Detecting Indian Right-Wing Propaganda On Twitter Using Multi-Disciplinary, Multi-Modal Approach

[1]Ameenur Rahman, [2]Dr.K Nagi Reddy, [3]Mohammer Fazal, [4]B Nagalakshmi

[1] Independent Researcher, [2] Professor, HOD & Dean Academics, [3] Independent Researcher, [4] Assistant Professor

[1]Department of Information Technology, [2]Department of Information Technology, [3]Department of Information Technology, [4]Department of Information Technology

[1]Lords Institute of Engineering and Technology, Hyderabad, India, [2]Lords Institute of Engineering and Technology, Hyderabad, India, [3]Lords Institute of Engineering and Technology, Hyderabad, India, [4]Lords Institute of Engineering and Technology, Hyderabad, India

Abstract: In the context of Indian democracy, Pervasive Right-wing digital Propaganda is used by political actors as a strategic tool to accomplish their political and ideological agendas, undermining the freedom of speech and belief, both promised and upheld as fundamental rights in our constitution. The primary objective of this paper is to understand the digital discourse of the right-wing in the context of Twitter. And, build an ANN and CNN-BiLSTM RNN to detect and predict the right-wing nature of the tweet. The model will be trained on a real-time 51000 dataset of various propaganda and non-propaganda Twitter handles. The ANN model yields an accuracy of 97% after hyperparameter tuning, and CNN-BiLSTM RNN yields an 87% AUC score after hyperparameter tuning. Analysing the empirical data drawn from the exploratory data analysis and insights drawn from the models, the paper will further highlight the trends of right-wing account usage and their digital propaganda, calling for a new approach to using metadata to understand right-wing tendencies

Index Terms - Indian Digital right-wing propaganda, real-time dataset, ANN, CNN-BiLSTM, metadata analysis.

I. INTRODUCTION

With the rise of digital devices and social media in Indian society after 2014, there has been an increase in concerns regarding political propaganda, as the World Economic Forum states disinformation and misinformation exacerbate the current communal polarization, and ranks India in the top 5 threatened by disinformation [41]. One of the purposes of social media algorithm is to perform "great clustering" and this creates many echo chambers and the information is readily accepted based on the trust on the source rather than the authenticity of the message [6], [2] itself thereby fostering and exacerbating the communal tensions and communal problems that currently exist in our society.

Further exacerbating the problem is a research gap that only focuses on the textual analysis of the tweet other than some papers that try to solve this problem with a different approach [20], [21], [22], [19]. Additionally, the discourse of right-wing propaganda on Twitter and social media still remains unclear in empirical analysis and the trends in the digital discourse of the right wing.

This research focuses primarily on classifying based on the metadata on the tweet and understanding the discourse of the right wing on Twitter and its underlying trends. Assisting the metadata model will be another model that will be trained on the textual analysis of the tweet's text.

1. To state the exact key contributions
2. Understanding the digital discourse of the right wing on Twitter.
3. Training two different algorithms for a multimodal approach.
4. Derive insights from the dataset and its trends.
5. Ponder over solutions from multiple disciplines.

2. RELATED WORK
2.1 SOCIAL SCIENCES AND OTHER DISCIPLINES
[37] Their paper highlights the typology, taxonomies, and the importance of timely detection of counterfacts. Additionally distinguishes them into misinformation (unintentional), disinformation (intentional), and mal information (unintentional). Finally, the paper briefly examines other datasets and themes in counterfacts in Rumor, Satire, Fake news and clickbait.[38] Examines the digital discourse of the Indian right wing (they use sarcasm and mockery as their tool in propaganda) and highlights their insecurities regarding big tech of social media, whose HQ is not in India, and their fallacy of the platform KOO.[6],[7],[2],[3],[5],[30] All of the papers emphasize on how the user locks himself into informational groups and silos and believes the propaganda based on the "source" rather than the authenticity of the news, and how entertainment furthers the information groups, enabling character assassinations and steering the democratic discourse into elitist benefits.[5] Emphasizes the higher penetration of right-wing posts and the composition of the posts, and not all of those consisted of text, but all of them consisted of metadata.

