

# Utilizing deep learning models for ternary classification in COVID-19 infodemic detection

DIGITAL HEALTH  
Volume 10: 1–12  
© The Author(s) 2024  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20552076241284773  
journals.sagepub.com/home/dhj



Jia Luo<sup>1,2</sup> , Didier El Baz<sup>3</sup> and Lei Shi<sup>4</sup>

## Abstract

**Objective:** To address the complexities of distinguishing truth from falsehood in the context of the COVID-19 infodemic, this paper focuses on utilizing deep learning models for infodemic ternary classification detection.

**Methods:** Eight commonly used deep learning models are employed to categorize collected records as true, false, or uncertain. These models include fastText, three models based on recurrent neural networks, two models based on convolutional neural networks, and two transformer-based models.

**Results:** Precision, recall, and *F1*-score metrics for each category, along with overall accuracy, are presented to establish benchmark results. Additionally, a comprehensive analysis of the confusion matrix is conducted to provide insights into the models' performance.

**Conclusion:** Given the limited availability of infodemic records and the relatively modest size of the two tested data sets, models with pretrained embeddings or simpler architectures tend to outperform their more complex counterparts. This highlights the potential efficiency of pretrained or simpler models for ternary classification in COVID-19 infodemic detection and underscores the need for further research in this area.

## Keywords

Deep learning models, infodemic data, ternary classification problem, COVID-19, benchmark result

Submission date: 15 April 2024; Acceptance date: 3 September 2024

## Introduction

The term “infodemic” is a portmanteau combining “information” and “epidemic.” Coined by David J. Rothkopf in 2003 during the SARS outbreak, an infodemic refers to the rapid and extensive spread of both accurate and inaccurate information, often exacerbated by digital communication technologies.<sup>1</sup> The World Health Organization formally recognized the term during the COVID-19 pandemic to describe an overwhelming amount of information in both digital and physical environments during a disease outbreak.<sup>2</sup> Some scholars argue that continuous discussions about infodemics can heighten public anxiety, potentially eroding trust in authoritative sources, and increasing skepticism toward official information channels. Conversely, other scholars contend that information control and censorship measures aimed at combating infodemics may infringe

upon freedom of expression and limit access to diverse sources of information. Thus, effectively addressing infodemics presents a complex challenge.<sup>3</sup>

The onset of the COVID-19 pandemic witnessed a significant upsurge in the infodemic phenomenon. Misinformation

<sup>1</sup>College of Economics and Management, Beijing University of Technology, Beijing, China

<sup>2</sup>Chongqing Research Institute, Beijing University of Technology, Chongqing, China

<sup>3</sup>LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France

<sup>4</sup>State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing, China

### Corresponding author:

Lei Shi, State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing 100024, China.  
Email: leiky\_shi@cuc.edu.cn



during this period spanned from false cures and prevention methods to conspiracy theories regarding the virus's origins. This surge in misleading information presented considerable challenges for public health efforts, as inaccuracies had the potential to undermine official guidelines,<sup>4</sup> fuel panic,<sup>5</sup> and even induce harmful behaviors.<sup>6</sup> Effectively managing an infodemic requires active measures to promote accurate information and counter misinformation. However, there are inherent challenges associated with distinguishing truth from falsehood in the context of specific information during the outbreak. Some theories, such as the mode of transmission of COVID-19<sup>7</sup> and the characteristics of the virus,<sup>8</sup> are still in flux. Additionally, it is difficult to draw direct conclusions about certain events from reports.<sup>9</sup> Therefore, it is crucial to consider infodemic detection as a ternary classification, where one group contains records requiring further checking and those without a certain conclusion at that time.

Currently, misinformation detection heavily relies on manual monitoring, where suspected cases are forwarded to experts for verification. This approach is time-consuming and cannot keep up with the rapid spread of infodemic, especially as social media accelerates its dissemination. Moreover, detecting misinformation on social media becomes even more challenging when considering the poor quality of user-generated content, the complex semantics of natural language, and the high dimensionality of textual data, especially when malicious entities can frequently manipulate and change their writing style to mimic trustworthy content.<sup>10</sup> Hence, we suggest it is imperative to develop automated techniques for misinformation detection to address the infodemic more efficiently.

Several approaches have been employed to automatically combat misinformation, and deep learning algorithms have proven superior to traditional methods in most cases.<sup>11</sup> Nonetheless, these algorithms become less accurate at detecting COVID-19 fake news due to their lack of specific knowledge about the incidents. After the outbreak, methods dedicated to detecting COVID-19 fake news have emerged. However, all these approaches center on misinformation detection as a binary classification problem, neglecting the inherent challenge of distinguishing the truth from falsehood. Recognizing the challenges associated with identifying misinformation through manual monitoring in certain cases, it is advisable to consider infodemic detection as a ternary classification, categorizing information into true, false, and uncertain.<sup>12</sup>

Therefore, this paper is driven by the need to improve the accuracy and efficiency of detecting true, false, and uncertain information in the context of an infodemic. It is a preliminary study that employs deep learning models for infodemic ternary classification detection and evaluates their performance using two open-source data sets. The main contributions of our work are summarized as follows:

1. Eight frequently used deep learning models are employed for COVID-19 infodemic ternary classification, categorizing records into three groups: true, false, and uncertain.
2. Precision, recall, and *F1*-score for each group, along with the overall accuracy, are presented for the eight models on the English data set and the Chinese data set, establishing a benchmark result.
3. An analysis of the confusion matrix for eight models on two data sets is conducted to gain a deeper understanding of their performance across the three groups.

