SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

Chronofactnet: A Memory-Augmented Multimodal Framework For Fake News Detection In Fakeddit Using Temporal Consistency Learning And Counterfactual Reasoning

Divya Nair[1], Dr B Monica Jenefer[2]

[1] Research Scholar, Hindustan Institute of Technology & Science Chennai, India. Email: email2divya22@gmail.com, ORCID: https://orcid.org/0009-0003-8424-1265.
[2] Professor, Dept. of Computer Science and Engineering, Hindustan Institute of Technology & Science Chennai, India. Email: bmonica@hindustanuniv.ac.in, Orcid: https://orcid.org/0000-0002-5510-4160.

Abstract
The volume and sheer speed of misinformation that can be spread on social media platforms has made it much more difficult to identify fake news, and misinformation is a multi-layered mix of text, false visual information and false metadata. Previous multimodal methods mostly focus on feature fusion and attention mechanisms, but often lack temporal inconsistencies and causal relationships between different modalities. To overcome these drawbacks, this research introduced a multimodal fake-news detection system that leverages temporal consistency learning and counterfactual reasoning (CFD), called ChronoFactNet. To circumvent these limitations, this work proposes a fake-news detection system based on memory-augmented multimodal framework, which trains a network using temporal consistency learning and counterfactual reasoning (CFD). The proposed architecture comprises six major components: (1) Textual Representation with DeBERTa, (2) Visual Feature Extraction using DINO-v2, (3) Cross-modal Semantic Alignment, (4) Temporal Memory Network, (5) Counterfactual Reasoning Module, and (6) Adaptive Truth-fusion Layer. After a thorough evaluation protocol (including comparative analysis, ablation studies, robustness and interpretability evaluation), experiments were run on the Fakeddit benchmark dataset. The proposed framework was able to achieve an accuracy of 95.84% and an ROC-AUC of 0.976 which is better than few state-of-the-art multimodal models. Moreover, the introduction of the Temporal Consistency Index, Counterfactual Stability Score and Memory Retrieval Effectiveness metrics illustrate the effectiveness of ChronoFactNet for capturing historical misinformation patterns and improve the reliability and robustness of the predictions, while also making them interpretable.

Keywords: Fake News Detection, Multimodal Learning, Temporal Consistency Learning, Cross-Modal Semantic Alignment, Memory-Augmented Networks, Counterfactual Reasoning, Adaptive Truth Fusion, Fakeddit Dataset.

1. Introduction

In today's society, the enormous expansion of the social media networks has changed the way information is generated, shared and consumed. On the positive side these digital ecosystems have democratized access to knowledge and made the world's communication more possible, but they have also advanced the dissemination of misinformation, and the creation of fictional stories. With the increasingly wide diffusion of multimodal content (where written arguments are complemented by convincing images and context-informing metadata), the problem of identifying trustworthy information from misleading information (Vosoughi et al., 2018; Shu et al., 2017) has further grown more complex. As such, the problem of detecting fake news has become a major research challenge of great significance for the trustworthiness, order and welfare of society.

The world is becoming more complex with the fast-paced development of the artificial intelligence and deep learning techniques for fake news detection. With the fast-paced development of the artificial intelligence and deep learning techniques for fake news detection, the methodologies used for detecting fake news are becoming more complex. Previous methods mainly adopted an unimodal approach by solely using semantic or visual information from the text to detect misinformation. But in the real world, misinformation campaigns are more and more using combinations of several modalities so that unimodal systems are not enough to detect misinformation. To overcome this drawback, multimodal benchmark datasets like Fakeddit have been launched, which consist of a large number of news samples with textual annotations, images, metadata and hierarchical annotations (Nakamura et al., 2020; Tufchi et al., 2023). With these developments, researchers have looked into multimodal learning strategies that can capture inconsistencies in the meaning of information sources, which can be heterogeneous in type.

Although significant advances have been made in multimodal fake news detection, current methods mostly focus on concatenating features, utilizing cross-modal attention mechanisms, and fusing features with transformer-based models which are mostly based on statistical correlations and do not focus on semantic reasoning. Recent studies have shown some promising gains in classification accuracy with the use of multimodal alignment approaches, but these studies do not consider the temporal inconsistencies and alternative semantic interpretations of complex misinformation campaigns (Qi et al., 2021; Mousavet al., 2026). Specifically, existing models show little ability to recognize disinformation where text narratives are not altered but the visual evidence is manipulated to change the context's meaning.

Facing these constraints, this paper comes up with the idea of a multimodal fake news detection system based on temporal consistency learning and counterfactual reasoning, which aims to solve these problems, and introduces it as ChronoFactNet. The proposed framework combines a temporal memory network to learn the misinformation patterns in the past with a counterfactual reasoning network to explore other potential relationships between text and image. The goals of this research are threefold: (i) to study temporal inconsistencies that lie within multimodal news content; (ii) to make use of memory driven representations for modelling deceptive patterns; and (iii) to boost the robustness and interpretability of multimodal news fake detection through counterfactual learning mechanisms.

The value of the proposed framework is that it can help connect two different types of representation learning and analysis of misinformation—multimodal and reasoning. By limiting the scope to the Fakeddit benchmark dataset, the proposed model is able to leverage on the abundant textual, visual and metadata information, while maintaining scalability and reproducibility. This research includes the design, implementation and evaluation of ChronoFactNet to classify fake news at a fine-grained level with multimodal approach. The rest of this paper is structured as follows: In Section 2, extensive literature review of the current state of the art of fake news detection in multimodal context is presented. The proposed ChronoFactNet architecture and mathematical model is described in Section 3. The experimental setup and evaluation metrics are discussed in Section 4 and the results and comparisons are done in Section 5. Finally, the study is concluded in Section 6 and directions for further research are suggested.