2.2 DATA SCIENCE
[20],[21],[11],[14] These papers mostly are based on textual analysis and some of them based on hybrid/multimodal analysis [22],[21],[20],[14],[11], most the accuracies of these papers range from 34% to 99% based on multiple machine learning, deep learning and transformers based methods to solve this problem.[19] Highlighted that the data from Twitter is very noisy[15] who built an ANN to identify fake news profiles.[16] Proposed CSI model that uses RNN to understand the behaviors of text and identify fake news.[18] Presented a multi-disciplinary approach incorporating sociology, psychology in their approach to propose a better solution.[17] Investigated and identified the authenticity and user's credibility in unsupervised ways. Fact checking websites include Alt News[30], Boom Live[31] and others [32],[33],[34],[35] are seen reliable. Finally, many datasets including PHEME, Twitter 15, 16 and many others were reviewed for this research none of those are based in India and more focus propaganda detection.

3. RESEARCH METHODOLOGY
3.1 DATASET COLLECTION
The dataset was collected from 55 and 51 right ideological and non/anti right ideological accounts of 51000 size and then divided into two independent datasets. The ANN dataset contains 51000 records of Conversation Controlled, possibly sensitive, source, type, isPinned, isQuote, isReply, isRetweet, media/0, media/1, lang, likeCount, replyCount, quoteCount, retweetCount, and target. Some columns were exempted from the dataset for computation optimization. The CNN-BiLSTM dataset contains 48000 records of text and target.

3.2 Dataset Preprocessing
Both datasets were balanced after being cleaned for duplicates, empty values, and NaN values; if any of these conditions were found, the row was purged. The text from CNN-BiLSTM's dataset was translated to English using the Google Translate library for computation purposes

3.3 Project Architecture
ANN classifier is designed mathematically to mimic the brain's pattern recognition and is better than the human brain in some aspects.
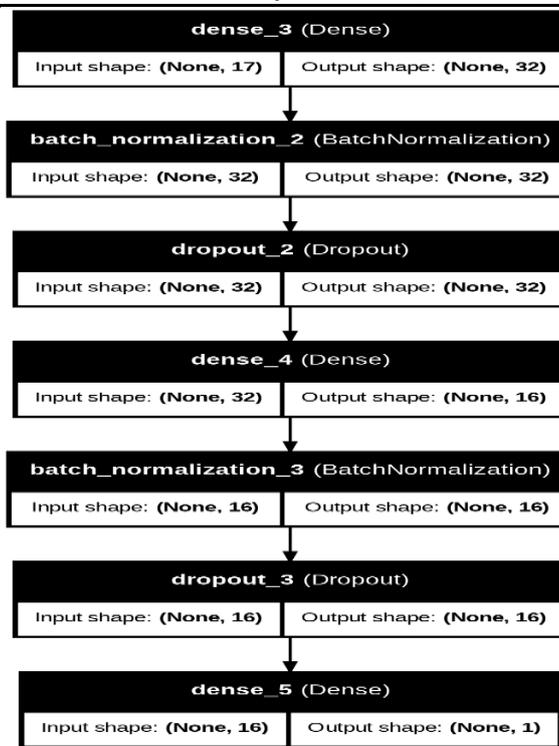The architecture of our ANN is:

dense_3 (Dense)	
Input shape: **(None, 17)**	Output shape: **(None, 32)**

batch_normalization_2 (BatchNormalization)	
Input shape: **(None, 32)**	Output shape: **(None, 32)**

dropout_2 (Dropout)	
Input shape: **(None, 32)**	Output shape: **(None, 32)**

dense_4 (Dense)	
Input shape: **(None, 32)**	Output shape: **(None, 16)**

batch_normalization_3 (BatchNormalization)	
Input shape: **(None, 16)**	Output shape: **(None, 16)**

dropout_3 (Dropout)	
Input shape: **(None, 16)**	Output shape: **(None, 16)**

dense_5 (Dense)	
Input shape: **(None, 16)**	Output shape: **(None, 1)**

figure 1: the architecture of ann

Equations

The mathematical notation of MLP is:

$$Y_j^l = \alpha((y_j^{(l-1)} + \omega_{(i,j)}^l) + b_j^l) \qquad (1)$$

Where:

Y_j^l = is the output.