## Related works

Misinformation on the Internet is generally treated as fake news or rumors.<sup>13</sup> Many deep learning-based misinformation detection methods have been proposed and attracted significant attention recently. Zhang et al.<sup>14</sup> studied a graph attention network-based model that combined both sentiment and external knowledge comparison to meet the needs of fake news classification. Liu et al.<sup>15</sup> displayed a framework for detecting fake news by leveraging a graph neural network to jointly model the content, emotional information, and propagation structure of news conversations. Xie et al.<sup>16</sup> studied a social network fake news detection method by introducing the concept of gatekeepers into social network fake news detection and presenting a recurrent neural network (RNN)-based gatekeeping behavior model. Shrestha et al.<sup>17</sup> studied a role-relational graph convolutional network to exploit inter-relationships between stories, sources, and final users for jointly estimating the credibility degree of each entity and detecting fake news. Yang et al.<sup>18</sup> proposed a multimodal relationship-aware attention network for fake news detection where the captured text and image representations were input into the relationship-aware attention network. Chen et al.<sup>19</sup> developed a deep semantic-aware graph convolutional network for Cantonese rumor detection in social networks, which integrated the global structural information and the local semantic features. Chen et al.<sup>20</sup> presented a syntactic multilevel interaction network model that incorporated syntactic dependency relationships and a multilevel interaction network for rumor detection. Luvembe et al.<sup>21</sup> displayed a unified complementary attention fusion with an optimized deep neural network that captured subtle cross-modal relationships for multimodal fake news detection.

Misinformation around the COVID-19 is considered as the first social media infodemic.<sup>22</sup> Therefore, it has drawn the attention of deep learning models specifically designed for categorizing the COVID-19 infodemic as either true or false. Bangyal et al.<sup>23</sup> constructed a semantic model and applied it along with eight machine-learning algorithms and four deep-learning algorithms to identify COVID-19-related fake news, concluding that bidirectional

long short-term memory (BiLSTM) and convolutional neural network (CNN) exhibited the best performance. Paka et al.<sup>24</sup> introduced a cross-stitch-based semi-supervised end-to-end neural attention model for COVID-19 fake news detection, which demonstrated partial generalization to emerging fake news by incorporating relevant external knowledge into its learning process. Malla et al.<sup>25</sup> proposed a fusion technique-based ensemble deep learning model to detect fraudulent tweets during the COVID-19 epidemic where the fusion vector multiplication was designed to enhance the model's effectiveness. Chen et al.<sup>26</sup> used multiple deep learning model frameworks to detect misinformation in Chinese and English, comparing them based on different text feature selections, with BiLSTM producing the best detection results for COVID-19 fake news. Alghamdi et al.<sup>27</sup> evaluated various downstream neural network approaches for COVID-19 fake news detection where CT-BERT + BiGRU outperformed others with its effectiveness in capturing context and generating informative representations for downstream tasks. Xia et al.<sup>28</sup> studied a hybrid CNN-BiLSTM-AM model with an outlier knowledge management framework of generation–spread–identification–refutation for detecting COVID-19 fake news.

To the best of our knowledge, all deep learning models specifically developed for detecting COVID-19 infodemic have treated it as a binary classification problem. This choice is rooted in the fact that the majority of collected records used in the analysis and detection of the infodemic phenomenon are typically labeled as either true or false.<sup>29</sup> However, a few studies have taken a different approach, classifying these records into three to five groups to achieve a more comprehensive understanding of the COVID-19 infodemic and its impact with greater granularity. Haouari et al.<sup>30</sup> introduced an Arabic COVID-19 Twitter data set, wherein each tweet was labeled as either true, false, or categorized as other. Cheng et al.<sup>31</sup> assembled an English COVID-19 rumor data set by gathering news and tweets, which were then manually labeled as true, false, or unverified. Kim et al.<sup>32</sup> produced a data set comprising English claims and associated tweets, categorized into four groups: COVID true, COVID fake, non-COVID true, and non-COVID fake. Luo et al.<sup>33</sup> gathered widely spread Chinese infodemic content during the COVID-19 outbreak from Weibo and WeChat, classifying each record as true, false, or questionable after a four-time adjustment. Dharawat et al.<sup>34</sup> published a data set for assessing health risks associated with COVID-19-related social media posts, consisting of English tweets and tokens. Each entry in the data set was classified into one of five categories: real news/claims, not severe, possibly severe, highly severe, or refutes/rebuts misinformation. Luo et al.<sup>35</sup> released two balanced infodemic data sets by refining previously collected social media textual data with annotations from healthcare workers where all records

were categorized into three distinct groups: true, false, and uncertain.

The summary of misinformation detection in the literature is presented in Table 1. It is evident that numerous deep learning-based misinformation detection methods have been proposed and have garnered considerable interest. While the infodemic gained public attention following the COVID-19 pandemic, it has also drawn the focus of deep learning models specifically designed for infodemic detection. However, these models are limited in categorizing the COVID-19 infodemic as either true or false. Given the complexity of the infodemic, it is crucial to approach it as a multiclass classification problem. The above-mentioned data sets, which classify COVID-19 infodemic records into three to five groups, provide a starting point for research in this domain. Therefore, this paper aims to employ deep learning models for infodemic ternary classification, categorizing information into true, false, and uncertain. It specifically focuses on English and Chinese, as they are recognized as the two most common languages used on the Internet.<sup>36</sup>

## Methodology

This study centers on numerical analysis, specifically utilizing publicly available data sets in English and Chinese languages. The English data set was published on October 1, 2020<sup>37</sup> sourced from public fact-verification websites, Twitter API, and online tools.<sup>38,39</sup> The Chinese data set consists of records collected until April 10, 2020<sup>33</sup> sourced from manually verified Weibo posts, WeChat mini-program “Jiaozhen,” and authoritative sources.<sup>40,41,42</sup> All experiments are conducted on a MacBookPro equipped with a Dual-Core Intel Core i5 processor, using PyTorch, an open-source machine learning library.

## Data preparation

The two balanced data sets proposed in Luo et al.<sup>35</sup> are used in this research, as balanced data sets can significantly reduce bias in deep learning models. For model tuning, 10% of each data set is randomly selected, while the remaining 90% is randomly divided in a 3:1 ratio for training and testing. The details are presented in Table 2.

