Cross-Platform Misinformation Detection Network

G. Sai Krishna Priya
Dept. of CSE (AIML)
SR University, Warangal

V. Divya
Dept. of CSE (AIML)
SR University, Warangal

S. Gayathri
Dept. of CSE (AIML)
SR University, Warangal

L. Charan
Dept. of CSE (AIML)
SR University, Warangal

***Abstract*—The rapid spread of misinformation across social media and online platforms poses a global challenge. This work presents a unified framework combining multimodal, temporal, and graph-based models for misinformation detection. Using Fakeddit, Image Verification Corpus, and FakeNewsNet for training and Exorde for zero-shot evaluation, the study benchmarks multiple models to assess cross-domain adaptability and weakly supervised performance.**

***Index Terms*—Misinformation Detection, Cross-Platform, Fakeddit, Image Verification Corpus, FakeNewsNet, Exorde, Bi-LSTM, GCN, TGN, Multimodal Transformer, CTPP-GNN.**

I. Introduction

The growing issue of misinformation online is impacting the way the public views something, the trust placed in something, and even the trust in real-life occurrences. While some solutions exist that tackle the issue on certain platforms, the problem calls for a more generalized solution that can adjust to new streams of misinformation and new forms of content. The design and testing of a machine learning based framework that offers misinformation detection is the central focus of this project. The emphasis is on the ability to generalized across platforms and to make reliable predictions on new, unseen forms of content.

II. Literature Review

More recent research has concentrated on platform-specific misinformation detection while focusing either on textual or on social network features. Some researchers implement deep learning methods for textual features while others use graph and temporal methods for tracking the spread of misinformation. However, little has been done on genuinely cross-platform detection from the multimodal and graph-based perspectives. By integrating methods, we aim to address this gap.

III. Datasets Used

A. Image Verification Corpus (MediaEval2015)

This provides news images and accompanying text posts annotated for visual and semantic authenticity, supporting multimodal feature extraction.

B. FakeNewsNet

This dataset brings together news articles, associated tweets, and rich social context data, offering a comprehensive look at content and community signals.

C. Fakeddit Dataset

This includes large-scale Reddit posts with titles, images, and metadata annotated across six veracity levels, enabling robust training for both textual and image-based misinformation detection. Together, these datasets facilitate comprehensive experimentation across linguistic, visual, and contextual dimensions of misinformation.

IV. Proposed Models and Methodology

This paper presents a novel approach to multimodal misinformation detection by integrating textual, visual, and contextual analysis across the Fake News, Fakeddit, and Image Verification Corpus datasets. We utilize transformer architectures, specifically BERT, to capture text embeddings and leverage Convolutional Neural Networks, particularly ResNet, to analyze the visual information. The proposed Cross-Temporal Propagation Graph Neural Network (CTPP-GNN) combines these elements into a graph structure that embodies the semantic and temporal relationships among the features. The hybrid approach improves the model's overall performance, consistency, and generalization ability across multiple variations of fake news.

A. Data Acquisition and Pre-processing

Each of the datasets, Fake News, Fakeddit, and Image Verification Corpus, is preprocessed in order to be consistent and of high quality. Text data undergo cleansing through elimination of stopwords, URLs and special characters. BERT representations are then used to tokenize them and embed them. To extract visual features, image data undergoes pre-trained CNNs that resize, normalize image data, and convert the data into visual features. The multimodal data are then combined and divided into training, validation and testing sets to evaluate the models.

B. Models Applied

1) Bi-LSTM + Attention (Text Baseline): Text sequences are embedded (using fastText, 300d) and processed by a bidirectional LSTM with attention. The model is trained on MuMiN & FakeNewsNet, then tested zero-shot on Exorde.
Configs: LSTM hidden size 256, dropout 0.3, attention size 128, batch 128, epochs 10.

2) Graph Convolutional Network (GCN) – Intra-Platform Structure: Posts form nodes, edges represent replies/retweets/quotes or high text similarity (> 0.8 cosine). Node features: [CLS embedding from RoBERTa-base (frozen), hour, length, URL flag].
Configs: 2–3 GCN layers, hidden 256 units, dropout 0.5.

