COVID-19 Fake News Detection : A Comparative Study of NLP Models (LSTM, GRU, RNN, and BERT)

Jean Robin Raherisambatra
African Master's in Machine Intelligence (AMMI)
AIMS Senegal
jrraherisambatra@aimsammi.org

Supervised by

Dr. James Bentham
Newcastle University
James.Bentham@newcastle.ac.uk

July 23, 2024

Abstract

In this work, we explore the application of various Natural Language Processing (NLP) models for detecting fake news related to COVID-19. Using a dataset of COVID-19-related news articles, we implemented and compared the performance of Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRUs), and Transformers. Our results demonstrate that advanced models such as Transformers outperform traditional RNN-based approaches in terms of accuracy and robustness. The BERT model achieved an impressive test accuracy of 95.61%, the LSTM model achieved a test accuracy of 91.79%, the GRU model achieved a test accuracy of 91.65%, and the RNN model achieved a test accuracy of 87.19%. This highlights the effectiveness of Transformers, particularly BERT, in distinguishing between fake and real news in the context of COVID-19.

1 Introduction

The proliferation of fake news, especially in the health domain, poses significant risks to public health and safety. Misinformation can lead to harmful behaviors, undermine public health efforts, and cause widespread panic. Automated detection of fake news using Natural Language Processing (NLP) techniques is a crucial step towards mitigating this issue. This paper investigates the effectiveness of various deep learning models for the binary classification of COVID-19-related news articles into fake or real categories. We compare traditional Recurrent Neural Networks (RNNs) and their advanced variants, such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs), with the state-of-the-art Transformer model, BERT. Our study aims to highlight the strengths and limitations of each model in handling the complexity and nuances of COVID-19-related misinformation.

2 Related Work

Prior research on fake news and NLP detection has predominantly focused on general news articles [1, 2, 5, 15, 23], leveraging models like RNNs, LSTMs, GRUs, and Transformers. Recent advancements in Transformers, particularly the BERT model, have shown promising results in various NLP tasks. However, there is limited research specifically targeting fake news detection in the health domain [11, 13, 19].

There is also some work on COVID-19 specifically [3, 8, 9].

3 Dataset Description

We utilized the dataset described in `https://link.springer.com/chapter/10.1007/978-3-030-73696-5_3` : "Fighting an Infodemic: COVID-19 Fake News Dataset" [16, 17]. This dataset contains a collection of COVID-19-related articles labeled as fake or real. The dataset includes various features such as the article text, providing a rich resource for training and evaluating NLP models.

4 Methodology

4.1 Data Preprocessing

We performed several preprocessing steps on the dataset:

- Lowercasing the article texts.
- Removing URLs and special characters.
- Modifying stop words by adding unnecessary elements.

- Removing stop words from the article texts.
- Lemmatize the words in the article texts.

4.2 Model Architectures

We implemented four types of models for this study (look at Figure 1):

- **RNN**: A basic recurrent neural network that processes the text data sequentially, which is known for its simplicity but can struggle with long-term dependencies due to issues like vanishing gradients[14, 24].
- **LSTM**: An advanced RNN variant designed to capture long-term dependencies more effectively by using memory cells and gating mechanisms to control the flow of information [12, 22].
- **GRU**: A simplified version of LSTM that balances performance and computational efficiency by using fewer gates, making it faster to train while maintaining effectiveness in learning long-term dependencies [4, 20].
- **Transformer (BERT)**: A state-of-the-art model utilizing self-attention mechanisms to handle long-range dependencies and parallel processing, which enables it to capture contextual information from both directions of a text sequence simultaneously [6].