A = activation function

$y_j^{(l-1)}$ = the output of the previous layer.

$\omega_{(i,j)}^l$ = the weights.

b_j^l = the biases

The mathematical notation for tanh:

BiLSTM overcomes the gradient vanishing problem paired with CNN for feature extraction, churning out a computationally optimized model that yields better results.

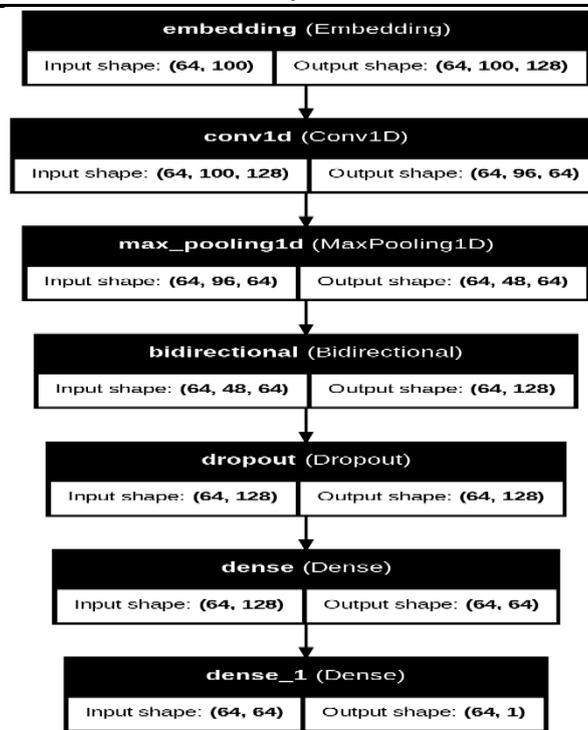the architecture of the CNN-BiLSTM is shown in Figure 2.

figure 2: the architecture of cnn

The mathematical formulation of Bi-LSTM is:

$$y_t = w_y \cdot h_t + b_y \qquad (2)$$

$$h_t = [\vec{h}_t; \overleftarrow{h}_t] \qquad (3)$$

Where:

- y_t = is the output
- w_y = weights
- $\vec{h}_t = \text{LSTM}_{forward}(x_1, x_2, \ldots, x_t)$
- $\overleftarrow{h}_t = \text{LSTM}_{backward}(x_n, x_{n-1}, \ldots, x_t)$

4. RESULTS

4.1 Results of Descriptive Statics of Study Variables

The initial hypothesis states the right wing is coordinated and well organised, yielding them more engagement than the non-right, which will be proved right in further

The metrics of the ANN model are shown in table 1.

table 1: the metrics of ann

METRIC	VALUES
ACCURACY	97
PRECISION	0.9720850740600075
RECALL	0.9959143968871595
F1 SCORE	0.9838554679992312
AUC ROC SCORE	0.9972804170960152

The hyperparameters that were included are:

1. Activation function:
   1.1 tanh: better generalization than Relu.
2. Overfitting tools:
   2.1 11, l2 regularization.
   2.2 Batch normalization.
   2.3 Dropout
   2.4 Early stop
   2.5 ReduceLROnPlateau: makes learning rate dynamic.
3. optimizer:
   3.1 rmsprop:
1. more stable than ADAM
      2. Fewer hyperparameters to tune (light model)

The metrics of CNN-BiLSTM are shown in table 2.

table 2: the metrics of cnn

METRIC	VALUES
ACCURACY	79
PRECISION	79
RECALL	79
F1 SCORE	79
AUC ROC SCORE	87

The hyperparameters that were included are:

1. Activation function:
   1.1 tanh: better generalization than Relu.
2. Overfitting tools:
   2.1 11,l2 regularization.
   2.2 Dropout
   2.3 Early stop
   2.4 ReduceLROnPlateau: makes learning rate dynamic.