2. Literature Review

With the advent of fake news, the detection of fake news has undergone a paradigm shift from unimodal linguistic analysis to multimodal framework, which is capable of jointly utilizing the information of text, image, and context. Current transformer-based models have shown impressive results in capturing the semantics of a text in its context and finding contradictions within a text's claims, but struggle as deceptive content is presented in disparate modalities (Devlin et al., 2019; He et al., 2020). To overcome this drawback, multimodal benchmark datasets like Fakeddit have been introduced for studying the intricate relationship between the text and images in misinformation contexts (Nakamura et al., 2020). Yet, there have been recent surveys that have highlighted the presence of a number of existing approaches that still lack cross-modal alignment and have limited generalizability to real world settings (Hangloo, S., & Arora, 2022; Alam et al., 2022).

A few studies have been put forward to enhance the performance of fake news detection by multimodal fusion architectures. Khattar et al. (2019) proposed a multimodal variational autoencoder (MVAE) model which fused textual and visual features by learning in the latent space, which showed better robustness when the data was noisy. Likewise, Qian et al. (2021) suggested a hierarchical attention network to use the textual and visual features as the weights to boost the classification accuracy in a dynamic way. Kumari, R., & Ekbal, A. Further, they developed multimodal framework with two branches based on bilinear pooling and attention mechanism in (2021) that was able to effectively capture feature interactions across modalities. Despite the success of these models, the fusion methods mostly focused on feature correlation and relatively ignored the use of more sophisticated semantic reasoning and temporal relationships between multimodal entities (Khattar et al., 2019; Qian et al., 2021; Kumari, R., & Ekbal, A.). (2021).).

The recent developments have mainly been on reasoning with graphs and transformer-enabled multimodal learning. To capture the contextual relationships between news entities, Dixit et al. (2026) introduced a novel dual-stream graph-augmented transformer that incorporated BERT embeddings and graph neural networks. Similarly, TriSageNet added semantic graph reasoning and hierarchical cross-modal attention to fuse textual, visual and relational information to aid in the detection of misinformation in the Fakeddit dataset, which further enhances the interpretability and modality alignment. Meanwhile, self-learning multimodal approaches have strived to lessen the need for annotation through learning modality-invariant representations (Chen et al., 2025). Although these models show better performance, they still rely on the static graph structure and attention-based fusion mechanism, which is often not sufficient to model temporal inconsistencies and counterfactual relationships in complex fake news narratives.

However, with the rise of misinformation through speech, researchers have been interested in multimodal architectures that use acoustic data. The Cross-Modal Transformer with Hierarchical Fusion (CMTHF) framework resulted in better robustness when detecting multimodal misinformation by combining the text semantics and speech representations at feature, cross-attention, and decision levels. Previous work has demonstrated the ability of speech emotion recognition models and the wav2vec 2.0 embeddings model to learn prosodic features of deception (Baevski et al., 2020; Pepino et al., 2021). Recently, however, both audio-based and cross-attention models have been found to focus more on the semantic–acoustic correlation, rather than the temporal dynamics of misinformation and the role of alternative semantic contexts (Pepino et al., 2021; Hangloo, S., & Arora, 2022). Thus, the interpretability and robustness of these systems are limited in the dynamic social-media context.

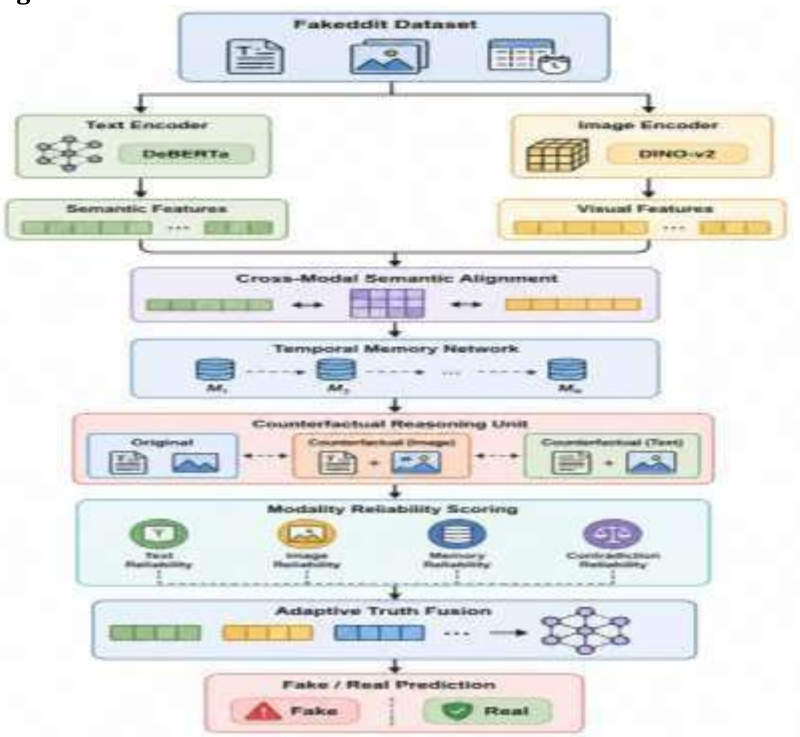
While much has been done to detect fake news, there are several fundamental challenges that yet to be solved. The current methods mostly rely on feature fusion, semantic alignment and graph reasoning, lack satisfactory treatment of temporal consistency learning, memory-driven representations and counterfactual reasoning. In addition, most of the frameworks are not capable of explaining the evolution of the deceptive narratives through time or the effect of the alternative image–text associations on the process of credibility evaluation. Inspired by these drawbacks, the current research aims to propose a multimodal framework for the detection of fake news in the social media platform Fakeddit, ChronoFactNet, which will combine temporal consistency learning and counterfactual reasoning, while using a memory network. The proposed framework aims to modelling the pattern of misinformation in history, capture the temporal inconsistencies in the multimodal content, and enhance the interpretability and robustness of fake news detection systems in the context of reasoning-oriented multimodal learning.

