JOURNAL OF
EMERGING INVESTIGATORS

Article

Innovative fake health news detection: Integrating emotional features into graph neural networks

Yiheng Wang[1], James Wang[1]
[1] Princeton International School of Mathematics and Science, Princeton, New Jersey

SUMMARY

In the digital age, the rapid dissemination of fake health news on social media has posed a serious threat to public health decision-making. Artificial intelligence (AI) has been increasingly used in detecting fake health news by analyzing textual content through models such as recurrent neural networks, long short-term memory, and large language models. However, sequence-based AI models often neglect social context information, limiting their detection performance. To address this limitation, we propose a graph-based model, X-Health News Detection (X-HND), which enhances health-related fake news detection by jointly analyzing textual content and propagation information describing how news spreads among users in the social network. Additionally, we constructed a specialized health-related dataset, Health-News Dataset (HNDataset), for this model. The proposed model integrates graph-based propagation structures with emotion-aware textual features within a domain-specific health dataset, enabling it to capture both how information spreads and how it is expressed. We hypothesize that graph structures incorporating spread paths carry valuable information for fake news detection, while emotional cues in text also improve accuracy. Experimental results comparing different models and datasets indicate that incorporating graph-structured propagation information and emotion features improves fake news detection performance. This research offers a promising new approach for enhancing health misinformation detection through emotionally informed graph modeling.

INTRODUCTION

Over the past decade, social media platforms such as X (formerly Twitter), Facebook, and TikTok have emerged as the primary news sources for a growing number of users due to their immediacy, accessibility, and user-generated content. While this shift enables faster access to information, it has also introduced new challenges—most notably, the rapid dissemination of fake news. Fake news undermines public distrust and deepens social polarization (1). According to a 2024 study by Reuters institute, 13–27% of social media users find it difficult to identify trustworthy news on social media platforms (2). Health-related fake news can cause even more harm to society, as it can influence individuals' decisions regarding medical treatments and public health behaviors. Furthermore, due to the presence of complex professional terminology and domain-specific knowledge, the credibility of health-related news is more difficult for the general public to assess.

Fake health news on social media is characterized by its large volume, diverse formats, and real-time nature, making it particularly difficult to detect and contain (3). Additionally, the proliferation of artificial intelligence (AI) tools, such as deepfake generators and text-based content creation models, has drastically reduced the cost of producing convincing false content (4). Automated bots further exacerbate this issue by accelerating the dissemination of misinformation, sometimes within minutes (5). Fake news detection methods can be broadly categorized into two main categories: content-based analysis and social context-based analysis (6). The former focuses on analyzing the text or content of the news itself, while the latter involves examining user interactions, propagation dynamics, and stance signals related to the news. Recent studies have shown that incorporating socio-contextual information with textual content enhances the overall accuracy of fake health news detection (7). Prior research has also shown that fake news often displays distinct social behaviors, such as highly emotional involvement patterns or coordinated sharing by bots (8). Manual factchecking is time-consuming and cannot keep up with the volume of misinformation online. In response, researchers have turned to AI and machine learning-based detection methods. While models such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and large language models like ChatGPT have shown promise in text processing, they lack the ability to incorporate social context information, limiting their effectiveness in fake news detection (6,7).

In contrast, Graph Neural Networks (GNNs) are capable of capturing relational patterns, such as user behaviors and news propagation structures, and can combine the relational information with content analysis, showing greater promise for misinformation detection (9). GNN-based methods model the social context of news by capturing different types of interactions, such as user-user, user-news, and news-news relations (10). However, a limitation of current GNN-based models lies in their underutilization of textual content features, such as emotion, linguistic style, and inter-text similarity (11). Furthermore, while many studies focus on designing complex model structures, relatively little attention has been paid to enhancing model performance through improved dataset construction and preprocessing.

To address these gaps, this study proposes a novel graph-based model, X-HND (X-Health News Detection), aimed to detect health-related fake news on social media by integrating both content features and social context. The model leverages a style-based content analysis approach and a propagation-based context analysis framework. Central to the model is a custom-built dataset, the Health-News Dataset (HNDataset), consisting of health-related posts collected from X. We hypothesized that (i) graph-based models can improve the accuracy of health-related fake news detection, (ii) training on a domain-specific dataset outperforms a general-purpose dataset, and (iii) integrating emotion-based features into graph neural networks improves fake health news detection performance. The third hypothesis is motivated by prior research demonstrating that emotional intensity plays a critical role in the spread and amplification of misinformation, particularly in the health domain (8). Validating these hypotheses may provide valuable insights into designing effective fake health news detection models and building high-quality datasets with rich features.

Using standard evaluation metrics, we identified the most effective model configurations for real-world deployment. The finalized X-HND model achieves strong performance, offering a promising approach for combating health misinformation on social media and contributing to the growing body of knowledge on automated misinformation detection.