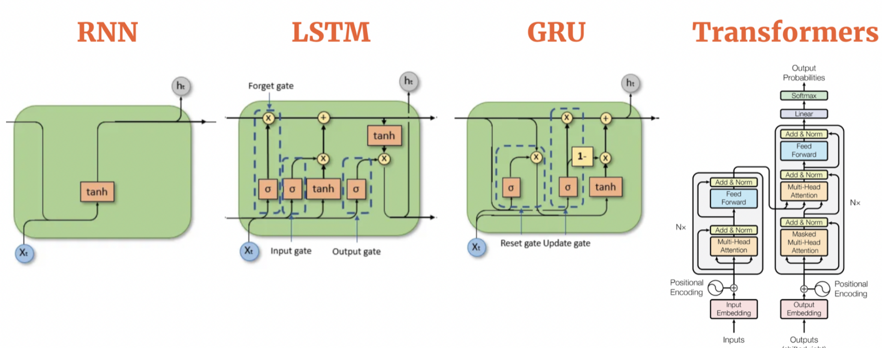

Figure 1: Comparing different sequences models : RNN, LSTM, GRU and Transformers

Each model was fine-tuned and optimized for the task of fake news detection.

4.3 Model Enhancements

Enhancements were added to the LSTM, GRU, and RNN models to improve their performance in detecting fake news:

- **Bidirectional Layers:** Provide context from both past and future sequences, enhancing understanding of sentence context and improving model performance in NLP tasks.
- **Batch Normalization:** Stabilizes training by normalizing layer inputs, accelerating convergence and improving generalization in deep learning models.
- **Dropout Layers:** Prevent overfitting by randomly dropping input units during training, improving model robustness and performance on unseen data.
- **Dense Layers:** Follow recurrent layers to extract complex features and enhance model capacity for learning intricate patterns in data.
- **L2 Regularization:** Adds a penalty to the loss function, encouraging smaller weights and reducing overfitting, thereby improving model generalization.

4.4 Handling Class Imbalance

The dataset is imbalanced, with a higher number of real news articles (5600) compared to fake ones (5100). To address this, we used the Synthetic Minority Over-sampling Technique (SMOTE) to oversample the minority class [18]. Additionally, for the RNN, LSTM, and GRU models, class weights were applied to handle the class imbalance during training [7, 10, 21].

4.5 Training Process

We split the dataset into training and test sets. The models were trained for 20 epochs using binary cross-entropy loss and optimized using Adam optimizer. Early stopping was employed to prevent overfitting.

5 Results and Discussion

5.1 Visualization with Word Clouds

In an effort to gain deeper insights into the data, we used word clouds to visualize the most frequent words in both cleaned real and fake news articles. Figure 2,Figure 3 shows the word clouds for fake and real news, respectively.