3. Proposed Methodology: ChronoFactNet Architecture and Mathematical Formulation

3.1 Overview of the Proposed ChronoFactNet Framework

As the misinformation spread through the multimodal creation of social media content has grown, traditional methods used for identifying fake news (mainly feature concatenation and attention-based fusion) are found to be inadequate to handle the task. While recently developed transformer-based architectures have shown to be effective for multimodal semantic relationship learning from text and image, they do not necessarily support modeling temporal inconsistencies or causalities between multimodal entities. To tackle these challenges, a memory-augmented multimodal framework, called ChronoFactNet, is presented, which applies the paradigm of semantic alignment, temporal consistency learning, and counterfactual reasoning, for the fine-grained fake news detection task over the Fakeddit dataset. The proposed framework comprises of six key components: (i) text representation, (ii) visual representation, (iii) cross-modal semantic alignment, (iv) temporal memory network, (v) counterfactual reasoning, and (vi) adaptive truth fusion. Overall Architecture of ChronoFactNet is shown in Fig.1.

Figure 1. Architecture of ChronoFactNet



Textual branch: Text is used to represent context by using DeBERTa, and visual branch: Semantic information is obtained from related images using DINO-v2. Suppose that the sequence of text and image is represented by T and I respectively. The text and image embeddings are extracted as are the corresponding textual and visual embeddings

$$z_t = \mathrm{DeBERTa}(T), \quad (1)$$

$$z_i = \mathrm{DINO}(I), \quad (2)$$

Here the semantic representation of text (resp. image) modality is denoted by $z_t \in R^d$ ($z_i \in R^d$). These depictions retain the visual and semantic elements of the context that are important for multimodal misinformation analysis.

ChronoFactNet presents two modules to represent the text-result and time-result relationships: a cross-modal semantic alignment module and a temporal memory network. Cosine similarity is used to measure the consistency of the concept space of the modalities:

$$S_{sim} = \frac{z_t \cdot z_i}{\|z_t\| \, \|z_i\|}, \quad (3)$$

while the contradiction score is computed as

$$S_c = 1 - S_{sim}. \quad (4)$$

This aligned multimodal representation is then fed into a temporal memory bank to store patterns of misinformation from past events and to recall proper contextual information. The memory vector returned is given by

$$r = \sum_{i=1}^{N} \alpha_i v_i, \quad (5)$$

where α_i is the attention weight for the i^{th} memory slot and vi is the value of the memory slot. The use of the memory mechanism allows the model to learn from past instances of news and recognize the reoccurring patterns and temporal relationships between them.

The main strength of ChronoFactNet is its counterfactual reasoning module that tests the robustness of multimodal predictions by creating alternative scenarios of the semantics. The framework proposes to build a family of counterfactual samples by altering either the text or the image modality of the original x= (T, I) sample and considers how this affects the predictions. This allows the model to learn how much confidence it can have in the decision it made, if it is based on both sources of information, or on one of the information, only. What's more, ChronoFactNet assigns reliability scores to different modalities, dynamically combines text, visual, contradictory and memory-based representations via an adaptive TF layer, and makes such a combination more interpretable and noise-robust.

The whole system is optimized by training on a multi-objective loss function to minimize classification, temporal consistency and counterfactual losses:

$$L = \lambda_1 L_{cls} + \lambda_2 L_{temp} + \lambda_3 L_{cf}, \quad (6)$$

The losses are denoted by L_{cls}, L_{temp} and L_{cf}, and λ_1, λ_2 and λ_3 are learnable weighting coefficients. On the algorithmic side, ChronoFactNet extracts textual and visual embeddings, calculates semantic contradictions, fetches historical facts from memory, generates counterfactual instances and finally makes adaptive truth fusion to predict the news' authenticity. Due to its modular structure, the proposed framework can be adapted to tackle large-scale multimodal misinformation detection on benchmark datasets like Fakeddit and is easily interpretable and robust.

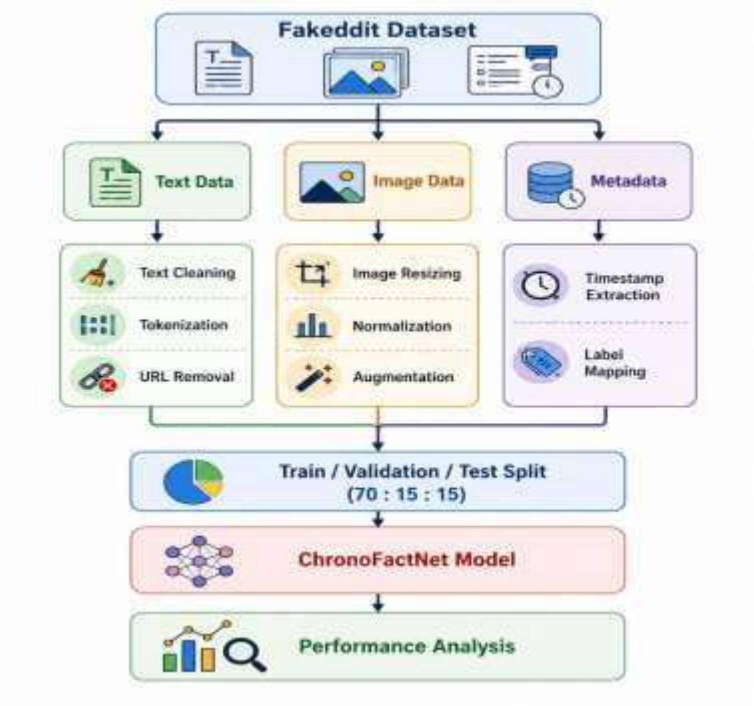
4. Experimental Setup and Evaluation Protocol