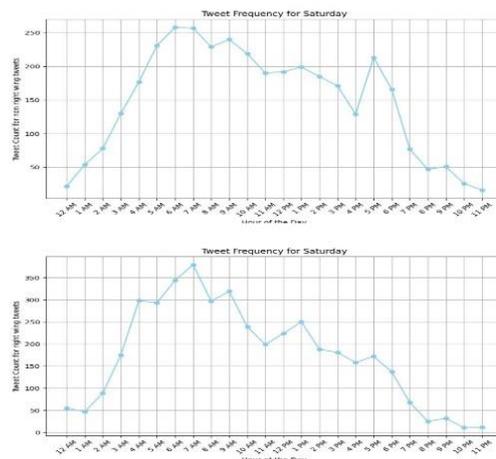The posting frequencies of both wings on Saturday are shown in Fig. 3.

figure 3: posting frequencies of both wings on Saturday

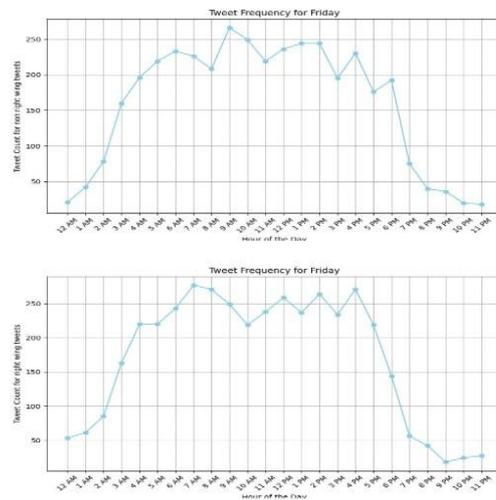The posting frequencies of both wings on Friday are shown in Fig. 4.

figure 4: posting frequencies of both wings on saturday

The posting frequency on a weekly basis for both wings is shown in Fig. 5.
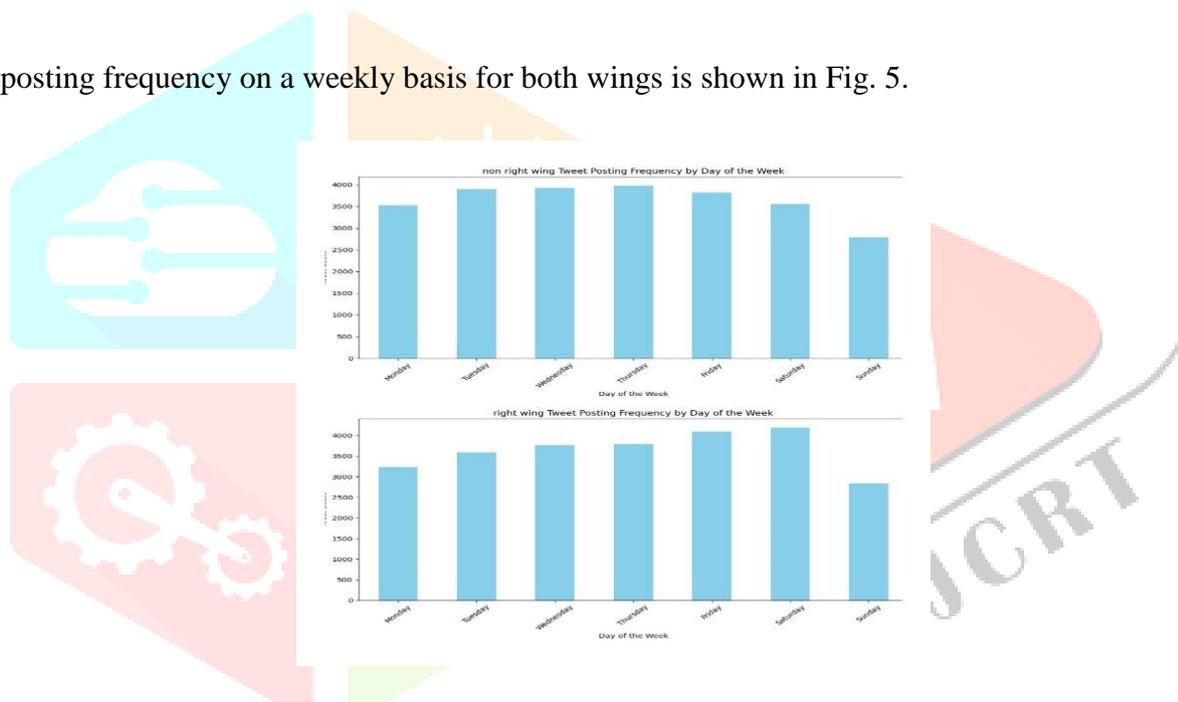