The English data originate from the training set published by Patwa et al.<sup>37</sup> Three healthcare workers were engaged to manually classify these records into three categories: true, false, and uncertain. Their evaluations relied solely on their own judgment, without reference to external sources, and the assigned label for each record was determined by a majority vote. To address the limited number of instances in the true category (830 records), an equal number of 830 records was selected from both the false and uncertain categories.

**Table 1.** Summary of misinformation detection in the literature.

Authors	Types of misinformation	Types of classification	Methods	Languages
Zhang et al. (2023)	General misinformation	Four-way classification and binary classification	A sentiment mixed heterogeneous network	English
Liu et al. (2023)	General misinformation	Four-way classification and binary classification	An emotion-aware graph neural network	English, Chinese
Xie et al. (2023)	General misinformation	Binary classification	A recurrent neural network-based gatekeeping behavior model	English, Chinese
Shrestha et al. (2023)	General misinformation	Six-way classification and binary classification	Role-relational graph convolutional networks	English
Yang et al. (2024)	General misinformation	Binary classification	A multimodal relationship-aware attention network	English, Chinese
Chen et al. (2024)	General misinformation	Binary classification	A deep semantic-aware graph convolutional network	Cantonese
Chen et al. (2024)	General misinformation	Four-way classification and binary classification	A syntactic multilevel interaction network	English, Chinese
Luvembe et al. (2024)	General misinformation	Binary classification	An optimized deep neural network	English
Bangyal et al. (2021)	COVID-19 infodemic	Binary classification	Semantic model based deep learning approaches	English
Paka et al. (2021)	COVID-19 infodemic	Binary classification	A cross-stitch semi-supervised neural attention model	English
Malla et al. (2022)	COVID-19 infodemic	Binary classification	Pretrained transformer models	English
Chen et al. (2023)	COVID-19 infodemic	Binary classification	Text feature selection based deep learning models	English, Chinese
Alghamdi et al. (2023)	COVID-19 infodemic	Binary classification	Transformer-based models	English
Xia et al. (2023)	COVID-19 infodemic	Binary classification	A hybrid CNN-BiLSTM-AM model	English
Haouari et al. (2020)	COVID-19 infodemic	Ternary classification	-	Arabic
Cheng et al. (2021)	COVID-19 infodemic	Ternary classification	-	English
Kim et al. (2021)	COVID-19 infodemic	Four-way classification	-	English
Luo et al. (2021)	COVID-19 infodemic	Ternary classification	-	Chinese
Dharawat et al. (2022)	COVID-19 infodemic	Five-way classification	-	English
Luo et al. (2023)	COVID-19 infodemic	Ternary classification	-	English, Chinese

**Table 2.** Statistics of the COVID-19 infodemic ternary classification detection data sets.

		Uncertain	False	True	Total
English data set	Training set	576	554	551	1681
	Validation set	75	92	82	249
	Test set	179	184	197	560
	Total	830	830	830	2490
Chinese data set	Training set	288	191	234	713
	Validation set	41	26	38	105
	Test set	106	64	67	237
	Total	435	281	339	1055

The Chinese data are sourced from Luo et al.,<sup>33</sup> where all instances are classified as either strongly related health records or weakly related health records based on their content. The strongly related health records are further subdivided into categories such as prevention measures, general virus knowledge, and treatment information. The weakly related health records are further subdivided into categories such as local measures, national measures, patient information, and others. By examining the properties of the collected records, the initially imbalanced data set was adjusted over four rounds, resulting in 435 records labeled as questionable, 281 as false, and 339 as true. The final labels achieved high intercoder reliability with healthcare workers' annotations. Therefore, the classification results were retained, with the label questionable being replaced by uncertain.

### Research models

Eight commonly used deep learning models are employed for the ternary classification of the COVID-19 infodemic in this research. These include fastText, three models based on RNNs, two models based on CNNs, and two transformer-based models. The dropout rate and early stopping criteria for each model are determined through experimentation and observation to achieve the optimal balance between preventing overfitting and maintaining high performance.

- FastText<sup>43</sup>: It is a library for learning word embeddings and text classification, representing each word as a bag of character n-grams. To enhance the solution's quality, the concatenation of the standard unigram

average with bigram and trigram vector averages is employed. In this study, the number of hidden units in the hidden layer is set to a fixed value of 64.

- TextRNN<sup>44</sup>: The BiLSTM is selected for TextRNN. The LSTM is a widely utilized RNN architecture while the bidirectional structure enables the network to incorporate both backward and forward information at every time step. In this study, the size of the hidden units in the BiLSTM is set to a fixed value of 64, and the number of hidden layers is fixed at 4.
- TextRNN\_Attn<sup>45</sup>: This model is an extension of TextRNN with a neural attention mechanism where an attention layer is introduced after the BiLSTM layer. The attention layer produces a weight vector, which is then utilized to merge word-level features from each time step into a sentence-level feature vector by multiplication.
- TextRCNN<sup>46</sup>: This model is an extension of TextRNN with a max-pooling layer, which is applied after computing representations for all words. The representation of each word is formed by concatenating the left-side context vector, the word embedding, and the right-side context vector where the BiLSTM is employed to capture both the left and right contexts of a word.
- TextCNN<sup>47</sup>: This model employs convolutional operations on input text sequences to capture local patterns and features. It consists of a single convolutional layer followed by a max-pooling layer. In this study, the filter sizes are configured as 2, 3, and 4, with 64 filters for each size.
- DPCNN<sup>48</sup>: This model is an extension of TextCNN by increasing the depth of the network. It iteratively alternates between a convolution block and a down-sampling layer. Therefore, the size of internal data diminishes in a pyramid shape and the final layer aggregates internal data for each record into a single vector.
- Transformer<sup>49</sup>: This model relies solely on attention mechanisms, specifically self-attention and multihead attention, allowing it to capture long-range dependencies in input sequences efficiently. In this study, it is configured with a hidden size of 768, an intermediate size of 3072, 12 attention heads, and 12 hidden layers.
- BERT<sup>50</sup>: This model is an extension of the transformer, specifically the bidirectional encoder representations from transformers. In this study, the BERT-Base, cased version is applied for COVID-19 infodemic ternary classification detection in English, while the BERT-Base, Chinese version is used for processing Chinese records.