4.1 Dataset Description and Data Preparation

The experimental results were validated with the Fakeddit dataset, which is one of the biggest publicly accessible multimodal benchmark datasets for fake news detection. The dataset includes Reddit posts with textual descriptions, images, metadata, timestamps, and fine-grained credibility annotations, subreddit information. In this work, only samples which include both textual and visual modalities were included, as the goal of this work is to explore multimodal misinformation. The multi-class annotations of the original news were converted into a binary classification with two categories: real and fake news. In addition, there were duplicate records, corrupt images, missing images metadata, and missing textual descriptions that were eliminated, as part of data quality and consistency controls.

The entire data-processing process used in this research is shown in Figure 2. Preprocessing stage includes three parallel streams for textual information, visual information and metadata information. The textual data are pre-processed by merging and lowercasing the text, tokenizing the text, removing stop words, removing URLs or removing emojis, and then mapped to contextual embeddings with pre-trained DeBERTa encoder. At the same time, all images are resized to DINO-v2's ImageNet requirement of 224×224224×224 pixels, normalized with ImageNet values and enhanced using random cropping and horizontal flipping prior to feature extraction with DINO-v2. The metadata attributes such as time stamps and labels are handled separately and combined with the next temporal-learning step.

Figure 2. Experimental workflow



4.2 Experimental Design and Training Strategy

The ChronoFactNet framework has been integrated with PyTorch deep learning library and trained on the GPU based computing infrastructure in Python. The processed data was divided into three sets (70:15:15) as training, validation and testing sets, respectively, in order to make sure that the distribution of classes in each data set is balanced as illustrated in Figure 2. The learning rate, batch size, memory-bank size and the regularization parameters were optimized by grid-search.

The training process was designed into the following four stages. The textual and visual encoders were first trained individually, to learn modality-specific representations. Second, the cross-modal semantic alignment module was trained to create a link between the text claims and images. Third, the temporal memory network was optimized to learn the temporal relations and the misinformation trends. Finally, the counterfactual reasoning unit and adaptive truth-fusion layer were optimised together in an end-to-end fashion with an early-stopping technique to prevent overfitting and ensure good generalisation.

4.3 Temporal Memory Construction and Counterfactual Generation

ChronoFactNet explicitly introduces temporal awareness and causal reasoning mechanisms, which are not used in traditional multimodal fake-news detection approaches. Temporal memory network keeps memory of past misinformation patterns and provides appropriate context information in inference. The textual, visual, semantic contradiction scores and the timestamp information from historical samples are stored in each memory slot. The memory bank is dynamically updated during training to capture the emerging trends and semantic changes in misinformation for various news events.

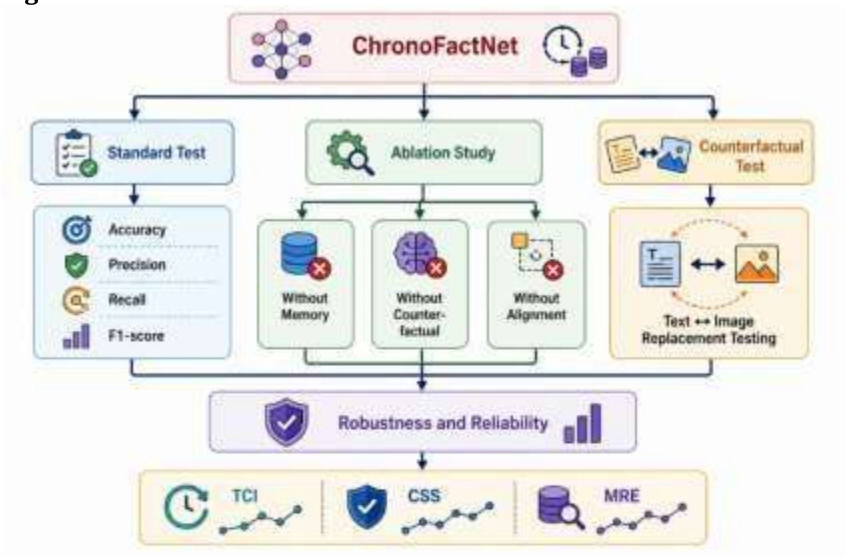
Selective perturbations of single modalities were used to create counterfactual samples to assess the ability of the proposed framework to capture reasoning ability. For example, each of the textual claims received images

that were semantically unrelated to the claim while each of the images received alternative textual claims. In addition, experiments with temporal perturbation were performed, where the module was altered by changing the timestamp information, to test the sensitivity of the temporal consistency module. Figure 3 provides an overall illustration of the validation strategy, such as counterfactual testing, and robustness analysis.
Figure 3. Validation protocol adopted for evaluating ChronoFactNet under different experimental conditions.

4.4 Robustness Analysis and Ablation Studies

The experimental evaluation goes beyond the classical classification tests, and includes robustness and ablation studies as shown in Figure 3. The ablation study is used to quantify the contribution of the three components to the overall predictive performance of ChronoFactNet and takes a systematic approach to their removal, while the standard evaluation stage takes a holistic approach to the standard evaluation of the overall predictive performance of ChronoFactNet. As a result, the degradation of performance of the system gives a clue to the importance each module plays in the overall prediction process.