RESULTS

To evaluate the proposed X-HND framework, we conducted a set of experiments focusing on model architecture, dataset choice, and feature design. The rationale for this experimental design was to examine how model structure, health-specific dataset, and emotion-based feature augmentation influenced fake news detection performance. Accordingly, the experiments consisted of three parts. Part I compared the effectiveness of the graph-based models with sequence-based models. Part II evaluated the necessity of constructing a domain-specific dataset for fake health news detection. Part III explored how different dataset construction and preprocessing strategies influenced the performance of the proposed X-HND model.

Part I: Comparison of graph-based and sequence-based models

We hypothesized that graph-based architectures that incorporate social propagation information would outperform sequence-based models (RNN, LSTM, Transformer) that relied solely on textual features for fake news detection. To evaluate the effectiveness of the graph-based model, we compared various sequence-based models with the X-HND graph-based fake news detection model by training and testing these models on HNDataset-BASE, a health-specific baseline dataset of social media posts and assessing their performance on an identical fake news classification task. Model performance was evaluated using accuracy, precision, recall, and F1 score. Accuracy indicated overall classification correctness, precision measured the reliability of fake news predictions, recall reflected the model's ability to identify fake news, and the F1 score summarized precision and recall into a single balanced metric.

The detection accuracy of the RNN model was 0.55, LSTM achieved 0.65, and the Transformer-based model (Chat-GPT) reached 0.69. The graph-based model outperformed all others with an accuracy of 0.71 (**Figure 1**). Consistent trends were observed across other evaluation metrics: precision increased from 0.60 (RNN) to 0.75 (graph-based), recall from 0.55 to 0.70, and F1 score from 0.57 to 0.72. These results demonstrated a progressive improvement from simpler to more complex architectures, with the graph-based model showing the best performance, supporting the hypothesis that integrating social context through graph structures enhanced fake news detection.

Part II: Evaluating the necessity of a customized health dataset

Having established the advantage of graph-based models over sequence-based approaches, we next examined whether the use of a health-specific dataset further improved fake health news detection performance. We hypothesized that training on a health-specific dataset would improve fake health news detection accuracy compared to training on a general-purpose news dataset, due to domain-specific semantic patterns and contextual features. In this experiment, two datasets with distinct domains were used. The publicly available UPFD (User Preference-aware Fake News Detection) Politifact dataset focused on political news, whereas HNDataset focused on health-related social media content, such as medical and public health topics, as reflected in the word clouds (**Figure 2, 3**) (12,13). The UPFD Politifact dataset contained 314 graphs, while HNDataset contained 244 graphs (**Table 1**). To assess the impact of dataset customization, we conducted a comparative experiment using our X-HND model under two training conditions. In the first setting, the model was trained on the UPFD Politifact dataset and tested on the customized HNDataset-BASE. In the second setting, both training and testing were performed on HNDataset-BASE. We split the dataset into 60% training and 40% testing, a common practice that balanced learning and evaluation (14). The larger training set helped the model capture meaningful patterns, while the testing set remained large enough to provide reliable performance estimates. In both cases, X-HND employed the same core architecture, with the only difference being the choice of training data.

The model trained on the UPFD dataset achieved a detection accuracy of 0.45 when applied to fake news in the HNDataset. In contrast, the model trained directly on the

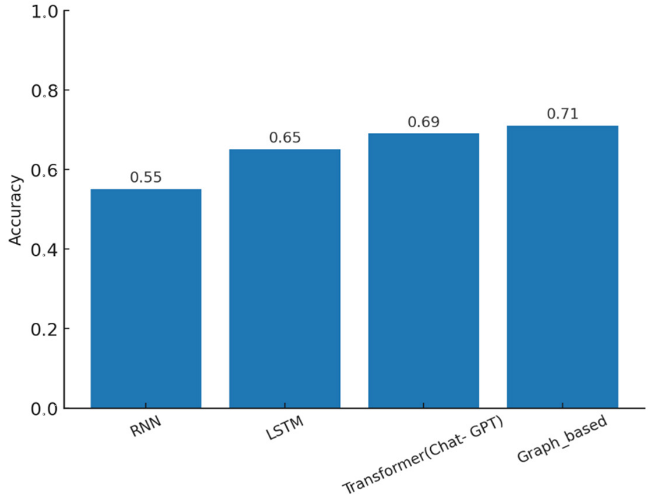

Figure 1. Comparison of graph-based and sequence-based models for health fake news detection. Detection accuracy of various machine learning models, including RNN, LSTM, Transformer-based models (e.g., ChatGPT), and the proposed graph-based X-HND model. The four models were all trained and tested on HNDataset-BASE, and performance was assessed on an identical fake news classification task. The graph-based model achieved the highest accuracy (0.71), outperforming the Transformer (0.69), LSTM (0.65), and RNN (0.55) models.

https://doi.org/10.59720/25-130


(A) (B)

Figure 2. Word clouds of true and fake news in the Health-News dataset. Word clouds were generated from the root news content in the baseline Health-News dataset (HNDataset-BASE). The most frequent words found in (A) true health news and (B) fake health news.