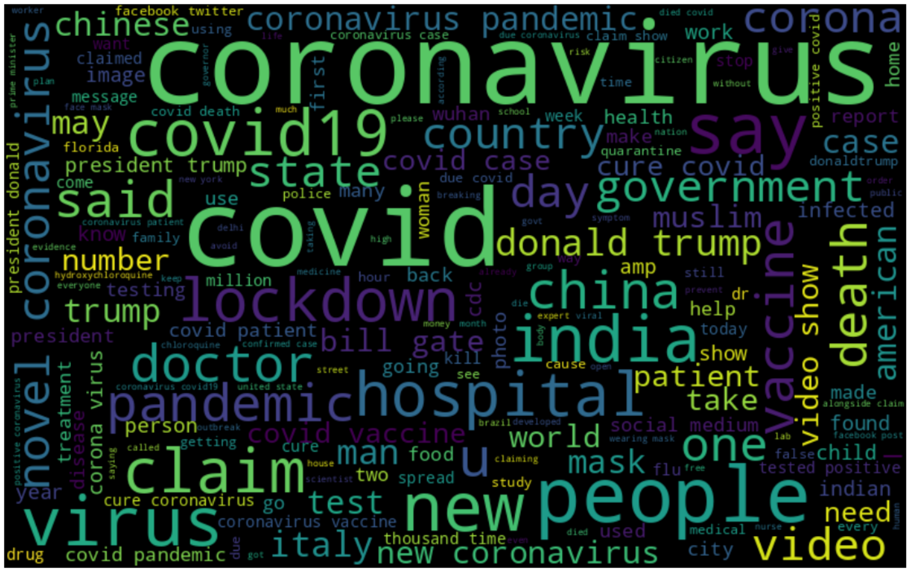

Figure 2: Word Cloud - Fake News articles

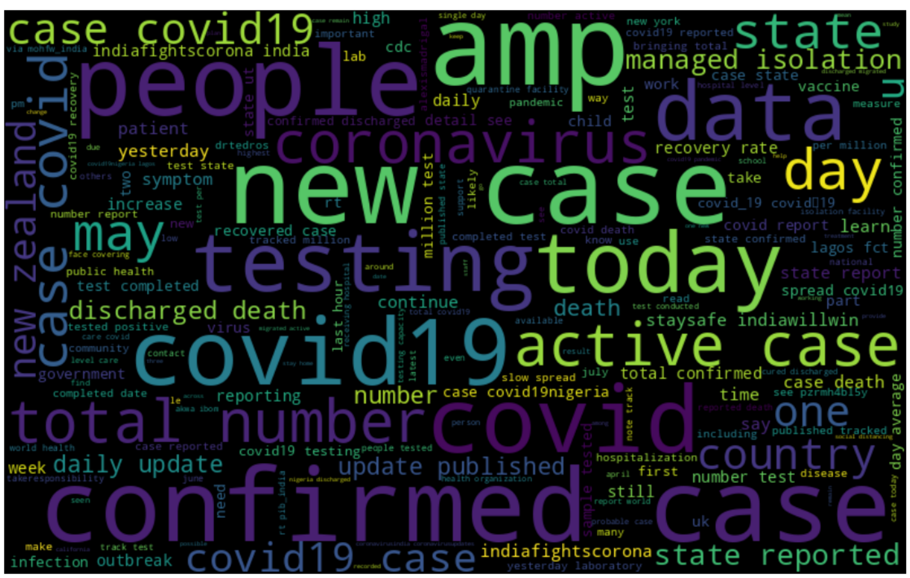

Figure 3: Word Cloud - Real News articles

The word clouds reveal common terms and phrases in both real and fake news tweets, providing valuable insights into the nature of the content.

5.2 Accuracy and Loss Curves

The training and validation accuracy and loss curves for each model are presented in Figures 1 to 4. These curves illustrate the learning dynamics and potential overfitting or underfitting issues.