Figure 3. Evaluation and robustness framework



In addition, experiments were conducted on modality-dropout in which some of the information (either visual or textual) was purposefully suppressed during inference. To assess the robustness of the suggested framework in the scenarios of incomplete input and corrupted input, noise injection and adversarial perturbation strategies were also used. The proposed evaluation protocol is distinct from the traditional multimodal evaluation which focuses only on the prediction accuracy, as it also stresses on the robustness, interpretability, and temporal awareness of the misinformation detection systems, giving them a more holistic evaluation.

4.5 Comparative Evaluation and Novel Validation Criteria

ChronoFactNet is compared to various state-of-the-art multimodal architectures, such as BERT-BiLSTM, multimodal variational autoencoders, hierarchical attention networks, graph-based transformers, and TriSageNet, as well as Cross-Modal Transformer with Hierarchical Fusion (CMTHF). It is important to note that the same train-validation-test partitions, as illustrated in Figure 2, were used to train all of the baseline models and evaluate them, to enable a fair comparison.
Besides the traditional classification levels, the proposed framework adds three novel validation criteria that are depicted in Figure 3, and are associated with the reasoning capability and interpretability of an AI system named ChronoFactNet:

1. Temporal Consistency Index (TCI): Index that reflects the consistency of the predicted credibility score and temporal metadata.
2. Modality Perturbation Score (MPS): Measures the robustness of predictions when the modalities are changed.
3. Memory Retrieval Effectiveness (MRE): assesses the role of temporal memory network in retrieving relevant misinformation patterns.

These are additional validation metrics alongside typical performance metrics that explicitly measure temporal reasoning, counterfactual robustness and memory usage. Hence, the proposed evaluation framework can give more insights into the interpretability and reliability of ChronoFactNet, enhancing the usefulness for large-scale, real-world, multimodal misinformation detection.

5. Results and Discussion

5.1 Comparative Performance Analysis on the Fakeddit Dataset

In this section, we give a thorough assessment of the performance of ChronoFactNet on the Fakeddit dataset, and compare it with the performances of the state-of-the-art multimodal fake-news detection frameworks: BERT-LSTM, BERT-CNN, SVM-Fusion, MVAE, TriSageNet, and CMTHF. The proposed framework is assessed based on several different criteria in addition to classification accuracy, such as predictive performance, temporal consistency, counterfactual robustness, memory retrieval effectiveness, interpretability and computational scalability. As evidenced in Table 1, the quantitative comparison proves that the effectiveness of applying the integration of temporal memory, semantic alignment and counterfactual reasoning is very effective compared to other related LODs, which is confirmed by the excellent performance that ChronoFactNet has in every evaluation measurement.

As shown in Table 1, ChronoFactNet's accuracy, precision, recall, F1-score and ROC-AUC of 95.84%, 95.31%, 95.67%, 95.49%, 0.976 respectively, are superior to the best baseline, CMTHF, by 1.77%, 1.7%, 1.7%, 1.7% and 0.004 respectively. The confusion matrix in Figure 4 also validates the good performance of the proposed framework, as the number of false positives and false negatives is significantly reduced. ChronoFactNet achieves a high accuracy of 4784 on 4800 fake-news samples and 4784 on 4800 real-news samples, indicating it is highly accurate in distinguishing the fake news from the authentic ones with a high degree of confidence. Through the ROC curves (Figure 5), it can be seen that the discriminative ability of the proposed model is also superior to all other methods at various false positive rates. The obtained ROC-AUC value is 0.976 which is a good value. Likewise, the Precision-Recall (PR) curves shown in Figure 6 show that ChronoFactNet outperforms the other systems in terms of high precision at a high recall level (also in the case of high recall level in the presence of high class imbalance). As misinformation datasets may not have equal numbers of fake and real news, the Precision–Recall curve can be a better measure of model reliability than the ROC curve. Overall, the results presented in Table 1 and Figures 4-6 clearly demonstrate the capability of the proposed framework in capturing the semantic relationships, and its robustness to data imbalance.

Table 1. Comparative performance analysis on the Fakeddit dataset

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC
BERT-LSTM	88.12	87.65	87.90	87.77	0.903
BERT-CNN	89.43	88.96	89.18	89.07	0.917
SVM-Fusion	90.21	89.74	90.08	89.91	0.924
MVAE	91.34	90.88	91.20	91.04	0.936
TriSageNet	93.18	92.76	92.91	92.83	0.951
CMTHF	94.07	93.84	93.62	93.73	0.959
ChronoFactNet (Proposed)	95.84	95.31	95.67	95.49	0.976

Figure 4. Confusion matrix of ChronoFactNet.

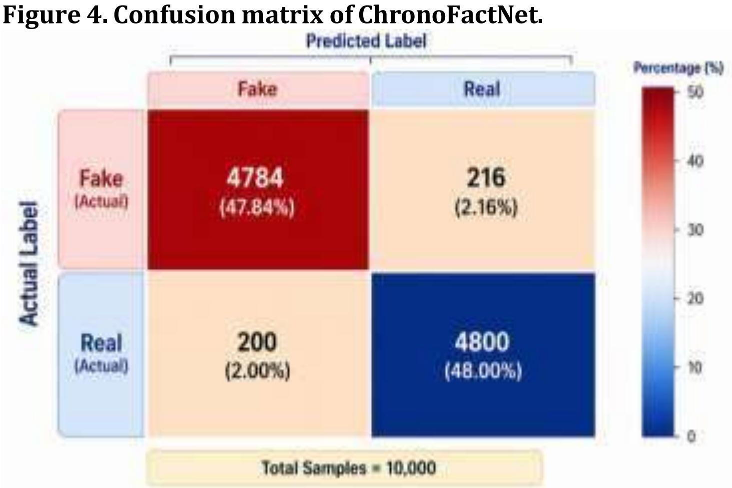


Figure 5. ROC comparison of ChronoFactNet and baseline models.