HNDataset achieved a much higher detection accuracy of 0.71 on the same dataset (**Figure 4**). The observed improvement was substantial, with an increase of 26 percentage points in accuracy. These results highlighted the importance of constructing a domain-specific dataset tailored to health misinformation detection. General-purpose political fake news datasets, such as Politifact, failed to transfer effectively to the health domain.

Part III: Effect of emotion-based feature augmentation on fake health news detection performance

While the results in the previous subsection highlighted the importance of using health-specific training data, we further investigated whether augmenting emotion-based features could provide additional performance gains. We tested the hypothesis that adding emotion-based node embeddings would improve fake health news detection performance compared to baseline (no augmentation). To this end, two versions of the dataset were used: HNDataset-BASE (baseline dataset) and HNDataset-Emotion, which included additional numerical features that quantified the emotional signals expressed in the text of each news post. The X-HND model was trained and tested on these two datasets. The model structure and parameters were the same, differing only in the dataset used.

We compared the performance of various models using different convolutional layers—Graph Convolution Network (GCN), Graph Sample and Aggregate (SAGE), and Graph Attention Networks (GAT)—on the BASE and Emotion versions of the HNDataset (**Table 2**) (15-17). The model trained on HNDataset-Emotion consistently outperformed the baseline dataset across all metrics. For instance, in terms of accuracy, the combination of GCN and HNDataset-Emotion achieved an accuracy of 0.80. In contrast, the GCN, SAGE, and GAT models trained on HNDataset-BASE each yielded an accuracy of 0.71. Regarding precision, the GAT + HNDataset-Emotion model reached 0.79, whereas the GAT + HNDataset-BASE model achieved only 0.65. As for recall, the GCN + HNDataset-Emotion combination attained 0.72, whereas recall values across the other model–dataset configurations ranged from 0.61 to 0.68. The combination of GCN + HNDataset-Emotion achieved the highest F1 score of 0.73, while GAT + HNDataset-BASE performed the worst (F1 = 0.63). Performance across different convolutional layers remained relatively stable, suggesting that dataset features contributed more to model performance than the choice of GNN layer. Overall, these results confirmed that emotion-based embeddings enhanced model performance. Moreover, differences in GNN layer types had minimal impact relative to dataset quality. The best performance was achieved with GCN + HNDataset-Emotion.

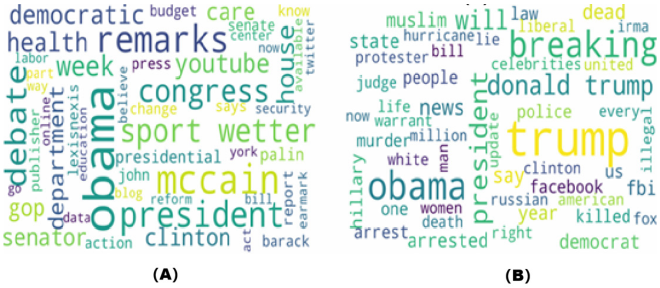
(A) (B)

Figure 3. Word clouds of true and fake news in the Politifact dataset. Word clouds were generated from the root news content in the Politifact dataset. The most frequent words found in (A) true political news and (B) fake political news. Compared to the Health-News dataset (Figure 2), Politifact exhibits distinct vocabulary trends, underscoring the differences in semantic characteristics across domains.

A two-sample independent *t*-test was conducted to compare the detection accuracies of the model trained on the baseline HNDataset and the emotion-enriched HNDataset. The mean accuracy for HNDataset-BASE was 0.71 (SD=0.0190), while for HNDataset-Emotion, it was 0.80 (SD=0.0178). The difference was statistically significant (t = 25.99, p < 0.001), supporting the hypothesis that emotion-based features improved model performance (**Figure 5**).

DISCUSSION

This study focused on improving fake health news detection on social media through graph-based modeling and textual features. The results confirmed our hypotheses and showed that graph-based models, health-specific datasets, and emotion-based feature augmentation each contributed to improved detection performance.

First, we compared the effectiveness of graph-based and sequence-based models. We found progressive improvement from simpler to more complex architectures, with the graph-based model performing best overall (**Figure 1**). Text-based models are inherently limited in their ability to understand deep semantics. Their performance is constrained by the variability of language, context fragmentation, and lack of structural relationships. In contrast, graph-based models integrate heterogeneous elements—such as characters, events, time, and location—into a unified structure of nodes and edges (9, 10). This allows them to capture complex relationships, including causal, temporal, and hierarchical dependencies. Moreover, graph topological properties such as node degree, clustering coefficient, and shortest path provide valuable insights into propagation patterns and influence distribution. These structural cues enhance the model's capacity to detect fake news more effectively. Overall, our results validated the superiority of graph-based architectures in fake health news detection tasks.