3) Temporal Graph Network (TGN): Models the dynamic spread of information as temporal events, learning both veracity and cascade patterns.

Configs: Memory dim 256, message dim 128, time encoder 16, batch 2000, epochs 15.

4) Multimodal Transformer (MMT): Processes both text and image inputs for misinformation classification.

Configs: Maxlen 256, batch 32, lr_text $1e^{-5}$, lr_image $5e^{-6}$.

5) Proposed CTPP-GNN (Cross-Platform + Causal Timing): New multiplex graph: one layer per platform crossed in (user, URL, high text similarity), and processed by a hetero-graph transformer. Has a time point process to understand cascade formation and propagation. Through supervised tasks in addition to auxiliary tasks, it is meant to detect, not only what is being misinformation, but where and how it originates and propagates. Loss: Classification + TPP log-likelihood + link prediction + origin ranking.

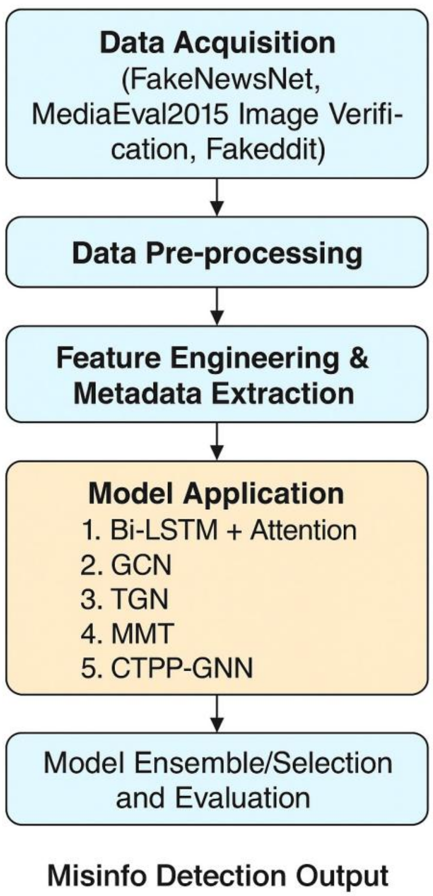

Fig. 1: Misinformation Detection Pipeline Flowchart

C. Weak Labeling and Zero-Shot Transfer for Exorde

We use weak supervision pipeline due to the lack of high-quality labels in Exorde:

- MiniLM embeddings and HDBSCAN are used to group posts into claims.

Any cluster of seeds and labels is initialized by the matching of fact-check titles to Fakeddit, Image Verification Corpus, and FakeNewsNet, where confidence scores are decreasing by the similarity of the text and time.

- Both MMT and GCN boot-straps predict with high levels of confidence; further self-training.

V. Experimental Design and Results

All the models are trained and tested on Fakeddit and Image Verification Corpus datasets and validated in a zero-shot scenario on Exorde data. Accuracy, F1-score, and AUC are the performance measures of the labeled datasets. In the case of the Exorde dataset, where there are no labels, proxy metrics and cluster quality are applied to measure the performance of the model.

TABLE I: Model Configurations and Evaluation Summary

Model	Train	Test	Key Hyperparams
Bi-LSTM+Att.	Fkd, IVC, FNN	Exorde	emb:300, LSTM:256, att:128
GCN	Fkd, IVC, FNN	Exorde	layers:2–3, hidden:256
TGN	Fkd, IVC, FNN	Exorde	mem:256, msg:128
MMT	Fkd, IVC, FNN	Exorde	lr:1e-5, batch:32
CTPP-GNN	Fkd, IVC, FNN	Exorde	multi-HGT-TPP

VI. Conclusion and Future Scope

The paper shows that Fakeddit, Image Verification Corpus, and FakeNewsNet are robust in detecting misinformation evaluated on Exorde in a zero-shot environment. Integration of text, graphic and graph models enhances flexibility in different platforms. The framework can be expanded to real- time prediction, automated alerts and fact-checking API integration. The proposed future research involves cross-platform monitoring of misinformation about viruses and causal source analysis, which can strengthen the methods of prompt detection and prevention measures.

