Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 13, No. 4, 2025, 427-440.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

Multilingual COVID-19 Fake News Detection with Hybrid Capsule Neural Networks

Mohammad Hadi Goldani, Saeedeh Momtazi* and Reza Safabakhsh

Computer Engineering Department, Amirkabir University of Technology, Tehran, Iran.

Article Info

Article History:
Received 21 January 2025
Revised 26 May 2025
Accepted 04 Juy 2025

DOI:10.22044/jadm.2025.15646.2681

Keywords:
Fake News Detection, COVID-19, Dynamic transformer, Deep Convolutional Neural Networks, Capsule Neural Networks.

**Corresponding author: momtazi@aut.ac.ir (S. Momtazi).*

Abstract

The widespread use of web-based forums and social media has led to an increase in news consumption. To mitigate the impact of fake news on users' health-related decisions, it is crucial to develop machine learning models that can automatically detect and combat fake news. In this paper, we propose a novel multilingual model, Hybrid CapsNet, which incorporates a dynamic transformer model for COVID-19 fake news detection in both English and Persian. Our model incorporates two dynamic pre-trained representation models that incrementally update word embeddings during training: dynamic RoBERTa for English and dynamic ParsBERT for Persian. Additionally, it features two parallel classifiers with a new loss function, margin loss. By utilizing a dynamic transformer and both Deep Convolutional Neural Networks (DCNNs) and Capsule Neural Networks (CapsNets), we achieve better performance than state-of-the-art baselines. To evaluate the proposed model, we use two recent COVID-19 datasets in English and Persian. Our results, in terms of F1-score, demonstrate the effectiveness of the Hybrid CapsNet model. The model outperforms existing baselines, suggesting its potential as an effective tool for detecting and combating COVID-19-related fake news in multiple languages. Overall, our study underscores the importance of developing effective machine learning models for combating fake news during critical events, such as the COVID-19 pandemic. The proposed model has the potential to be applied to other languages and domains, serving as a valuable tool for protecting public health and safety.

1. Introduction

The healthcare systems in our societies have faced significant challenges due to the SARS-CoV-2 (COVID-19) outbreak. Social media facilitates the easy, convenient, and rapid dissemination of news [1], making it a pervasive source of information. However, platforms like these have both constructive and destructive impacts. As an integral part of culture and society, social media is a double-edged sword [2]. Individuals may manipulate and spread misinformation for profit or entertainment in the form of fake news [3]. Fake news has become a widespread problem in the 21st century, particularly since the 2016 US presidential election [4,5,6]. COVID-19 has become a hotbed for the spread of online misinformation, ranging from conspiracy theories to fake cures and treatments. During the SARS-CoV-2 (COVID-19) pandemic, fake news spread globally, causing widespread chaos. Understanding and addressing fake news is crucial for community well-being, both during disease resurgences and in similar situations [7,8]. The global spread of fake news since the outbreak has compounded the problem [9,10]. Fake news about COVID-19, which often focuses on health issues, has significantly impacted people's behavior, leading to less effective healthcare measures during

the pandemic [11,12]. For example, Nigerian individuals were misled by news articles claiming that chloroquine, a malaria drug, was an effective treatment for COVID-19, resulting in overdoses [13]. Similarly, the fake news claim that "Alcohol is a cure for COVID-19" led to numerous hospitalizations and deaths in Iran [14].

A global infodemic has emerged due to the COVID-19 pandemic, driven by a rapid rise in political and medical fake news. Despite the increasing popularity of social media platforms for news consumption, a growing percentage of the population distrusts the reliability of information shared on these platforms. The prevalence of misinformation, including fake news, on social media poses a significant risk to audiences worldwide. Consequently, detecting and mitigating the effects of fake news is a pressing concern in research utilizing linguistic and deep learning models to address this issue [15,16].