Second, we examined the necessity of domain-specific

https://doi.org/10.59720/25-130

Dataset	Graphs (Fake)	Total Nodes	Total Edges	Average. Nodes Per Graph
HNDataset	244 (143)	2684	2440	11
Politifact	314 (157)	41054	40740	131

Table 1. Structural properties of the Health-News dataset (HNDataset) and the Politifact dataset. The structural properties of the HNDataset and the Politifact dataset, including the number of graphs, total nodes, total edges, and average number of nodes per graph. The HNDataset contains 244 graphs focused on health misinformation, while the Politifact dataset contains 314 graphs centered on political misinformation.

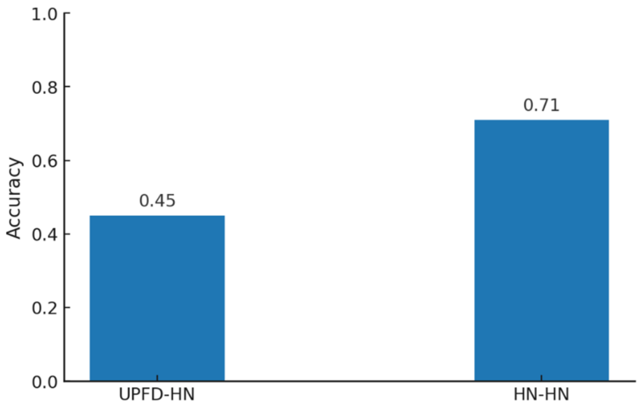

Figure 4. Training on a domain-specific health dataset improves fake news detection accuracy. Detection accuracy of the X-HND model under two different training settings. The model was evaluated under two conditions: (i) trained on the general-purpose UPFD Politifact dataset and tested on the domain-specific Health-News dataset (HNDataset-BASE), and (ii) trained and tested on HNDataset-BASE. The model trained on HNDataset-BASE achieved higher accuracy (0.71) compared to the model trained on the Politifact dataset (0.45).

dataset construction. Deep learning models trained on the UPFD Politifact dataset exhibited poor generalization when evaluated on HNDataset-BASE, with test accuracy remaining below 50%, despite the two datasets sharing similar graph structures and node feature dimensions (**Figure 4**). In contrast, models trained directly on HNDataset-BASE demonstrated effective learning, achieving higher accuracy. These results highlight the necessity of constructing a domain-specific dataset for detecting fake health news. Existing general-purpose datasets, such as Politifact, fail to capture the domain-specific semantic patterns present in health-related misinformation. This performance gap is likely attributable to content differences between the datasets, as visualized in the word clouds of root news (i.e., the original source posts of each news propagation graph) in HNDataset and Politifact (**Figure 2, 3**). Word clouds were generated by computing word frequencies from the root news texts and visualizing words proportionally to their frequency.

Finally, we explored how dataset enrichment strategies influence model performance. Models trained with HNDataset-Emotion consistently outperformed the baseline dataset across all metrics (**Table 2**). This result affirms that emotion embeddings enhance model performance in fake news detection. In addition to dataset variations, we compared the performance of different GNN layers: GCN, SAGE, and GAT. Results showed that neither SAGE nor GAT consistently outperformed GCN. In most cases, GCN-based models yielded the best performance. This may be attributed to the relatively small size and low feature diversity of graphs in HNDataset, improving the efficacy of simpler convolutional architectures like GCN. SAGE, which relies on neighbor sampling, may be less effective in small-scale graphs, while GAT layers, which require edge features to fully leverage attention mechanisms, are constrained by the lack of edge features in the current dataset. We also conducted an error analysis of the GCN-based model trained on two variants of the HNDataset: HNDataset-BASE and HNDataset-Emotion. When trained on HNDataset-BASE, the model misclassified a relatively large portion of true posts as fake, indicating a bias toward the fake label. In contrast, the model trained on HNDataset-Emotion achieved higher accuracy, revealing that the inclusion of emotional features reduced confusion between true and fake instances.

In summary, our experimental results demonstrate that (i) graph-based architectures outperform sequence-based models in detecting fake health news, (ii) emotion-based feature enrichment provides improvements in detection accuracy and robustness, and (iii) dataset construction and preprocessing quality play a crucial role in model performance, often more so than the complexity of the GNN architecture itself. Our results provided strong empirical support for the hypothesis that emotion-based feature augmentation could enhance the accuracy of fake health news detection. Statistical analysis further confirmed that the improvements attributed to emotion feature integration were highly significant ($p < 0.001$). This finding aligns with existing research that emphasizes the important role of emotional language in the virality and credibility perception of fake news (8). Emotional cues, such as expressions of fear, anger, or joy, often amplify the reach and perceived authenticity of misinformation, making them crucial signals for automated detection systems.