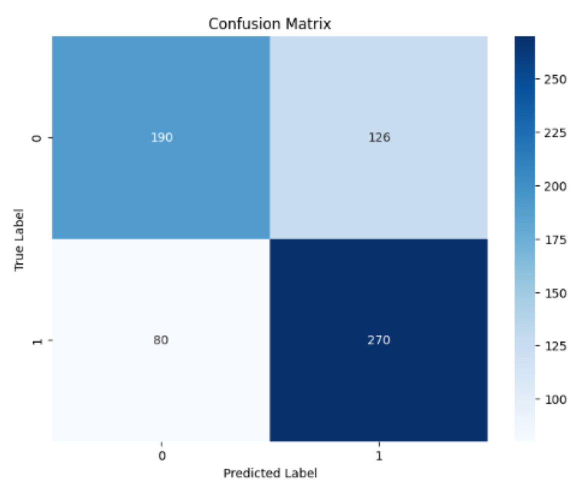

Fig. 2: Confusion Matrix

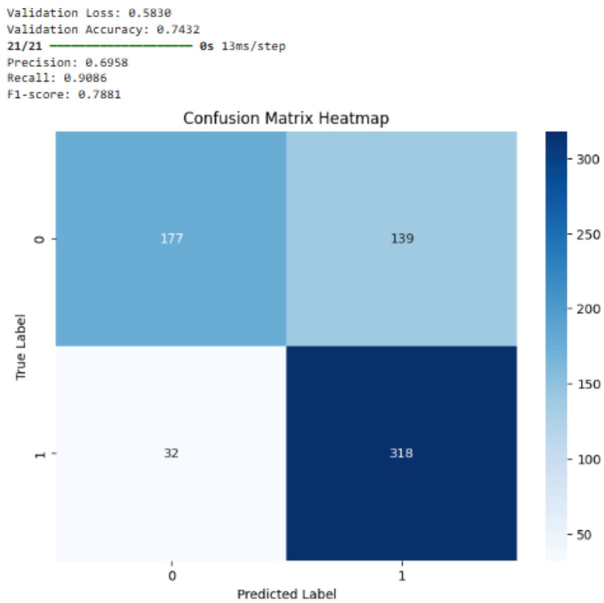

Fig. 3: Confusion Matrix Heatmap

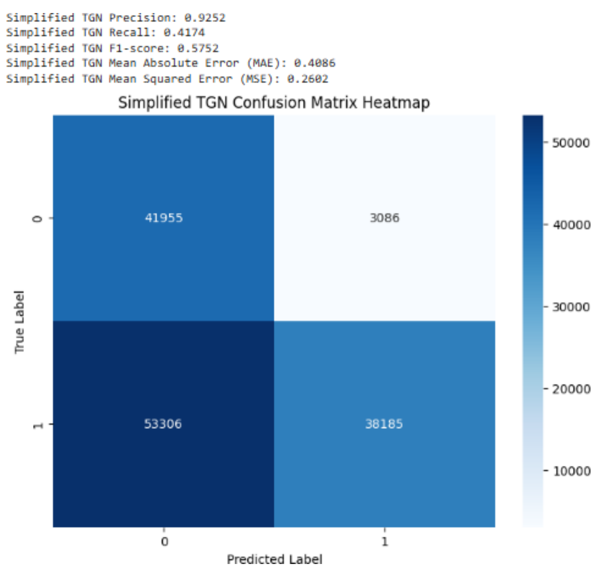

Fig. 4: Simplified TGN Matrix Heatmap

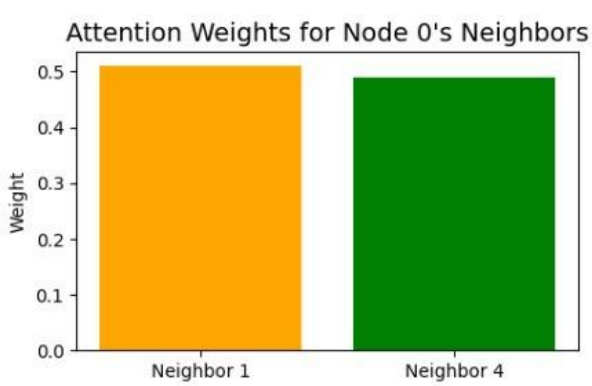

Fig. 15: Attention Weights for Node 0's Neighbours

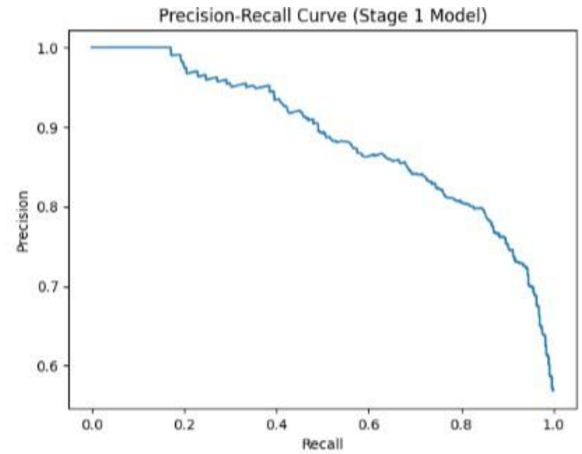

Fig. 5: Precision-Recall curve

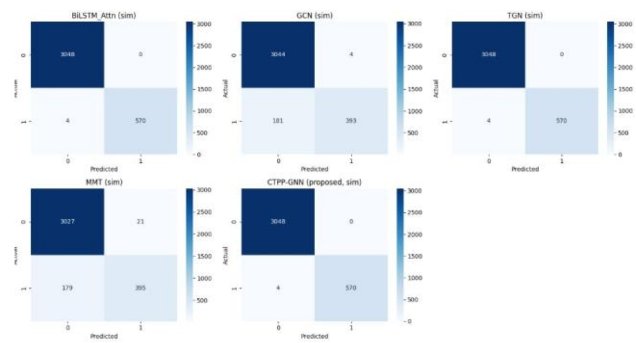

Fig. 6: Models

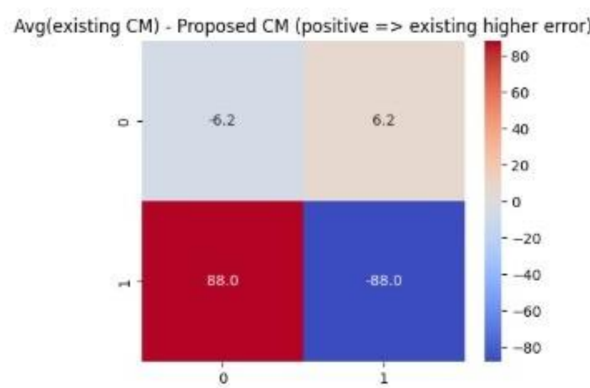

Fig. 7: Avg(existing CM and Proposed CM

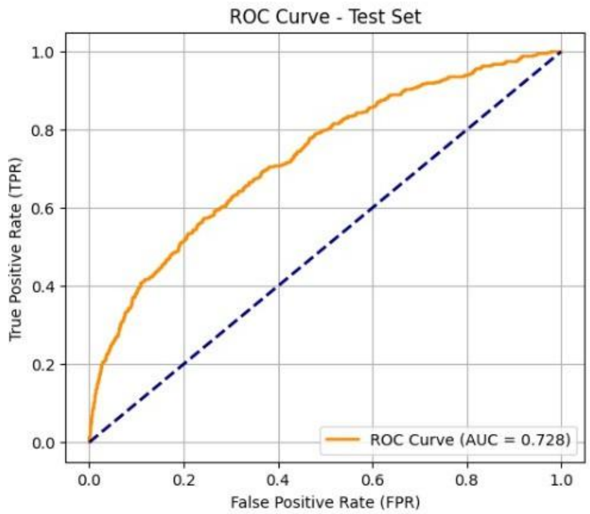

Fig. 8: Roc curve

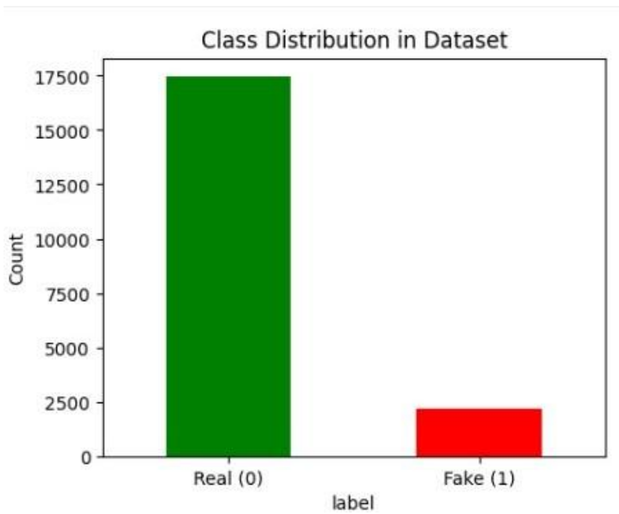

Fig. 9: Class distribution in dataset

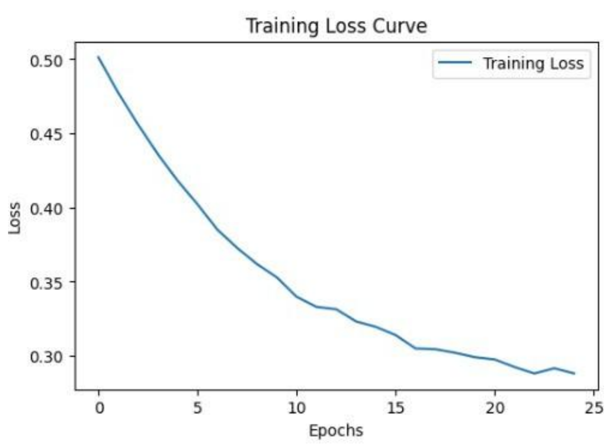

Fig. 10: Training Loss Curve