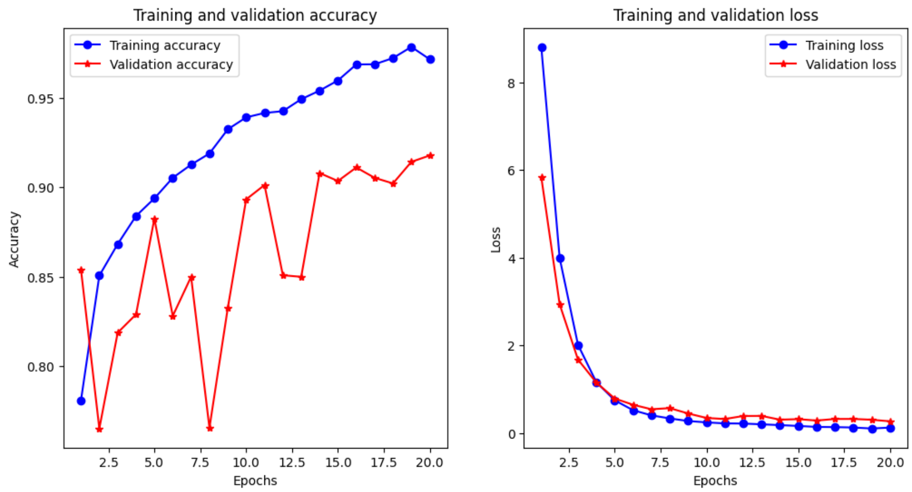

Figure 4: Training and Validation Accuracy/Loss Curves for LSTM Model

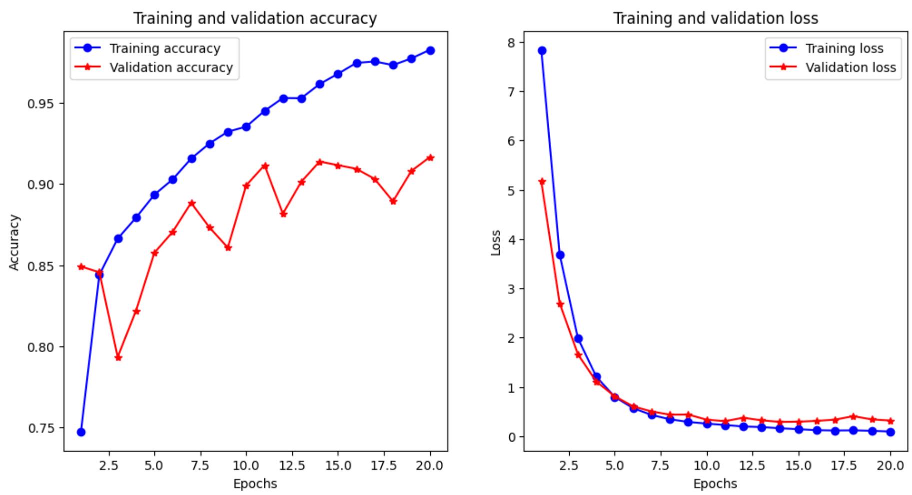

Figure 5: Training and Validation Accuracy/Loss Curves for GRU Model

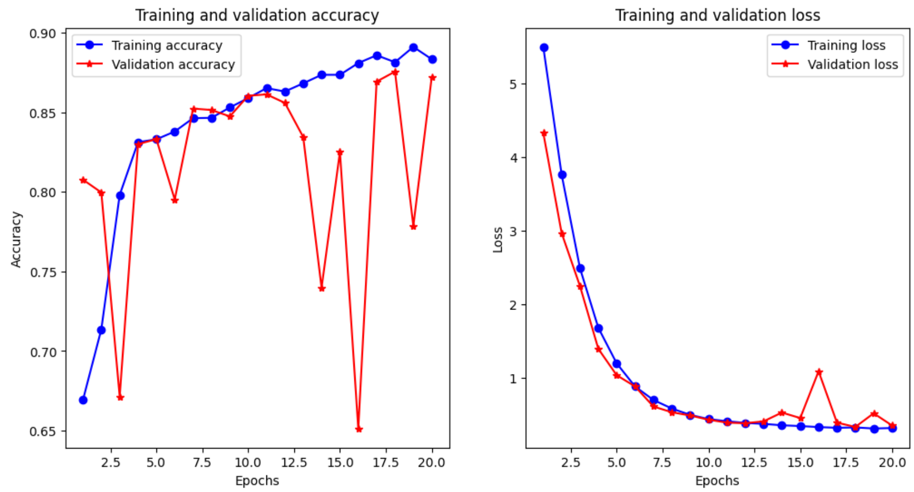

Figure 6: Training and Validation Accuracy/Loss Curves for RNN Model

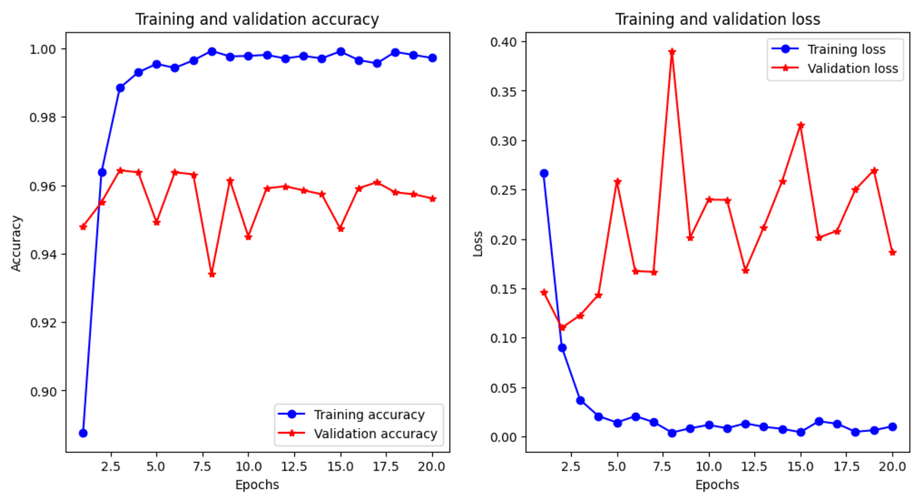

Figure 7: Training and Validation Accuracy/Loss Curves for BERT Model

5.3 Evaluation Metrics

We used the following metrics to evaluate model performance:

- **Accuracy**: The ratio of correctly predicted instances to the total instances.
- **Precision**: The ratio of true positive predictions to the sum of true positive and false positive predictions.