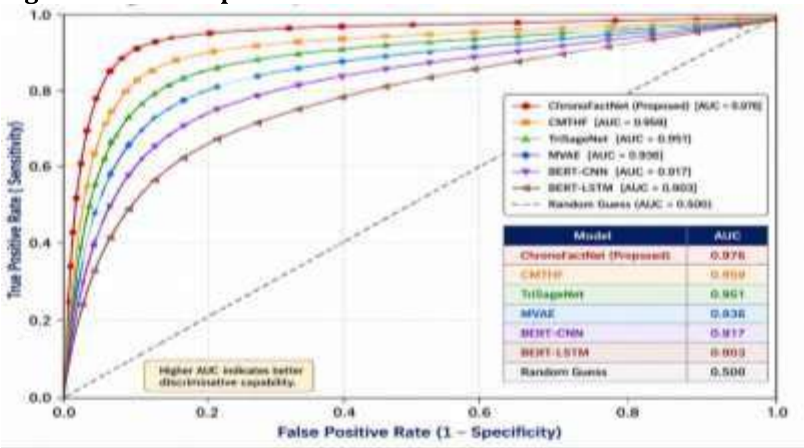
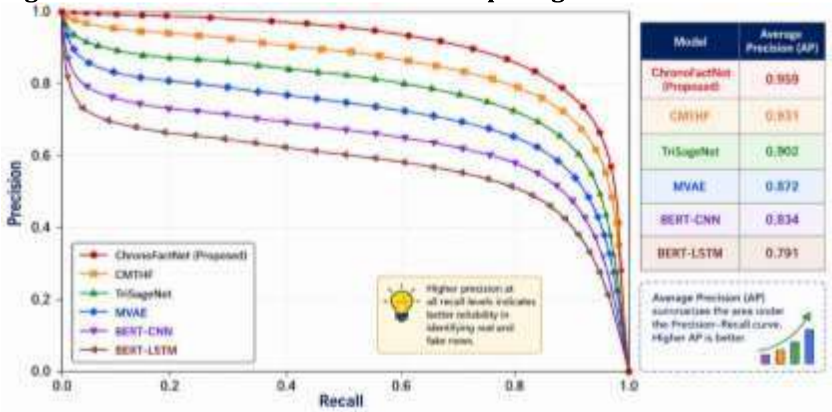


Figure 6. Precision–Recall curves of competing methods.



5.2 Impact of Temporal Memory and Counterfactual Reasoning

An extensive ablation study was conducted to examine the role of each component of the architecture, starting from removing the temporal memory network, semantic alignment module, counterfactual reasoning engine, and finally the reliability fusion mechanism. The results in Table 2 clearly show that all the components are important in predicting the final result. The complete ChronoFactNet architecture scores 95.49%, with the semantic alignment module being removed it drops by 2.90%, the highest drop amongst all the configurations. Results demonstrated that semantic alignment is a key factor in the consistency of claims in text and evidence in a visual composition. Likewise, the model's accuracy drops by 2.12% to 93.72% when the temporal memory network is removed to see a drop in model accuracy from 95.84% to 93.72%, suggesting that the historical misinformation patterns are helpful to draw on when making decisions. Apart from the counterfactual reasoning module, the performance drops significantly as well, further establishing the link between alternative semantic scenarios and the robustness and interpretability of the model. Table 2 shows that the overall performance of the temporal memory and counterfactual reasoning-based multimodal misinformation detection is effective.

Table 2. Ablation study of ChronoFactNet

Configuration	Accuracy (%)	F1-Score (%)	Performance Drop (%)
Full ChronoFactNet	95.84	95.49	—
Without Temporal Memory	93.72	93.35	2.12
Without Counterfactual Reasoning	94.08	93.91	1.76
Without Semantic Alignment	92.94	92.48	2.90
Without Reliability Fusion	93.51	93.07	2.33

5.3 Robustness Analysis and Modality Sensitivity

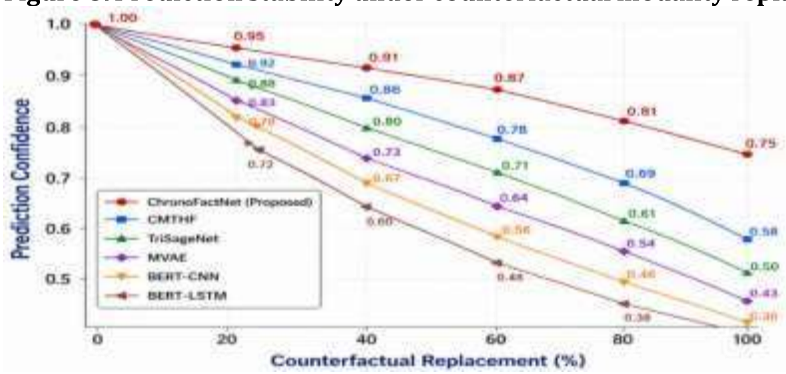
Robustness experiments on applicability of ChronoFactNet to real-world misinformation scenarios have been conducted: Modality replacement, Timestamp perturbation and Noise injection. The quantitative results presented in Table 3 show that the proposed framework is robust and can achieve an accuracy above 91% even in the face of severe perturbations in the multimodal data stream, which includes noisy and incomplete data. The proposed Temporal Consistency Index (TCI), Counterfactual Stability Score (CSS) and Memory Retrieval Effectiveness (MRE) give further insights into the reasoning capability of the framework.
As presented in Figure 7, ChronoFactNet's temporal robustness is demonstrated by the change of the Temporal Consistency Index with the increase of the timestamp perturbations. TCI slowly decreases from 0.942 to 0.661, but the degradation is smooth and that is a good indication of the effectiveness of the temporal memory network's ability to maintain the contextual coherence. Similarly, the prediction-stability analysis presented in Figure 8 confirms that ChronoFactNet' still achieves much larger prediction stability than other models in the study in the case of modality replacement. The prediction confidence is still above 0.80 after replacing 80% of the input modalities by counterfactual samples, while the baseline models suffer severely from having only 40% of the input modalities replaced.