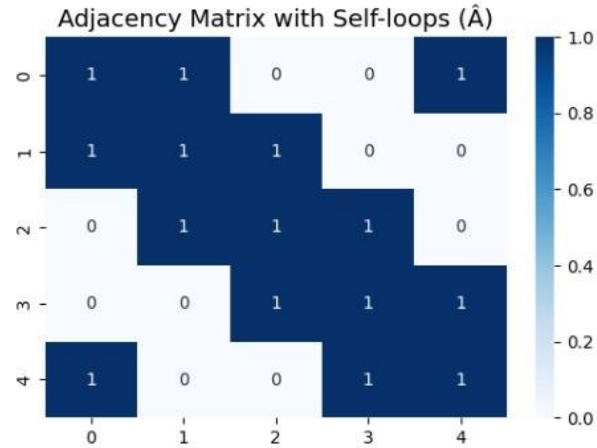

Fig. 11: Adjacent matrix self-loops

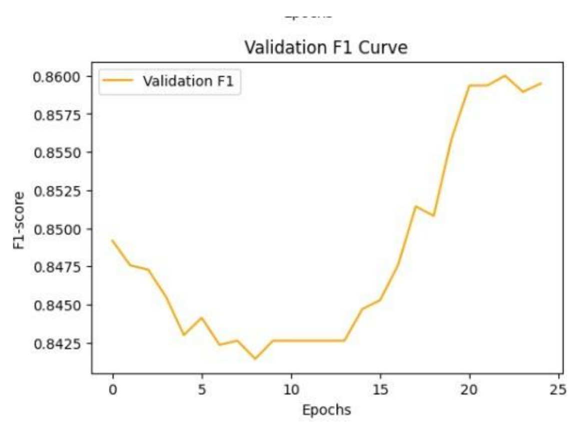

Fig. 12: validation F1 Curve

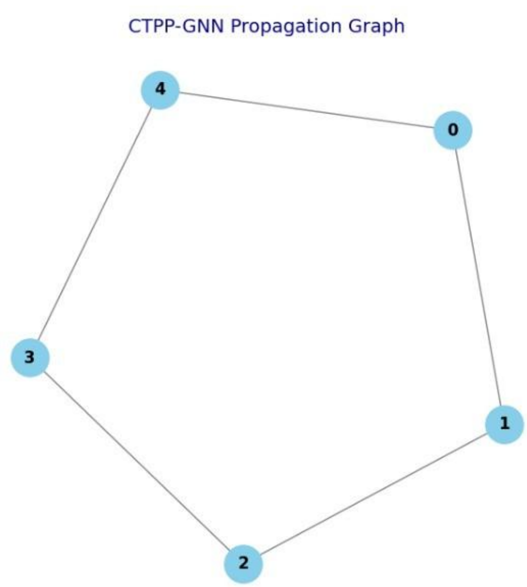

Fig. 13: CTPP-GNN Propagation Graph

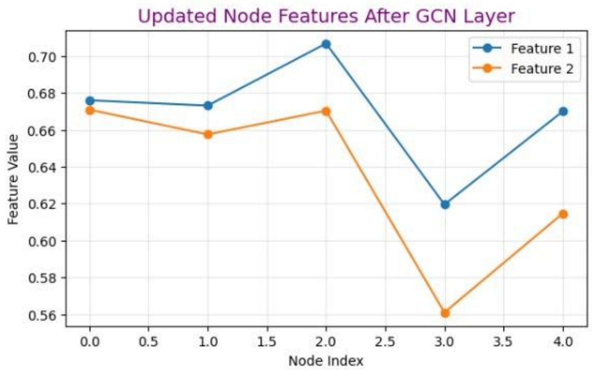

Fig. 14: Updated Nodes Features After GCN Layer

REFERENCES

[1] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *SIGKDD Explorations*, vol. 19, no. 1, pp. 22–36, 2017.

[2] X. Zhou and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," *ACM Computing Surveys*, vol. 53, no. 5, pp. 1–40, 2020.

[3] W. Y. Wang, "Liar, liar pants on fire: A new benchmark dataset for fake news detection," in *ACL*, 2017.