This paper proposes a novel model based on a dynamic transformer and a Hybrid CapsNet for detecting COVID-19 fake news in English and Persian languages. The proposed model enhances the DCNN and CapsNet architectures by incorporating margin loss. We also compare various word representation layers in the English language and ultimately utilize the robustly optimized Bidirectional Encoder Representations from Transformers (RoBERTa) [17] in our proposed model. We demonstrate that the proposed models outperform state-of-the-art methods on the COVID-19 fake news dataset. We also discuss the effects of unigrams, subjectivity, and polarity in text for the detection of fake news in general. The remainder of this paper is organized as follows: Section 2 reviews related work on fake news detection in COVID-19 datasets. Section 3 presents the proposed model. Section 4 introduces the COVID-19 fake news detection datasets for English and Persian languages. Section 5 presents the experimental comparison with baseline classification and discussion. Finally, Section 6 concludes the paper's results.

2. Related Work

The widespread adoption of the internet and social media as primary sources of information and news has significantly intensified the challenge of combating fake news. This has driven extensive research into identifying fake news, leading to the development of various related tasks, such as sentiment analysis in the context of COVID-19 [18,19]. Our focus here is on detecting COVID-19-related fake news using supervised learning methods. By training models with both authentic and misleading news data through machine learning or deep learning approaches, we aim to predict the credibility of incoming news stories. This section discusses convolutional neural networks (CNNs), long short-term memory (LSTM) networks, and BERT models, which are commonly employed in studies addressing the detection of COVID-19 fake news. Among the most commonly used approaches in natural language processing (NLP), word embedding works by learning a low-dimensional vector representation of a word in the text [20].

The power of word embedding algorithms such as Word2Vec [21], FastText [22], and GloVe [23] in capturing semantic and syntactic word relationships has been proven. This capability has facilitated various NLP tasks, such as aspect extraction, part-of-speech tagging, and sentiment analysis [24]. The idea of distributional semantics states that words occurring in the same context tend to have similar meanings. In fact, word embeddings reveal hidden relationships between words that can be used during the training process. However, early embedding techniques mentioned above, which provide a static embedding vector for a word in various contexts, limit the semantic information about the word. Therefore, in recent years, different deep, transform-based word representation methods have been implemented by researchers. These techniques for generating vectors for a word in various contexts can consider textual information.

BERT [25] is a transformer-based, unsupervised deep model that has been trained on a massive corpus of text using two different scenarios [26]: (1) the masked language model that learns the relationships between the adjacency of words in a sentence, and (2) the prediction of the next sentence that learns relations between sentences. RoBERTa [17] was presented with some hyperparameter tuning and modifications to the learning process of BERT. Liu et al. [17] claimed that BERT was undertrained. Therefore, in addition to longer sequences of the large dataset for model training, they used longer batches for training the model. They also evaluated alternative training approaches and, as a result, claimed that removing the next sentence prediction loss can enhance the performance of the learning process. ParsBERT [27] is a BERT model designed specifically for the Persian language. It has been shown to outperform other architectures and multilingual models, achieving state-of-the-art performance. To overcome the limited availability of data for Persian NLP tasks, a large dataset was compiled for pretraining the model. ParsBERT has achieved

consistently high scores on all datasets, including existing and newly composed ones. It has also surpassed the performance of multilingual BERT and previous works in tasks such as sentiment analysis, text classification, and named entity recognition.

In recent years, different variations of CNNs have been used in the task of fake news detection [28,29,30,31,32]. A CNN architecture with convolutional and pooling layers can accurately extract features from local to global representations, which is a powerful representational capability of CNNs. To extract more robust features for the learning process, the CNN needs to be enhanced with additional identifying information. For this purpose, the similarity of the intracluster and intercluster differences of the learned features should be maximized. To achieve this goal, one of the most commonly used loss functions in CNNs for fake news tasks is margin loss [29]. Using this loss function helps mitigate overlapping problems and reduces the model's tendency to overfit [29].

After the success of CapsNets in NLP tasks [33], in recent years, different models based on CapsNets have been used for fake news detection [34,35,36]. CapsNet incorporates a standard convolutional layer known as an n-gram convolutional layer that performs the function of extracting features. Second, the primary capsule layer represents scalar value features in the form of capsules. In the convolutional capsule layer, the outputs of the previous layer are fed into a new layer referred to as the convolutional capsule layer. Every capsule in this layer is only connected to its local area in the layer below. Lastly, the feed-forward capsule layer is applied to the output of the previous layer. Each capsule in the output is considered to be a member of a specific class for this layer. To train the model, the maximum margin loss is employed [37].