These findings offer practical implications for future research in detection of misinformation. Specifically, they suggest that building datasets tailored to a specific domain and using richer text features, such as emotion-related information, can improve fake news detection capabilities. Overall, these findings suggest that integrating emotional understanding into graph neural network architectures is a promising direction for future work in misinformation detection, particularly in domains where public trust and emotional engagement are critical, such as health news. Our results also show that simpler models, when supported by

Model Layer Type	GCN		SAGE		GAT	
HNDataset Type	BASE	Emotion	BASE	Emotion	BASE	Emotion
Accuracy	0.71	0.80	0.71	0.76	0.71	0.78
Precision	0.69	0.75	0.71	0.71	0.65	0.79
Recall	0.61	0.72	0.64	0.68	0.61	0.66
F1	0.65	0.73	0.66	0.69	0.63	0.72

Table 2. Performance metrics of the X-Health News Detection (X-HND) model across different dataset variants and graph convolutional layers. The accuracy, precision, recall, and F1 scores of the X-HND model trained on two different Health-News Dataset (HNDataset) types (BASE or Emotion), using three different graph convolutional layers (GCN, SAGE, or GAT). The model trained on the emotion-enriched dataset (HNDataset-Emotion) consistently achieved the highest performance across all metrics. (Note: Accuracy is the average of the last 50 epochs; precision, recall, and F1-score are calculated from the final confusion matrix).

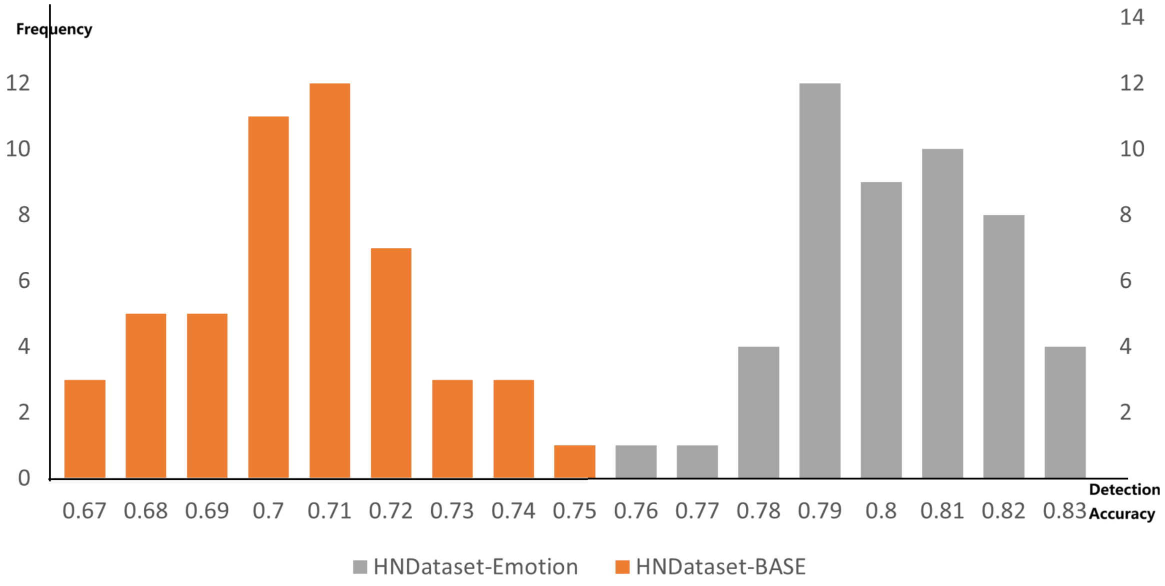

Figure 5. Detection accuracies of the X-HND model trained on HNDataset-BASE or HNDataset-Emotion. Distribution of detection accuracies for the X-HND model when trained and evaluated on two different datasets: HNDataset-BASE (orange) and HNDataset-Emotion (grey). Each model was trained and tested 50 times to obtain a robust distribution of accuracy values and sufficient sample size for statistical *t*-tests. The accuracy scores from HNDataset-BASE fall between 0.67 and 0.75, with a peak around 0.70–0.72. In contrast, the scores from HNDataset-Emotion are shifted to the right, with a peak between 0.79 and 0.82. The distinct separation in these distributions reflects a statistically significant difference in means ($t = 25.99$, $p < 0.0001$).

high-quality data, can achieve competitive performance in detecting health misinformation.