- **Recall**: The ratio of true positive predictions to the sum of true positive and false negative predictions.
- **F1-Score**: The harmonic mean of precision and recall.

5.4 Model Performance

The evaluation metrics for the models are summarized in Table 1.

- For the LSTM model, we achieved a test accuracy of 91.79%, with the precision, recall, and F1-score for class 0 (fake) being 0.90, 0.94, and 0.92 , respectively, and for class 1 (real) being 0.93, 0.90, and 0.92, respectively.
- For the RNN model, we achieved a test accuracy of 87.19%, with precision, recall, and F1-score for class 0 (fake) being 0.90, 0.84, and 0.87 , respectively, and for class 1 (real) being 0.85, 0.90, and 0.88, respectively.
- For the GRU model, we achieved a test accuracy of 91.65%, with precision, recall, and F1-score for class 0 (fake) being 0.90, 0.94, and 0.92, respectively, and for class 1 (real) being 0.94, 0.89, and 0.91, respectively.
- For the BERT model, we achieved a test accuracy of 95.61%, with precision, recall, and F1-score for class 0 (fake) being 0.95, 0.96, and 0.95, respectively, and for class 1 (real) being 0.97, 0.95, and 0.96, respectively.

Model	**Accuracy (%)**	**Precision**	**Recall**	**F1-Score**
		fake | real	fake | real	fake | real
RNN	87.19	0.90 | 0.85	0.84 | 0.90	0.87 | 0.88
LSTM	91.79	0.90 | 0.93	0.94 | 0.90	0.92 | 0.92
GRU	91.65	0.90 | 0.94	0.94 | 0.89	0.92 | 0.91
BERT	95.61	0.95 | 0.97	0.96 | 0.95	0.95 | 0.96

Table 1: Model Performance Comparison

5.5 Interpretation of Results

The accuracy and loss curves provide valuable insights into the performance of each model. The RNN initially exhibits rapid accuracy improvement during training but flattens out quickly, indicating potential overfitting due to its limited ability to capture long-term dependencies effectively. In contrast, both the LSTM and GRU models show smoother accuracy curves, suggesting better generalization capabilities and an improved ability to capture long-term dependencies. The GRU model slightly outperforms the LSTM, indicating it can effectively handle essential dependencies and patterns in the data. However, the BERT model stands out with its self-attention mechanism, demonstrating consistent accuracy improvement and lower validation loss over time. This highlights BERT's superior ability to handle complex dependencies and relationships within the data, resulting in higher accuracy and better generalization to unseen data. Overall, while traditional RNNs may struggle with intricate patterns, advanced models like LSTM, GRU, and especially BERT, offer significant improvements in capturing intricate patterns and ensuring robust performance.

5.6 Comparative Analysis

Our comparative analysis reveals significant performance differences among the models. The BERT model outperforms all others, achieving an impressive test accuracy of 95.61%. This model demonstrates exceptional precision, recall, and F1-scores for both fake and real news classes, with scores of 0.95, 0.96, and 0.95 for class 0 (fake) and 0.97, 0.95, and 0.96 for class 1 (real), respectively. These results underscore BERT's robustness and effectiveness in distinguishing between the two categories.

The LSTM model also performs well, with a test accuracy of 91.79%. It achieves balanced precision, recall, and F1-scores of 0.90, 0.94, and 0.92 for class 0 (fake) and 0.93, 0.90, and 0.92 for class 1 (real). This makes it a reliable choice for fake news detection.