### Evaluation metrics

Precision, recall, *F1*-score, and accuracy are used as evaluation metrics to assess the performance of eight deep learning models. Additionally, the confusion matrix is utilized to visualize the model's actual classifications. In the ternary

**Table 3.** Test results for COVID-19 infodemic ternary classification detection in English.

		Uncertain	False	True
FastText	Precision	0.6429	0.6396	0.6887
	Recall	0.4525	0.7717	0.7411
	<i>F1</i> -score	0.5311	0.6995	0.7139
	Accuracy	0.6589		
TextRNN	Precision	0.5732	0.6143	0.6833
	Recall	0.5028	0.7446	0.6244
	<i>F1</i> -score	0.5357	0.6732	0.6525
	Accuracy	0.6520		
TextRNN_Att	Precision	0.6357	0.6050	0.7150
	Recall	0.4581	0.7826	0.7005
	<i>F1</i> -score	0.5325	0.6825	0.7077
	Accuracy	0.6500		
TextRCNN	Precision	0.5789	0.6245	0.7020
	Recall	0.4302	0.7772	0.7056
	<i>F1</i> -score	0.4936	0.6925	0.7038
	Accuracy	0.6411		
TextCNN	Precision	0.5989	0.6554	0.7143
	Recall	0.6257	0.6304	0.7107
	<i>F1</i> -score	0.6120	0.6427	0.7125
	Accuracy	0.6571		
DPCNN	Precision	0.6144	0.6054	0.6216
	Recall	0.5251	0.6087	0.7005
	<i>F1</i> -score	0.5663	0.6070	0.6587
	Accuracy	0.6143		
Transformer	Precision	0.5064	0.5538	0.5780
	Recall	0.4413	0.5598	0.6396
	<i>F1</i> -score	0.4716	0.5568	0.6072
	Accuracy	0.5500		

(continued)



Table 3. Continued.

		Uncertain	False	True
BERT	Precision	0.7752	0.7150	0.7880
	Recall	0.5587	0.8315	0.8680
	F1-score	0.6494	0.7688	0.8261
	Accuracy	0.7571		

classification prediction task, the classifications are labeled as 0, 1, and 2, representing uncertain, false, and true, respectively.

- **Precision:** Precision for class  $i$  is the ratio of correctly predicted instances of class  $i$  to the total instances predicted as class  $i$ . It is formulated as

$$P_i = \frac{T_{ii}}{T_{ii} + \sum_{j \neq i} T_{ji}}. \quad (1)$$

- **Recall:** Recall for class  $i$  is the ratio of correctly predicted instances of class  $i$  to the total instances that actually belong to class  $i$ . It is formulated as

$$R_i = \frac{T_{ii}}{T_{ii} + \sum_{j \neq i} T_{ij}}. \quad (2)$$

- **F1-score:** It is the harmonic mean of precision and recall, providing a balance between the two metrics, which is formulated as

$$F1_i = 2 \times \frac{P_i \times R_i}{P_i + R_i}. \quad (3)$$

- **Accuracy:** It is the ratio of correctly predicted instances to the total number of instances which is formulated as

$$ACC = \frac{\sum_i T_{ii}}{\sum_{i,j} T_{ij}}. \quad (4)$$

## Statistical analysis

The test results for COVID-19 infodemic ternary classification detection in English are presented in Table 3, while the test results for Chinese records are displayed in Table 4. The confusion matrix for English records is presented in Figure 1, while the confusion matrix for Chinese records is displayed in Figure 2.

Regarding the test results for English records, the accuracy of fastText, TextRNN, TextRNN\_Att, TextRCNN, and TextCNN is ~65%, indicating a similar level of performance. However, the accuracy of DPCNN and transformer is low at 61.43% and 55.00%, respectively. Notably,

BERT achieves the highest accuracy at 75.71%. For records labeled as uncertain, the F1-score is low for TextRCNN and transformer at 49.36% and 47.16%, respectively, while it is high for TextCNN and BERT at 61.12% and 64.94%, respectively. For records labeled as false, the F1-score is low for TextCNN, DPCNN, and transformer at 64.27%, 60.70%, and 55.68%, respectively, while it is high for BERT at 76.88%. For records labeled as true, the F1-score is low for TextRNN, DPCNN, and transformer at 65.25%, 65.87%, and 60.72%, respectively, while it is high for BERT at 82.61%.

Concerning the test results for Chinese records, the accuracy of TextRNN, TextRNN\_Att, TextRCNN, and DPCNN ranges from 71% to 76%. The accuracy of transformer is low at 67.09%, while fastText, TextCNN, and BERT are high at 78.06%, 78.06%, and 81.01%, respectively. For records labeled as uncertain, the F1-score is low for transformer at 68.97%, while it is high for fastText, TextRNN\_Att, TextCNN, DPCNN, and BERT at 84.76%, 82.18%, 83.41%, 80.00%, and 84.85%, respectively. For records labeled as false, the F1-score is low for DPCNN and transformer at 58.82% and 55.74%, respectively, while it is high for BERT at 71.23%. For records labeled as true, the F1-score is low for TextRNN, TextRNN\_Att, TextRCNN, DPCNN, and transformer at 72.18%, 73.38%, 74.29%, 70.34%, and 73.83%, while it is high for BERT at 86.15%.