MATERIALS AND METHODS

Health-News dataset (HNDataset)

The HNDataset was specifically constructed for this study to capture both content and contextual propagation of health-related information on X. The dataset consisted of 244 graph-based samples, each representing a root health-related post and its corresponding reposting users and their historical activity. Data collection was conducted through the X API v2 using the Python Tweepy library, while root news was collected manually to ensure the quality of dataset. Specifically, root posts were required to focus on health-related topics such as medical treatments, public health issues, or health advice, contain factual or claim-based content rather than purely personal opinions, and have been reposted at least ten times to ensure sufficient propagation information. For each selected root post, ten reposting user IDs were randomly collected. Then, the most recent ten posts or reposts from each reposting user were retrieved to reflect their historical behavior on the social platform. The collected data were stored in the form of dictionary. All text data were tokenized with the Bidirectional Encoder Representations from Transformers (BERT) tokenizer and then embedded into 768-dimensional vectors using the pre-trained BERT model (18,19). Due to BERT's constraint on input length, any text exceeding 512 tokens was truncated. The 10 BERT vectors corresponding to each reposting user's posts were averaged to produce a single 768-dimensional vector representing the user's post history. Each graph consisted of 11 nodes: 1 central node representing the root post and 10 peripheral

https://doi.org/10.59720/25-130

nodes representing the users who reposted. All peripheral nodes were directly connected to the central node, forming a star-shaped graph. Data labeling was primarily conducted using GPT-4 with a structured prompt, followed by manual verification for quality control. The prompt evaluated content based on five prioritized criteria: authenticity (fact-checking), logical reasoning, emotionality (rational vs. irrational expression), reasoning complexity, and post length. News labeled "undetermined" were subsequently re-evaluated through close manual inspection and assigned a final label of either true or fake. Among the total of 244 samples used, 143 (58.6%) samples were labeled as fake news.

Two variants of this dataset were developed. The first variant was HNDataset-BASE. In this version, each node was represented solely by its 768-dimensional BERT embedding. This version served as a baseline for evaluating the contribution of additional features. The second variant was HNDataset-Emotion. To investigate the role of emotion in the spread of misinformation, we developed this emotion-aware variant of the dataset. For each root post, we used the Roberta-base-go_emotions model, a pretrained emotion classification model available via Hugging Face, to extract a 768-dimensional emotion embedding from the final hidden layer (20). Emotion embedding was employed as numerical feature to capture emotional signals for fake health news detection. This vector was concatenated with the original semantic BERT content embedding to form a 1536-dimensional vector, which was used as the node feature. No edge attributes were used in this version.

Politifact dataset

To provide a comparative benchmark, we employed the Politifact dataset, a component of the UPFD benchmark, available through the PyTorch Geometric library (12). Each graph in this dataset represented a political post along with user interaction data, including reposting user identity, user history, and post propagation. We used only the "BERT" variant of this dataset, which contained 768-dimensional BERT embeddings for both root posts and user history. A total of 314 graphs were used, among which 157 were labeled as fake news (**Table 1**).

Model structure

We designed a graph neural network model (X-HND) to classify fake and true news based on the graph structures described above. The core model comprised two graph convolution layers (GCN, SAGE, or GAT) for datasets with 768-dimensional node features. For the HNDataset-Emotion with 1536-dimensional node features (HNDataset-Emotion), an additional layer was added to facilitate effective dimensionality reduction. After convolution, a global mean pooling operation was applied to aggregate node features into a single graph-level embedding. It was followed by a dropout layer, then a linear layer, then another dropout layer and another linear layer. The dropout layer normalized the model to prevent overfitting, and the linear layer reduced the dimensionality of the data to two. Finally, a Softmax layer converted the two-dimensional data to possibility (**Figure 6**). HNDataset and UPFD Politifact dataset were used in this research as train datasets for this model. The X-HND