The GRU model achieves a test accuracy of 91.65%. It shows balanced performance with precision, recall, and F1-scores of 0.90, 0.94, and 0.92 for class 0 (fake) and 0.94, 0.89, and 0.91 for class 1 (real). This demonstrates the model's effectiveness and reliability, comparable to the LSTM model.

The RNN model achieves a test accuracy of 87.19%. It shows balanced performance with precision, recall, and F1-scores of 0.90, 0.84, and 0.87 for class 0 (fake) and 0.85, 0.90, and 0.88 for class 1 (real). However, it demonstrates limitations in handling the dataset's complexity compared to more advanced models.

5.7 Discussion on Model Performance

The superior performance of the BERT model can be attributed to its advanced architecture, particularly its ability to capture long-range dependencies and contextual information through self-attention. BERT's self-attention mechanism is crucial for handling complex dependencies, leading to robust performance across various evaluation metrics.

RNN-based models, while effective to some extent, struggle with long-term dependencies and often converge slower. Enhancements like dense and dropout layers improve performance and generalization.

LSTM and GRU models balance performance and efficiency. Bidirectional layers capture both past and future context, batch normalization stabilizes and accelerates training, and L2 regularization reduces overfitting, ensuring generalizability. Despite these enhancements, LSTM and GRU models still face challenges in capturing nuanced data relationships, unlike BERT's architecture.

6 Conclusion

This study demonstrated the effectiveness of various NLP models for detecting fake news related to COVID-19. Our experiments showed that the BERT model significantly outperforms traditional RNN-based models. This finding underscores the importance of leveraging advanced deep learning architectures for robust fake news detection in health-related content. Future research will focus on enhancing model robustness and exploring additional features to improve fake news detection.

GitHub Repository

All code developed and used in this project is available in the GitHub repository: `https://github.com/Robinjean/Project_Covid19_Fake_news_detection_using_RNN_LSTM_GRU_transformers/tree/main`.

Acknowledgments

First and foremost, I thank Almighty God for granting me the gift of life to undertake my studies at AMMI/AIMS Senegal. I am grateful to AMMI for the opportunity to pursue my master's. Special thanks to my supervisor, James Bentham, for his valuable feedback, guidance, and support throughout this research. I also acknowledge the authors of the dataset "Fighting an Infodemic: COVID-19 Fake News Dataset" for providing this valuable resource for our research.

References

[1] Vasu Agarwal, H Parveen Sultana, Srijan Malhotra, and Amitrajit Sarkar. Analysis of classifiers for fake news detection. *Procedia Computer Science*, 165:377–383, 2019.

[2] J Antony Vijay, H Anwar Basha, and J Arun Nehru. A dynamic approach for detecting the fake news using random forest classifier and nlp. In *Computational Methods and Data Engineering: Proceedings of ICMDE 2020, Volume 2*, pages 331–341. Springer, 2020.

[3] Waqas Haider Bangyal, Rukhma Qasim, Najeeb Ur Rehman, Zeeshan Ahmad, Hafsa Dar, Laiqa Rukhsar, Zahra Aman, and Jamil Ahmad. Detection of fake news text classification on covid-19 using deep learning approaches.

Computational and mathematical methods in medicine, 2021(1):5514220, 2021.

[4] Trapit Bansal, David Belanger, and Andrew McCallum. Ask the gru: Multi-task learning for deep text recommendations. In *proceedings of the 10th ACM Conference on Recommender Systems*, pages 107–114, 2016.

[5] Ganesh Gopal Devarajan, Senthil Murugan Nagarajan, Sardar Irfanullah Amanullah, S. A. Sahaaya Arul Mary, and Ali Kashif Bashir. Ai-assisted deep nlp-based approach for prediction of fake news from social media users. *IEEE Transactions on Computational Social Systems*, pages 1–11, 2023.

[6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[7] Guo Haixiang, Li Yijing, Jennifer Shang, Gu Mingyun, Huang Yuanyue, and Gong Bing. Learning from class-imbalanced data: Review of methods and applications. *Expert systems with applications*, 73:220–239, 2017.