Regarding the confusion matrix for English records, the overall performance of the eight models is best for records labeled as true, followed by those labeled as false, and worst for records labeled as uncertain. In detail, BERT achieves the best performance for records labeled as true, while TextRNN performs the worst. BERT achieves the best performance for records labeled as false, while transformer performs the worst. TextCNN achieves the best performance for records labeled as uncertain, while TextRCNN performs the worst. Moreover, more records are misclassified as uncertain than misclassified as false in the true group in most cases, and more records are misclassified as false than misclassified as true in the uncertain group in most cases. This tendency is not obvious in the false group.

Concerning the confusion matrix for Chinese records, the overall performance of the eight models is best for

**Table 4.** Test results for COVID-19 infodemic ternary classification detection in Chinese.

		Uncertain	False	True
FastText	Precision	0.8558	0.7000	0.7397
	Recall	0.8396	0.6562	0.8060
	<i>F1</i> -score	0.8476	0.6774	0.7714
	Accuracy	0.7806		
TextRNN	Precision	0.8261	0.5823	0.7273
	Recall	0.7170	0.7188	0.7164
	<i>F1</i> -score	0.7677	0.6434	0.7218
	Accuracy	0.7173		
TextRNN_Att	Precision	0.8646	0.6522	0.7083
	Recall	0.7830	0.7031	0.7612
	<i>F1</i> -score	0.8218	0.6767	0.7338
	Accuracy	0.7553		
TextRCNN	Precision	0.7778	0.6786	0.7123
	Recall	0.7925	0.5938	0.7761
	<i>F1</i> -score	0.7850	0.6333	0.7429
	Accuracy	0.7342		
TextCNN	Precision	0.8381	0.6875	0.7794
	Recall	0.8302	0.6875	0.7910
	<i>F1</i> -score	0.8341	0.6875	0.7852
	Accuracy	0.7806		
DPCNN	Precision	0.8077	0.6364	0.6538
	Recall	0.7925	0.5469	0.7612
	<i>F1</i> -score	0.8000	0.5882	0.7034
	Accuracy	0.7173		
Transformer	Precision	0.7216	0.5862	0.6707
	Recall	0.6604	0.5312	0.8209
	<i>F1</i> -score	0.6897	0.5574	0.7383
	Accuracy	0.6709		

(continued)



Table 4. Continued.

		Uncertain	False	True
BERT	Precision	0.9130	0.6341	0.8889
	Recall	0.7925	0.8215	0.8358
	F1-score	0.8485	0.7123	0.8615
	Accuracy	0.8101		