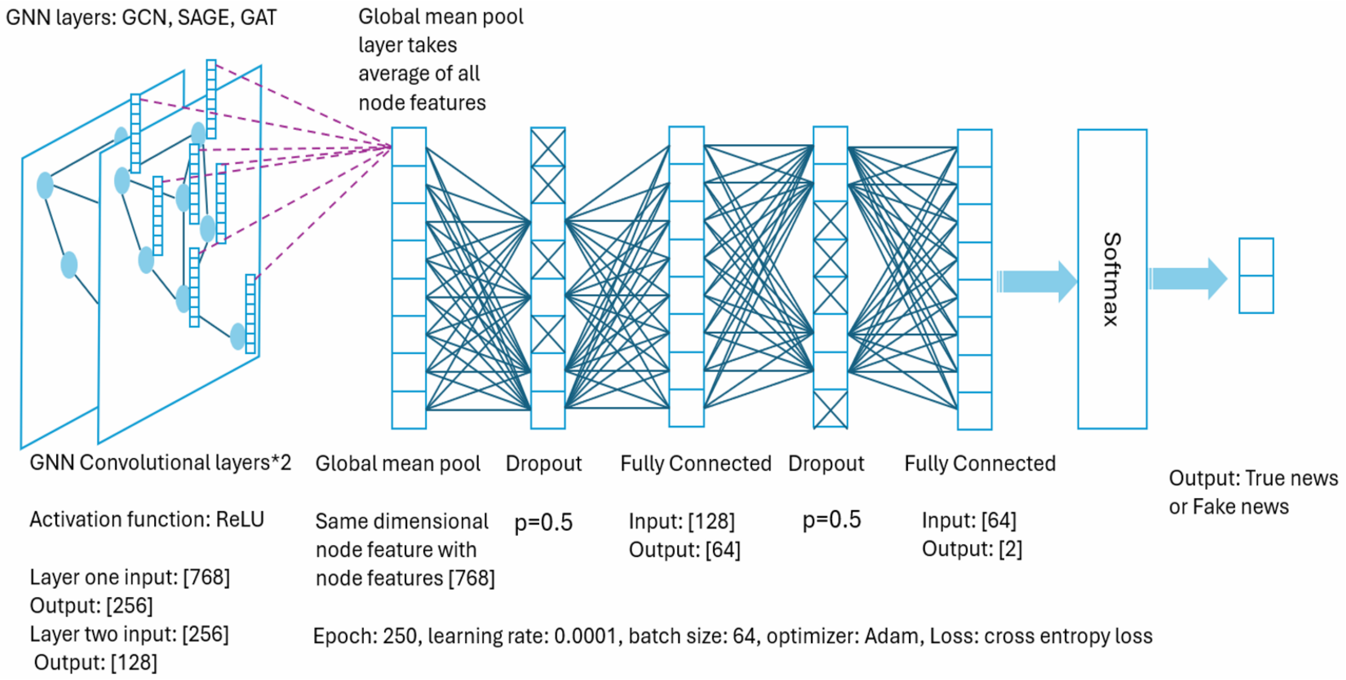

Figure 6. Architecture of the X-Health news detection (X-HND) model. Structure of the X-HND model used for health fake news detection. The model consists of two graph convolution layers (GCN, SAGE, or GAT), followed by a global mean pooling layer, two fully connected (linear) layers with dropout regularization, and a final Softmax output layer for binary classification. For datasets with emotion-enriched 1536-dimensional node features (HNDataset-Emotion), an additional dimensionality reduction layer was incorporated before graph convolution to accommodate the larger input feature size.

implementation, along with the HNDataset and raw news data, is available in a GitHub repository (13).

Model training strategies

The hyperparameters were chosen based on a combination of established best practices and preliminary experimentation. The model was trained using the Adam optimizer with a learning rate of 0.0001(21). The batch size was set to 64, and training proceeded for 250 epochs. A dropout rate of 0.5 was applied between layers to prevent overfitting, and the rest of the parameters were set default ($\beta 1=0.9$, $\beta 2=0.999$, $\varepsilon=1*10^{-8),}$ where $\beta 1$ and $\beta 2$ were exponential decay rates for estimating the first and second moments of the gradients, respectively, and ε was a small constant added for numerical stability to prevent division by zero during parameter updates. The activation function used throughout the model was ReLU (22). The training was conducted in Python 3.8.19 using standard PyTorch libraries.

Performance evaluation

To evaluate the model's classification performance, we adopted four standard metrics: accuracy, precision, recall, and F1 score (23). Accuracy represented the proportion of all correctly predicted instances relative to the total number of predictions. Precision assessed the proportion of true positive predictions among all positive predictions made by the model, while recall measured the proportion of actual positive cases that were correctly identified. These metrics were calculated using the following equations:

$$Accuracy = (tp + tn)/(tp + tn + fp + fn) \qquad \text{(Equation 1)}$$

$$Precision = tp/(tp + fp) \qquad \text{(Equation 2)}$$

$$Recall = tp/(tp + fn) \qquad \text{(Equation 3)}$$

F1 Score was the harmonic mean of precision and recall, calculated as:

$$F1 = 2 \cdot (precision \cdot recall)/(precision + recall) \qquad \text{(Equation 4)}$$

In these equations, *tp* (true positives) referred to the number of instances that were correctly predicted as positive by the model. *tn* (true negatives) represented the number of instances that were correctly identified as negative. *fp* (false positives) denoted the number of negative instances that were incorrectly predicted as positive, and *fn* (false negatives) indicated the number of positive instances that were incorrectly classified as negative. These four metrics ranged from 0 to 1, with higher values indicating better performance.