figure 5: posting frequency on a weekly basis of both wings

As seen, they pose a min-max function-like activity and spikes in the right are seen at 7 am, 12 pm, and 3 pm were whereas there are little but existent patterns in the left, but to most it is random. When a word cloud of the 500 most-used words was plotted, it was observed that both labels discussed similar topics. A few words are very frequent in the right wing, which is not the case in the non-right wing, where the words are too diverse. Many more topics were being discussed. Those tweets, when fed to Bert or any flavor of Bert, did not classify any of the tweets from either case as toxic or abusive or sarcastic, even when from the dataset these examples are present in it: "Abe Piddiyon, kitni baar Baap badloge?", "No D-STOCK is hurt here".

The word cloud of the right-wing tweets is shown in Fig. 6.

figure 6: right-wing cloud

The word cloud of the non-right-wing tweets is shown in Fig. 7.

figure 7: non-right-wing cloud

CONCLUSIONS:

The hypothesis is proved empirically that the posting frequency's pattern, with word cloud's same words being used and the topics likewise (which would be normal of non-right would be the same). this highly suggests that the right wing in their digital discourse/propaganda is very coordinated and well organized in deepening the already existing communal tensions. Right-wing and its actors share the same rights to exist in the digital and political sphere as much as any other ideology or an individual/entity does. the fault, dear reader, is not in our political/digital sphere but in us that we stoop to unethical persuasion, which is beneath us, problematized it must be. the best solution to this is to strengthen democracy and its people [20] to resist the propaganda and educate them enough about the evils of social media and propaganda. some of the personal solutions are to enforce laws against hate speech and sarcasm that promote anti-democratic ideas, deeper understanding and a vibrant approach to this problem from data science, and to produce more datasets on this topic. Finally, we would like to thank the head of the IT department of the Lords Institute of Engineering and Technology for approving and funding this project.

REFERENCES

[1] Diederik P.K., Jimmy L.B., "ADAM: A Method for Stochastic Optimization", ICLR, January 2015.

[2] Matteo S., "(Dis)Information, Neoliberalism and the Strength of Democracy in the Digital Age", Cambridge University Press, May 2021.

[3] Jonathan P.M., "Disinformation Society, Communication and Cosmopolitan Democracy", Cosmopolitan Civil Societies: An Interdisciplinary Journal, Vol. 9, No. 2, 2017.

[4] Altay S., Berriche M., Heuer H., Farkas J., Rathje S., "A Survey of Expert Views on Misinformation: Definitions, Determinants, Solutions, and Future of the Field", Zurich Open Repository and Archive.

[5] Md. Sayeed A.Z., "Social Media Fake News in India", Asian Journal for Public Opinion Research, Vol. 9, No. 1, February 2021.

[6] Simons G., "Fake News: As the Deeper Problem or a Symptom of a Deeper Problem?", 2017.

[7] Jamie B., "The People vs Tech: How the Internet is Killing Democracy (and How We Save It)", 2018.

[8] M.S. Looijenga, "The Detection of Fake Messages Using Machine Learning", B.S. Thesis, University of Twente, 2018.