[8] Abdullah Hamid, Nasrullah Shiekh, Naina Said, Kashif Ahmad, Asma Gul, Laiq Hassan, and Ala Al-Fuqaha. Fake news detection in social media using graph neural networks and nlp techniques: A covid-19 use-case. *arXiv preprint arXiv:2012.07517*, 2020.

[9] Celestine Iwendi, Senthilkumar Mohan, Suleman khan, Ebuka Ibeke, Ali Ahmadian, and Tiziana Ciano. Covid-19 fake news sentiment analysis. *Computers and Electrical Engineering*, 101:107967, 2022.

[10] Bartosz Krawczyk. Learning from imbalanced data: open challenges and future directions. *Progress in artificial intelligence*, 5(4):221–232, 2016.

[11] Santoshi Kumari, Harshitha K Reddy, Chandan S Kulkarni, and Vanukuri Gowthami. Debunking health fake news with domain specific pre-trained model. *Global Transitions Proceedings*, 2(2):267–272, 2021.

[12] Stephen Merity, Nitish Shirish Keskar, and Richard Socher. Regularizing and optimizing lstm language models. *arXiv preprint arXiv:1708.02182*, 2017.

[13] Claudio Tinoco Mesquita, Anderson Oliveira, Flávio Luiz Seixas, and Aline Paes. Infodemia, fake news and medicine: science and the quest for truth, 2020.

[14] Tomáš Mikolov, Stefan Kombrink, Lukáš Burget, Jan Černocký, and Sanjeev Khudanpur. Extensions of recurrent neural network language model. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5528–5531, 2011.

[15] Ray Oshikawa, Jing Qian, and William Yang Wang. A survey on natural language processing for fake news detection. *arXiv preprint arXiv:1811.00770*, 2018.

[16] Parth Patwa, Shivam Sharma, Srinivas PYKL, Vineeth Guptha, Gitanjali Kumari, Md Shad Akhtar, Asif Ekbal, Amitava Das, and Tanmoy Chakraborty. Fighting an infodemic: Covid-19 fake news dataset, 2020.

[17] Parth Patwa, Shivam Sharma, Srinivas Pykl, Vineeth Guptha, Gitanjali Kumari, Md Shad Akhtar, Asif Ekbal, Amitava Das, and Tanmoy Chakraborty. Fighting an infodemic: Covid-19 fake news dataset. In Tanmoy Chakraborty, Kai Shu, H. Russell Bernard, Huan Liu, and Md Shad Akhtar, editors, *Combating Online Hostile Posts in Regional Languages during Emergency Situation*, pages 21–29, Cham, 2021. Springer International Publishing.

[18] Russel Pears, Jacqui Finlay, and Andy M Connor. Synthetic minority over-sampling technique (smote) for predicting software build outcomes. *arXiv preprint arXiv:1407.2330*, 2014.

[19] Cristina M Pulido, Laura Ruiz-Eugenio, Gisela Redondo-Sama, and Beatriz Villarejo-Carballido. A new application of social impact in social media for overcoming fake news in health. *International journal of environmental research and public health*, 17(7):2430, 2020.

[20] Apeksha Shewalkar, Deepika Nyavanandi, and Simone A Ludwig. Performance evaluation of deep neural networks applied to speech recognition: Rnn, lstm and gru. *Journal of Artificial Intelligence and Soft Computing Research*, 9(4):235–245, 2019.

[21] Vimalraj S Spelmen and R Porkodi. A review on handling imbalanced data. In *2018 international conference on current trends towards converging technologies (ICCTCT)*, pages 1–11. IEEE, 2018.

[22] Martin Sundermeyer, Ralf Schlüter, and Hermann Ney. Lstm neural networks for language modeling. In *Interspeech*, volume 2012, pages 194–197, 2012.

[23] Aswini Thota, Priyanka Tilak, Simrat Ahluwalia, and Nibrat Lohia. Fake news detection: a deep learning approach. *SMU Data Science Review*, 1(3):10, 2018.

[24] Wenpeng Yin, Katharina Kann, Mo Yu, and Hinrich Schütze. Comparative study of cnn and rnn for natural language processing. *arXiv preprint arXiv:1702.01923*, 2017.