Received: April 28, 2025
Accepted: May 30, 2026
Published: July 03, 2026

REFERENCES

1. Ognyanova, Katherine, et al. "Misinformation in Action: Fake News Exposure Is Linked to Lower Trust in Media, Higher Trust in Government When Your Side Is in Power." *Harvard Kennedy School Misinformation Review*, June 2020, https://doi.org/10.37016/mr-2020-024
2. Newman, Nic, et al. "Reuters Institute Digital News Report 2024." *Reuters Institute*. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2024. Accessed 15 Jan. 2025.
3. Zhang, Xichen, and Ali A. Ghorbani. "An Overview of Online Fake News: Characterization, Detection, and Discussion." *Information Processing & Management*, vol. 57, no. 2, Mar. 2020, https://doi.org/10.1016/j.ipm.2019.03.004.
4. Chesney, Bobby, and Danielle Citron. "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security." *California Law Review* 107, no. 6 (2019): 1753–820. https://www.jstor.org/stable/26891938.
5. Shao, Chengcheng, et al.. "The Spread of Low-Credibility Content by Social Bots." *Nature Communications* 9 (2018): 4787. https://doi.org/10.1038/s41467-018-06930-7.
6. Shu, Kai, et al. "Fake News Detection on Social Media: A Data Mining Perspective." *SIGKDD Explorations*, vol. 19, no. 1, Sept. 2017, pp. 22–36, https://doi.org/10.1145/3137597.3137600.
7. Rishabh Upadhyay, et al.. "Leveraging Socio-contextual Information in BERT for Fake Health News Detection in Social Media." *Proceedings of the 3rd International Workshop on Open Challenges in Online Social Networks (OASIS '23)*, 4–8 Sept. 2023, Rome, Italy. ACM, 2023. https://doi.org/10.1145/3599696.3612902.
8. Farhoudinia, Bahareh, et al.. "Emotions Unveiled: Detecting COVID-19 Fake News on Social Media." *Humanities and Social Sciences Communications*, vol. 11, no. 1, May 2024, p. 640, https://doi.org/10.1057/s41599-024-03083-5.
9. Phan, H.T., et al. *"Fake News Detection: A Survey of Graph Neural Network Methods." Applied Soft Computing*, vol. 139, 2023, https://doi.org/10.1016/j.asoc.2023.110235.
10. Min, Erxue, et al. "Divide-and-Conquer: Post-User Interaction Network for Fake News Detection on Social Media." *Proceedings of the ACM Web Conference 2022*, Apr. 2022, pp. 1148–1158, https://doi.org/10.1145/3485447.3512163.
11. Patel, Anantram, and Vijay Kumar Sutrakar. "Advanced Text Analytics—Graph Neural Network for Fake News Detection in Social Media." *6th International Conference on Control, Communication and Computing (ICCC)*, India, 2025, https://ieeexplore.ieee.org/document/11077205.
12. Dou, Yingtong, et al. "User Preference-Aware Fake News Detection." *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 11 July 2021, ACM, Virtual Event, Canada, pp. 2051–2055. *ACM Digital Library*, https://doi.org/10.1145/3404835.3462990. Dataset available at: https://github.com/safe-graph/GNN-FakeNews.
13. Wang, Yiheng. *Fake-Health-News-Detection*. Version 1.0.0, 2025, https://github.com/frankyw2009/Fake-Health-News-Detection. Accessed 20 Nov. 2025.
14. James, Gareth, Daniela Witten, Trevor Hastie, and

Robert Tibshirani**.** *An Introduction to Statistical Learning with Applications in R.* Springer, 2013.

15. Hamilton, William L, et al. "Inductive Representation Learning on Large Graphs." *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, pp.1025–1035*,* https://dl.acm.org/doi/10.5555/3294771.3294869.
16. Kipf, Thomas N., and Max Welling. "Semi-Supervised Classification with Graph Convolutional Networks." *Proceedings of the International Conference on Learning Representations (ICLR),* 2017, https://openreview.net/pdf?id=SJU4ayYgl.
17. Veličković, Petar, et al. "Graph Attention Networks." *Proceedings of the International Conference on Learning Representations (ICLR)*, 2018, https://openreview.net/pdf?id=rJXMpikCZ.
18. Devlin, Jacob, et al. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, June 2019, pp. 4171–4186, https://doi.org/10.18653/v1/N19-1423.
19. Liu, Yinhan, et al. "RoBERTa: A Robustly Optimized BERT Pretraining Approach." *arXiv*, 2019, https://doi.org/10.48550/arXiv.1907.11692.
20. Lowe, Sam. "roberta-base-go_emotions." *Hugging Face*. https://huggingface.co/SamLowe/roberta-base-go_emotions. Accessed 10 Dec. 2024.
21. Kingma, Diederik P., and Jimmy Ba. "Adam: A Method for Stochastic Optimization." *Proceedings of the International Conference on Learning Representations (ICLR)*, 2015.
22. Glorot, Xavier, et al. "Deep Sparse Rectifier Neural Networks." *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2011, pp. 315–323.
23. Sokolova, Marina, and Guy Lapalme. "A Systematic Analysis of Performance Measures for Classification Tasks." *Information Processing & Management*, vol. 45, no. 4, 2009, pp. 427–437. https://doi.org/10.1016/j.ipm.2009.03.002.