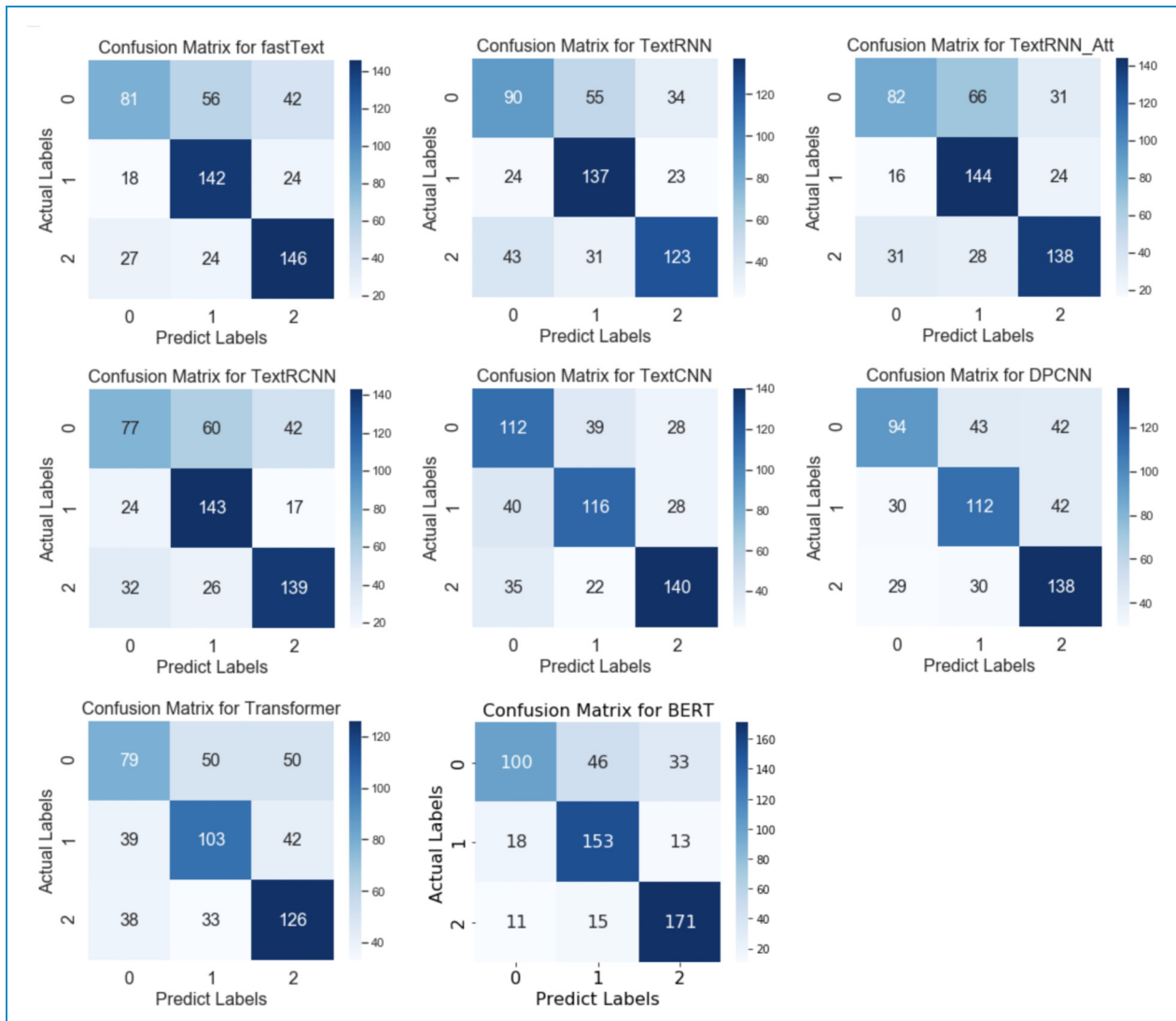
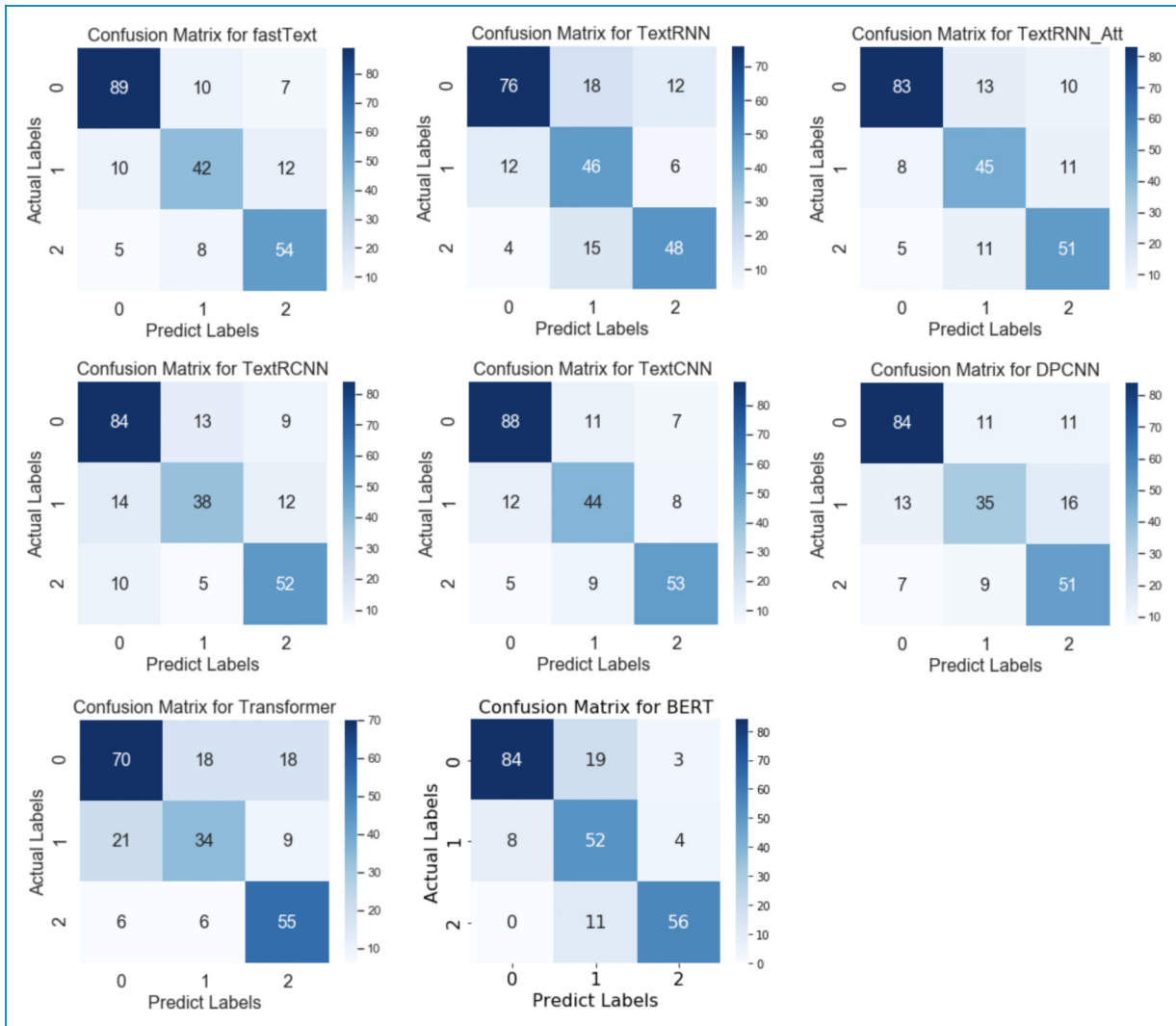


Figure 1. Confusion matrix of eight deep learning models for ternary classification of COVID-19 infodemic in English.

records labeled as uncertain, followed by those labeled as true, and worst for records labeled as false. In detail, fastText achieves the best performance for records labeled as uncertain, while transformer performs the worst. BERT

achieves the best performance for records labeled as true, while TextRNN performs the worst. BERT achieves the best performance for records labeled as false, while transformer performs the worst. Moreover, more records are



**Figure 2.** Confusion matrix of eight deep learning models for ternary classification of COVID-19 infodemic in Chinese.

misclassified as false than misclassified as true in the uncertain group in most cases, and more records are misclassified as false than misclassified as uncertain in the true group in most cases. This tendency is not obvious in the false group.

## Discussions

The overall performance of deep learning models for infodemic ternary classification detection is better on the Chinese data set than on the English data set. BERT achieves the highest accuracy and consistently obtains high  $F1$ -scores across all categories. This indicates BERT's robustness in understanding and classifying complex linguistic patterns and contexts related to infodemic content. On the opposite, transformer has the lowest accuracy on both data sets underscoring the need for language-specific model tuning or selection. Surprisingly, fastText demonstrates unexpectedly good performance just after BERT. Among the three RNN-based

models, there is a similar level of performance, with TextRNN achieving the best on the English data set and TextRNN\_Att excelling on the Chinese data set. Regarding the two CNN-based models, the shallow TextCNN outperforms the deep DPCNN on both data sets.