[9] O. Ajao, D. Bhowmik, S. Zargari, "Sentiment Aware Fake News Detection on Online Social Networks", in ICASSP 2019 – IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), IEEE, 2019, pp. 2507–2511.

[10] X. Zhou, J. Wu, R. Zafarani, "SAFE: Similarity-Aware Multi-Modal Fake News Detection", arXiv Preprint, Vol. 200304981, p. 2, 2020.

[11] X. Zhi, L. Xue, W. Zhi, Z. Li, B. Zhao, Y. Wang, Z. Shen, "Financial Fake News Detection with Multi-Fact CNN-LSTM Model", in 2021 IEEE 4th International Conference on Electronics Technology (ICET), IEEE, 2021, pp. 1338–1341.

[12] K.M. Fouad, S.F. Sabbeh, W. Medhat, "Arabic Fake News Detection Using Deep Learning", Computers, Materials & Continua, Vol. 71, No. 2, 2022.

[13] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, L. Su, J. Gao, "EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection", Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 849–857.

[14] Castellini J., Poggioni V., Sorbi G., "Fake Twitter Followers Detection by Denoising Autoencoder", 2017, pp. 195–202, doi:10.1145/3106426.310648.

[15] Ruchansky N., Seo S., Liu Y., "CSI: A Hybrid Deep Model for Fake News", 2017.

[16] Yang S., Shu K., Wang S., Gu R., Wu F., Liu H., "Unsupervised Fake News Detection on Social Media: A Generative Approach", Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, 2019, pp. 5644–5651, doi:10.1609/aaai.v33i01.3301564.

[17] Zhou X., Zafarani R., "Fake News: A Survey of Research, Detection Methods, and Opportunities", arXiv, abs/1812.00315, 2018.

[18] Srishti S., Mala S., Anil K.D., "Fake News Detection on Twitter", 2022.

[19] Shreya G., Prasenjit M., "Tweeted Fact vs Fiction: Identifying Vaccine Misinformation and Analyzing Dissent", 2023.

[20] Shreya G., Prasenjit M., "Catching Lies in the Act: A Framework for Early Misinformation Detection on Social Media", 2023.

[21] Shreya G., Prasenjit M., "How Early Can We Detect? Detecting Misinformation on Social Media Using User Profiling and Network Characteristics", 2023.

[22] "PHEME Dataset." https://www.kaggle.com/datasets/usharengaraju/pheme-dataset

[23] "Rumor Detection on Twitter15 & Twitter16 Data." https://github.com/OwenLeng/rumor-detection-include-twitter15-twitter16data

[24] "LIAR Dataset." https://github.com/tfs4/liar_dataset

[25] "CREDBANK Data." https://github.com/compsocial/CREDBANK-data

[26] "PolitiFact Fact-Check Dataset." https://www.kaggle.com/datasets/rmisra/politifact-fact-check-dataset

[27] "IFND Dataset." https://www.kaggle.com/datasets/sonalgarg174/ifnd-dataset

[28] "AltNews–MethodologyforFactChecking." https://www.altnews.in/methodology-for-fact-checking/

[29] "Boomlive Fact-Checking Platform." https://www.boomlive.in/

[30] "SM Hoax Slayer." https://smhoaxslayer.com/

[31] "The Quint WebQoof." https://www.thequint.com/news/webqoof

[32] "Factly." https://factly.in/

[33] "FactCheck.org." https://www.factcheck.org/

[34] Shubhangi R., Divya B., "A Review on Fake News Detection: 3T's Typology, Time of Detection,

Taxonomies", November 2022.

[35] Kiran V.B., Payal A., "Clip the Blue Bird: Discursive Strategies of Hindutva Digital Mobilization Against Twitter in India", March 2024.

[36] "Indian Right-Wing Propaganda Detection." https://github.com/theameendawoodi/indian-right-wing-propaganda-detection.git

[37] Swati C., "I Am a Troll", 2017.

[38] Global Economic Forum, "The Global Risk Report 2025", ISBN: 978-2-940631-30-8.