Patwa et al. [38] developed a COVID-19 dataset featuring a diverse collection of tweets that encompassed both factual and misleading information regarding the COVID-19 pandemic. This dataset comprises 10,700 posts gathered from social media platforms, including genuine news tweets sourced from 14 official Twitter accounts and fake news tweets collected from social media and fact-checking websites. The study focused on a binary classification task aimed at distinguishing between real and false news, evaluating four machine learning models: Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (LR), and Gradient Boosting Decision Tree (GDBT). The SVM model exhibited the highest performance, achieving an F1-score of 93.32% on the test set. Shifath et al. [39] developed a multi-model ensemble by integrating eight pre-trained transformer architectures, each enhanced with supplementary layers. These models were fine-tuned and evaluated on a dataset related to COVID-19, with the RoBERTa-CNN model achieving an F1 score of 96.49%. Wani et al. [32] conducted a study to investigate various supervised text classification algorithms for identifying fake news related to COVID-19. The study employed models such as CNN, LSTM, and BERT. Additionally, the research explored the significance of unsupervised learning by using a pre-trained language model and distributed word representations with an unlabeled dataset of COVID-19-related tweets. The study concluded that the proposed model demonstrated enhanced accuracy in detecting fake news. Samadi et al. [40] sought to integrate various classifiers, including MLP, SLP, and CNN, with several text representation models such as RoBERTa, BERT, Funnel Transformer, and GPT2. The research team compared the performance of these models, analyzed the results, and subsequently contrasted the performance of the other models against the top-performing one. Additionally, they incorporated a Gaussian noise layer with the text representation layer and CNN classifier, hypothesizing that this addition would improve learning, particularly for COVID-19 and other datasets. In their research, Vijjali et al. [41] developed a two-phase automated system to tackle the detection of fake news related to COVID-19. This system utilizes a specialized machine learning model tailored for fake news identification. The initial phase involves a novel fact-checking algorithm to gather relevant facts concerning user claims on COVID-19 topics. The subsequent phase evaluates the credibility of these claims by measuring the semantic similarity between the claims and the facts derived from a manually compiled COVID-19 information repository. The study assessed various models, including those that employ traditional text features and advanced Transformer models such as BERT and ALBERT, which provide contextualized representations. The results demonstrated that the pipelines incorporating BERT and ALBERT consistently yielded the most favorable outcomes in both phases of the system.

Koloski et al. [42] investigated various methods and techniques for identifying fake news related to COVID-19. They employed several methods to glean useful information from the data, including developing custom features that captured the statistical distribution of characters and words in

tweets. Through the analysis of character n-grams and word-based features, they identified possible spatial arrangements and significant patterns. Additionally, they utilized different BERT-based models to capture contextual information and distinguish between fake and authentic COVID-19 news. Their evaluation showed that the distilBERT tokenizer provided the best performance, achieving a high F1 score of 97.05%. Ghayoomi and Mousavian [43] developed a Persian text corpus centered on COVID-19, with a focus on identifying and annotating instances of fake news. They developed a detection model that combines the cross-lingual language model XLM-RoBERTa with parallel convolutional neural networks. To enhance the model's performance, they applied knowledge transfer by incorporating an English COVID-19 dataset and a general Persian fake news dataset.