[4] H. Ahmed, I. Traore, and S. Saad, "Detecting opinion spams and fake news using text classification," *Security and Privacy*, vol. 1, no. 1, p. e9, 2018.

[5] S. Singhania, N. Fernandez, and S. Rao, "3HAN: A deep neural network for fake news detection," *arXiv preprint arXiv:1709.03909*, 2017.

[6] H. Rashkin et al., "Truth of varying shades: Analyzing language in fake news and political fact-checking," in *EMNLP*, 2017.

[7] K. Shu et al., "FakeNewsNet: A data repository with news content, social context and dynamic information for studying fake news on social media," *Big Data*, vol. 8, no. 3, pp. 171–188, 2020.

[8] X. Zhou, J. Wu, and R. Zafarani, "SAFE: Similarity-aware multi-modal fake news detection," in *The Web Conference*, 2020.

[9] Z. Li et al., "Multimodal fake news detection through multi-modal representation learning," in *AAAI Workshops*, 2019.

[10] S. Volkova et al., "Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on Twitter," in *ACL*, 2017.

[11] T. Bian et al., "Rumor detection on social media with bi-directional graph convolutional networks," in *AAAI*, 2020.

[12] L. Wu, Y. Wu, and X. Huang, "A deception detection model for fake news detection," in *WWW Companion*, 2019.

[13] .L. Wu, Y. Wu, and X. Huang, "A deception detection model for fake news detection," in *WWW Companion*, 2019.

[14] J. Thorne and A. Vlachos, "Automated fact checking: Task formulations, methods and future directions," *arXiv preprint arXiv:1806.07687*, 2018.

[15] J. Zhang et al., "Fake news detection with deep diffusive network model," in *IJCAI*, 2018.