D. Rawat, (2025), "Decoding Fake News and Hate Speech: A Survey of Explainable AI Techniques", Vol. 57, No. 1, pg. 1-37, ACM Computing Surveys, Available: https://dl.acm.org/doi/full/10.1145/3711123