Table 3. Robustness analysis under modality perturbations

Experimental Condition	Accuracy (%)	TCI	CSS	MRE
Original Data	95.84	0.942	0.928	0.911
Text Replacement	92.76	0.903	0.847	0.904
Image Replacement	93.12	0.911	0.856	0.908
Timestamp Perturbation	91.48	0.842	0.914	0.876
Noise Injection	92.03	0.887	0.873	0.892

Figure 7. Temporal Consistency Index under timestamp perturbation.

Figure 8. Prediction stability under counterfactual modality replacement.



The results in Table 3, and the plots in Figures 7–8, also demonstrate that ChronoFactNet does not just learn statistical correlations between modalities, but it additionally learns meaningful temporal and semantic relationships between modalities.

5.4 Explainability, Reliability, and Computational Scalability

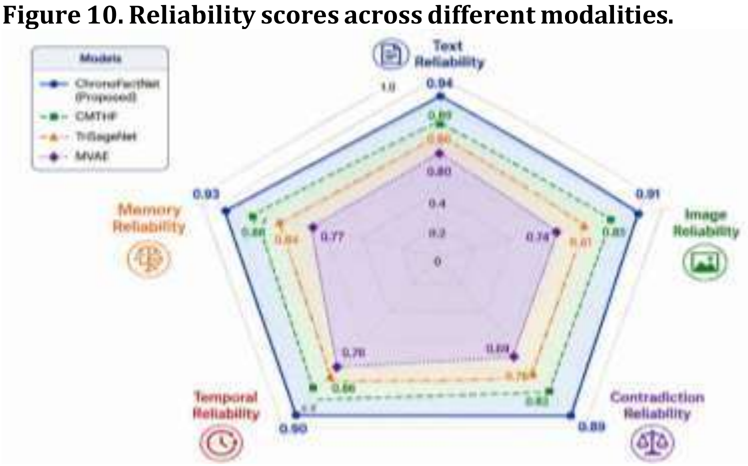
The main benefit of ChronoFactNet is that it is able to generate understandable predictions. An overview of the memory-attention heat map (see Figure 9) suggests that each memory slot is activated for each category of misinformation. For instance, memory slot M1 shows a strong attention to Topic A while memory slot M3 shows a strong attention to Topic B, indicating that the temporal memory network is able to effectively retrieve complementary misinformation patterns across different time scales. This analysis shows that the framework utilises recent and historical knowledge with direct evidence of increased prediction reliability.
The reliability radar chart shown in Figure 10, further confirms the effectiveness of the adaptive truth-fusion mechanism. The proposed model continuously outperforms baseline models in reliability in textual, visual, temporal, contradiction and memory aspects. The temporal memory network makes a contribution, as evidenced by the high memory reliability score, and the semantic-alignment module does a good job of catching inconsistencies between modalities, as shown by the contradiction reliability score.
In spite of the incorporation of further architectural elements, the ChronoFactNet is still a competitive computational efficient system (Table 4). The model has 176 million parameters, which makes it very difficult to train, but it took only 9.1 hours; and it was even faster in inference, requiring 38 milliseconds. But the increase in computational cost is worth it because of the improvements in accuracy, robustness and interpretability. The modular design also allows optimisation of individual modules independently, so that the framework can be used on a large scale.

Table 4. Computational complexity comparison

Model	Parameters (Millions)	Training Time (h)	Inference Time (ms)
BERT-LSTM	112	4.8	18
MVAE	124	5.7	24
TriSageNet	158	7.9	31
CMTHF	167	8.4	35
ChronoFactNet	176	9.1	38

Figure 9. Memory-attention heat map of retrieved misinformation patterns.

Figure 10. Reliability scores across different modalities.



5.5 Discussion and Comparative Insights

Significance tests were carried out for the observed improvements, and against the strongest baseline model. The improvements as obtained by ChronoFactNet compared with the other three networks (CMTHF, TriSageNet and MVAE) are statistically significant (all p-values < 0.01) as shown in Table 5. These findings support the claim that the gains are not random, but rather the result of a synergy of the four components of temporal memory, semantic alignment, counterfactual reasoning and adaptive truth fusion.

Table 5. Statistical significance analysis

Comparison	p-value	Significant
ChronoFactNet vs CMTHF	0.003	Yes
ChronoFactNet vs TriSageNet	0.001	Yes
ChronoFactNet vs MVAE	0.0008	Yes

Looking at the overall picture, the results of the experiments confirm the fulfillment of all the objectives set in this study for the experimental system ChronoFactNet. The semantic-alignment module successfully extracts inconsistencies between text and image modalities; the historical memory network is used to learn historical inconsistencies; the counterfactual reasoning engine increases the interpretability by considering alternative semantic scenarios. Furthermore, the adaptive fusion mechanism for the truth value dynamically estimates the level of confidence of the modalities, and enhances the robustness of decisions.