The tested deep learning models exhibit a greater proficiency in learning the features for records labeled as false and true on the English data set. In contrast, the tested models display a higher capability in learning the features for records labeled as uncertain and true on the Chinese data set. This discrepancy may stem from linguistic or contextual variations between the two languages, impacting the models' ability to generalize features effectively. Furthermore, BERT consistently achieves the best performance in most cases, while TextCNN and fastText also perform well in specific instances. Given the limited availability of infodemic records and the relatively modest size of the two tested data sets, models with pretrained

embeddings or simpler architectures tend to outperform more complex counterparts.

Some limitations of this study cannot be overlooked. Firstly, the limited availability and modest size of the tested infodemic data sets are highlighted. Such constraints affect the robustness of the findings and may impact the models' performance when applied to larger or more diverse data sets. Secondly, models demonstrate different capabilities across English and Chinese data sets. It is challenging to discern whether this performance discrepancy arises from linguistic and contextual variations or from the models themselves. Thirdly, only the efficiency metrics of the models are discussed, while their effectiveness is not considered. Finally, a specific set of deep learning models is evaluated. There may be other models or techniques that could perform better but are not included in this study.

## Conclusions

The infodemic has gained significant public attention in the wake of the COVID-19 pandemic. Given its complexity, it is crucial to address infodemic detection as a multiclass classification problem. This paper focused on the application of deep learning models for infodemic ternary classification, categorizing information into true, false, and uncertain. Firstly, eight commonly used models, including fastText, RNN-based models, CNN-based models, and transformer-based models, were tested on collected English and Chinese records. Secondly, precision, recall, and  $F1$ -score metrics for each category, along with overall accuracy, were presented to establish a benchmark. BERT demonstrated its robustness in understanding and classifying the complex linguistic patterns and contexts related to infodemic content. Thirdly, a comprehensive analysis of the confusion matrix was conducted to provide insights into the models' performance. In these evaluations, BERT consistently achieved the best results in most cases, while TextCNN and fastText also performed well in specific instances. Finally, a discussion highlighted the potential efficiency of pretrained or simpler models for infodemic ternary classification detection due to the limited availability of infodemic records.

**Author's note:** Lei Shi is also affiliated with Key Laboratory of Ethnic Language Intelligent Analysis and Security Governance of MOE, Minzu University of China, Beijing, China, and Key Laboratory of Education Informatization for Nationalities (Yunnan Normal University), Ministry of Education, Kunming, China.

**Contributorship:** Jia Luo initiated the idea, addressed the whole issues in the manuscript, and wrote the manuscript. Didier El Baz revised and polished the final edition of the manuscript. Lei Shi conducted the numerical experiments.

**Data availability:** Data will be made available on request.

**Declaration of conflicting interests:** The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding:** The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Natural Science Foundation of China (Grant No. 72104016), the Natural Science Foundation of Chongqing, China (Grant No. CSTB2023NSCQ-MSX0391), Beijing Natural Science Foundation (Grant No. 9242003), the R&D Program of the Beijing Municipal Education Commission (Grant No. SM202110005011), the Key Laboratory of Ethnic Language Intelligent Analysis and Security Governance of MOE (Grant No. 202306) and the Foundation of Key Laboratory of Education Informatization for Nationalities (Yunnan Normal University), Ministry of Education (Grant No. EIN2024C006).

**ORCID ID:** Jia Luo  <https://orcid.org/0000-0003-3914-8865>

## References

1. Simon FM and Camargo CQ. Autopsy of a metaphor: the origins, use and blind spots of the 'infodemic'. *New Media & Society* 2023; 25: 2219–2240.
2. [https://www.who.int/health-topics/infodemic#tab=tab\\_1](https://www.who.int/health-topics/infodemic#tab=tab_1).
3. <https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation>.
4. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters>.
5. <https://www.bbc.com/worklife/article/20200304-coronavirus-covid-19-update-why-people-are-stockpiling>.
6. <https://time.com/5835244/accidental-poisonings-trump/>.
7. Lewis D. Why the WHO took two years to say COVID is airborne. *Nature* 2022; 604: 26–31.
8. Libby P and Lüscher T. COVID-19 is, in the end, an endothelial disease. *Eur Heart J* 2020; 41: 3038–3044.
9. Kozlov M. How a rural school teacher became a top COVID sleuth. *Nature* 2023; 616: 636–637.
10. Alghamdi J., Lin Y. and Luo S. (2022, August). Modeling fake news detection using bert-cnn-bilstm architecture. In 2022 IEEE 5th International Conference on Multimedia Information Processing and Retrieval (MIPR). 354–357. IEEE.
11. Zhou X and Zafarani R. A survey of fake news: fundamental theories, detection methods, and opportunities. *ACM Comput Surv (CSUR)* 2020; 53: 1–40.
12. Luo J, Xue R and Hu J. COVID-19 infodemic on Chinese social media: a 4P framework, selective review and research directions. *Meas Control* 2020; 53: 2070–2079.
13. Bondielli A and Marcelloni F. A survey on fake news and rumour detection techniques. *Inf Sci* 2019; 497: 38–55.
14. Zhang H, Li Z, Liu S, et al. Do sentence-level sentiment interactions matter? Sentiment mixed heterogeneous network for