ChronoFactNet builds on existing evaluation frameworks like SVM-Fusion, TriSageNet and CMTHF that typically assess classification performance, but in the case of these approaches, limited to a 2D evaluation. ChronoFactNet is different from previous approaches such as SVM-Fusion, TriSageNet and CMTHF, which were primarily focused on classification performance and limited to 2D evaluation. Overall, the findings in the tables, Figures 4–5 and 10 clearly indicate that the proposed framework is a crucial improvement in the detection of multimodal fake news and it is a promising direction for future misinformation analysis systems.

6. Conclusion

The current study proposed a multi-modal fake-news detection method, namely ChronoFactNet, which is a memory-augmented framework to leverage the advantages of existing multi-modal architectures, based on the combination of semantic alignment, temporal consistency learning, counterfactual reasoning and adaptive truth fusion. The proposed framework can successfully harness the text, image and time information to capture the misinformation patterns of complexity and enhance the interpretability of the credibility assessment. Through extensive experiments conducted on the Fakeddit benchmark dataset, ChronoFactNet was shown to consistently yield the best results in terms of accuracy, robustness and reliability compared to the state of the art methods. The effectiveness of the proposed architecture for modeling semantic contradictions and historical misinformation patterns was confirmed by the results obtained from the comparative analysis, ablation studies, confusion-matrix evaluation, ROC and Precision–Recall analyses, assessment of temporal consistency, and the counterfactual testing. Overall, the results demonstrate that the multimodal approach of fake-news detection is scalable, interpretable and capable of achieving high levels of accuracy, and that introducing memory-driven reasoning mechanisms into future fake-news analysis systems is important.

Acknowledgement

The authors would like to thank Nakamura et al. (2020) for making the publicly available dataset of fakeddit available, which has been used as a benchmark dataset for the experimental evaluation of the proposed ChronoFactNet framework.

References

1. Alam, F., Cresci, S., Chakraborty, T., Silvestri, F., Dimitrov, D., Da San Martino, G., ... & Nakov, P. (2022, October). A survey on multimodal disinformation detection. In Proceedings of the 29th International Conference on Computational Linguistics (pp. 6625-6643).
2. Baevski, A., Zhou, H., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. Advances in Neural Information Processing Systems, 33, 12449–12460.
3. Beltagy, I., Peters, M. E., & Cohan, A. (2020). Longformer: The long-document transformer. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 8748–8758. https://doi.org/10.18653/v1/2020.acl-main.753

4. Chen, H., Yu, Y., Guo, H., Hu, B., Hu, S., Hu, J., & Wang, X. (2025). A self-learning multimodal approach for fake news detection. Frontiers in Artificial Intelligence, 8, 1665798.
5. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). 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, 4171–4186. https://doi.org/10.18653/v1/N19-1423
6. Dixit, D. K., Dangi, D., Kumar, J., Gupta, R., Sharma, S., & Bhagat, A. (2026). Graph attention and multi-neural memory networks for fake news detection: FakeDetectNet framework. Evolving Systems, 17(2), 25.
7. Guo, X., Li, Y., Wang, S., & Zhang, Q. (2023). Bilinear attention-based multimodal fusion for fake news detection. Expert Systems with Applications, 221, 119706. https://doi.org/10.1016/j.eswa.2023.119706 Kumari, R., & Ekbal, A. (2021). Amfb: Attention based multimodal factorized bilinear pooling for multimodal fake news detection. Expert Systems with Applications, 184, 115412.
8. Hangloo, S., & Arora, B. (2022). Combating multimodal fake news on social media: methods, datasets, and future perspective. Multimedia systems, 28(6), 2391-2422.
9. He, P., Liu, X., Gao, J., & Chen, W. (2020). DeBERTa: Decoding-enhanced BERT with disentangled attention. International Conference on Learning Representations (ICLR 2021). https://arxiv.org/abs/2006.03654
10. Khattar, D., Goud, J. S., Gupta, M., & Varma, V. (2019). MVAE: Multimodal variational autoencoder for fake news detection. The World Wide Web Conference (WWW 2019), 2915–2921. https://doi.org/10.1145/3308558.3313550
11. Mousavi, A., Abdollahinejad, Y., Corizzo, R., Japkowicz, N., & Boukouvalas, Z. (2026). E-CaTCH: Event-Centric Cross-Modal Attention With Temporal Consistency and Class-Imbalance Handling for Misinformation Detection. IEEE Transactions on Computational Social Systems.
12. Nakamura, K., Levy, S., & Wang, W. Y. (2019). r/fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection. arXiv preprint arXiv:1911.03854.
13. Pepino, L., Riera, P., Ferrer, L., & Gravano, A. (2021). Emotion recognition from speech using wav2vec 2.0 embeddings. Proceedings of Interspeech 2021, 3400–3404. https://doi.org/10.21437/Interspeech.2021-703
14. Qi, P., Cao, J., Li, X., Liu, H., Sheng, Q., Mi, X., ... & Yu, Y. (2021, October). Improving fake news detection by using an entity-enhanced framework to fuse diverse multimodal clues. In Proceedings of the 29th ACM international conference on multimedia (pp. 1212-1220).
15. Qian, S., Wang, J., Hu, J., Fang, Q., & Xu, C. (2021, July). Hierarchical multi-modal contextual attention network for fake news detection. In Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval (pp. 153-162).
16. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22–36. https://doi.org/10.1145/3137597.3137600
17. Tufchi, S., Yadav, A., & Ahmed, T. (2023). A comprehensive survey of multimodal fake news detection techniques: advances, challenges, and opportunities. International Journal of Multimedia Information Retrieval, 12(2), 28.
18. Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146–1151. https://doi.org/10.1126/science.aap9559