- fake news detection. *IEEE Trans Comput Soc Syst* 2024; 11: 5090–5100.
15. Liu F, Zhang X and Liu Q. An emotion-aware approach for fake news detection. *IEEE Trans Comput Soc Syst* 2024; 11: 3516–3524.
  16. Xie B and Li Q. Detecting fake news by RNN-based gate-keeping behavior model on social networks. *Expert Syst Appl* 2023; 231: 120716.
  17. Shrestha A, Duran J, Spezzano F, et al. Joint credibility estimation of news, user, and publisher via role-relational graph convolutional networks. *ACM Trans Web* 2023; 18: 1–24.
  18. Yang H, Zhang J, Zhang L, et al. MRAN: multimodal relationship-aware attention network for fake news detection. *Comput Stand Interfaces* 2024; 89: 103822.
  19. Chen X, Jian Y, Ke L, et al. A deep semantic-aware approach for Cantonese rumor detection in social networks with graph convolutional network. *Expert Syst Appl* 2024; 245: 123007.
  20. Chen Z, Zhuang F, Liao L, et al. A syntactic multi-level interaction network for rumor detection. *Neural Comput Appl* 2024; 36: 1713–1726.
  21. Luvembe AM, Li W, Li S, et al. CAF-ODNN: complementary attention fusion with optimized deep neural network for multimodal fake news detection. *Inf Process Manag* 2024; 61: 103653.
  22. Ahmad AR and Murad HR. The impact of social media on panic during the COVID-19 pandemic in Iraqi Kurdistan: online questionnaire study. *J Med Internet Res* 2020; 22: e19556.
  23. Bangyal WH, Qasim R, Ahmad Z, et al. Detection of fake news text classification on COVID-19 using deep learning approaches. *Comput Math Methods Med* 2021. DOI: 10.1155/2021/5514220
  24. Paka WS, Bansal R, Kaushik A, et al. Cross-SEAN: a cross-stitch semi-supervised neural attention model for COVID-19 fake news detection. *Appl Soft Comput* 2021; 107: 107393.
  25. Malla S and Alphonse PJA. Fake or real news about COVID-19? Pretrained transformer model to detect potential misleading news. *Eur Phys J Special Top* 2022; 231: 3347–3356.
  26. Chen MY, Lai YW and Lian JW. Using deep learning models to detect fake news about COVID-19. *ACM Trans Internet Technol* 2023; 23: 1–23.
  27. Alghamdi J, Lin Y and Luo S. Towards COVID-19 fake news detection using transformer-based models. *Knowl Based Syst* 2023; 274: 110642.
  28. Xia H, Wang Y, Zhang JZ, et al. COVID-19 fake news detection: a hybrid CNN-BiLSTM-AM model. *Technol Forecast Soc Change* 2023; 195: 122746.
  29. Murayama T. Dataset of fake news detection and fact verification: a survey. arXiv preprint arXiv:2111.03299. 2021.
  30. Haouari F, Hasanain M, Suwaileh R, et al. ArCOV19-rumors: Arabic COVID-19 twitter dataset for misinformation detection. arXiv preprint arXiv:2010.08768. 2020.
  31. Cheng M, Wang S, Yan X, et al. A COVID-19 rumor dataset. *Front Psychol* 2021; 12: 644801.
  32. Kim J, Aum J, Lee S, et al. FibVID: comprehensive fake news diffusion dataset during the COVID-19 period. *Telemat Inform* 2021; 64: 101688.
  33. Luo J, Xue R, Hu J, et al. Combating the infodemic: a Chinese infodemic dataset for misinformation identification. *In Healthcare* 2021; 9: 1094.MDPI
  34. Dharawat A, Lourentzou I, Morales A, et al. Drink bleach or do what now? covid-hera: a study of risk-informed health decision making in the presence of covid-19 misinformation. In Proceedings of the International AAAI Conference on Web and Social Media, 2022, May, Vol. 16, pp. 1218–1227.
  35. Luo J, Peng D, Shi L, et al. A comparative analysis of the COVID-19 infodemic in English and Chinese: insights from social media textual data. *Front Public Health* 2023; 11: 1281259.
  36. Statista. Available at [www.statista.com/statistics/262946/share-of-the-most-common-languages-on-the-internet](http://www.statista.com/statistics/262946/share-of-the-most-common-languages-on-the-internet). 2023.
  37. Patwa P, Sharma S, Pykl S, et al. Fighting an infodemic: Covid-19 fake news dataset. In Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021, Collocated with AAAI 2021, Virtual Event, February 8, 2021, Revised Selected Papers 1 (pp. 21–29). Springer International Publishing.
  38. <https://toolbox.google.com/factcheck/explorer>.
  39. <http://poy.nu/ifcnbot>.
  40. <https://www.tencent.com/zh-cn/responsibility/combat-covid-19-handbook.html>.
  41. <http://www.nhc.gov.cn/zycj/s7653p/202008/0a7bdf12bd4b46e5bd28ca7f9a7f5e5a.shtml>.
  42. [https://www.who.int/zh/emergencies/diseases/novel-coronavirus-2019?gclid=CjwKCAiAINf-BrB\\_EiwA2osbxRKB\\_bkVsu64Vrc2d4xOD75fOvcPIXwGzaEdwx5VXsn-0LcYYTx-0BoC\\_RjYQAvD\\_BwE](https://www.who.int/zh/emergencies/diseases/novel-coronavirus-2019?gclid=CjwKCAiAINf-BrB_EiwA2osbxRKB_bkVsu64Vrc2d4xOD75fOvcPIXwGzaEdwx5VXsn-0LcYYTx-0BoC_RjYQAvD_BwE).
  43. Joulin A, Grave E, Bojanowski P, et al. Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759. 2016.
  44. Liu P, Qiu X and Huang X. Recurrent neural network for text classification with multi-task learning. arXiv preprint arXiv:1605.05101. 2016.
  45. Zhou P, Shi W, Tian J, et al. Attention-based bidirectional long short-term memory networks for relation classification. In Proceedings of the 54th annual meeting of the association for computational linguistics. 2016, August. (volume 2: Short papers) (pp. 207–212).
  46. Lai S, Xu L, Liu K, et al. Recurrent convolutional neural networks for text classification. In Proceedings of the AAAI conference on artificial intelligence. 2015, February. (Vol. 29).
  47. Kim Y. Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882. 2014.
  48. Johnson R and Zhang T. Deep pyramid convolutional neural networks for text categorization. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017, July. (Vol. 1: Long Papers) (pp. 562–570).
  49. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: *Proceedings of the 31st international conference on neural information processing systems (NIPS'17)*, Long Beach, CA, USA, 4–9 December, pp. 6000–6010. Red Hook, NY, USA: Curran Associates Inc.
  50. Devlin J, Chang MW, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. 2018.