Vishwakarma et al. [44] introduced a novel framework known as WSCH-CNN, which employs Convolutional Neural Networks (CNN) for detecting fake news. This framework comprises a content model and a heading model, both designed to analyze linguistic patterns in news content to classify articles as real or fake. The models were evaluated using a range of datasets, including publicly available resources such as the Kaggle dataset and the fake news challenge dataset, as well as self-assembled real-world datasets that incorporate both text and multimedia data. The WSCH-CNN framework demonstrated high accuracy, achieving 85.06% on the multimedia dataset, 94.16% for the heading model, and 85.32% for the content model, thereby outperforming existing approaches in the field of fake news detection. Mehta et al. [45] introduced a classification approach based on BERT, a natural language processing framework. They fine-tuned BERT with domain-specific datasets and enhanced the model by incorporating human justification and metadata. The results demonstrated that BERT's contextual understanding is effective in classifying fake news, achieving significant improvements in binary classification and notable gains in six-label classification. This research highlights the importance of advanced techniques in combating the dissemination of fake news. Choudhary and Arora [46] utilized a linguistic model to represent news text for detecting fake news. This model captures syntactic, semantic, and readability features of the news content. Due to the computational demands of the linguistic model, the researchers employed a hierarchical neural network architecture to enhance feature representation. They assessed the effectiveness of their approach by comparing the results of sequential neural networks with those of other machine learning methods and LSTM-based models. The findings revealed that the hybrid neural network model outperformed the alternative methods.

Blackledge and Atapour-Abarghouei [47] employed transformer-based models to evaluate their efficacy in detecting fake news and their adaptability across diverse news topics and model architectures. The research findings indicated that these models are not inherently skilled at identifying news articles grounded in opinion and suspicion. To mitigate this limitation, the researchers introduced a two-step method: initially filtering out articles based on opinion and suspicion, followed by classifying the remaining content. The study demonstrates that this approach enhances the accuracy of the transformer-based models in detecting fake news.

The research by Ma et al. [48] presents a novel model, DC-CNN (Dual-channel CNN with Attention-pooling), aimed at detecting fake news. This model employs Skip-Gram and Fast-text methodologies to reduce noise in data and improve the handling of non-derived words. Additionally, it introduces a parallel dual-channel pooling layer as an alternative to the conventional CNN pooling layer. This layer consists of a Max-pooling channel to capture local information and an Attention-pooling channel equipped with a multi-head attention mechanism to extract contextual semantics and global dependencies. The model performance is evaluated using Covid-19-related fake news datasets, demonstrating its effectiveness in managing noisy data while maintaining a harmonious balance between global and local features.

Brown et al. [49] demonstrated that large language models (LLMs) can effectively identify misinformation, even without explicit training on the target datasets (zero-shot scenarios). Chen and Shu [50] investigated the risks of LLM-generated misinformation compared to human-written falsehoods, finding it more deceptive and harder to detect. They developed a taxonomy of LLM-generated misinformation and validated real-world creation methods. Their empirical analysis showed that both humans and automated detectors struggle more with AI-generated misinformation, suggesting greater societal risks. The study also discussed potential countermeasures.

Chen and Shu [51] investigated the dual role of LLMs in misinformation, highlighting their capacity as both detection tools and generators of

deceptive content. While LLMs can leverage world knowledge to counter misinformation, they also facilitate the mass production of convincing falsehoods. The study reviewed pre-LLM detection methods and current mitigation strategies, emphasizing the need for interdisciplinary collaboration to tackle emerging challenges. Wan et al. [52] proposed DELL, a framework enhancing LLM-based misinformation detection by integrating them into three stages: (1) generating simulated user-news interactions, (2) producing explanations for auxiliary tasks (e.g., sentiment analysis) and (3) merging expert predictions. Evaluated on seven datasets, DELL improved macro F1-scores up to 16.8% over state-of-the-art baselines, demonstrating the effectiveness of synthetic reactions and expert fusion. Garry et al. [53] highlighted the dangers of chat-based LLMs (e.g., ChatGPT) in spreading misinformation due to probabilistic outputs. These models exploit cognitive biases, making false claims appear credible. A harmful feedback loop emerges as AI-generated misinformation contaminates training data, potentially eroding trust in institutions. The study emphasized the urgent need to safeguard reliable information access. Liu et al. [54] provided a comprehensive tutorial on preventing and detecting LLM-generated misinformation. They categorized false content into factual errors and intentional deception, discussing prevention (e.g., AI alignment, retrieval-augmented generation) and detection methods (e.g., classifiers, watermarking). The paper also outlined the challenges of identifying AI-generated misinformation and proposed a comprehensive mitigation framework.

Leveraging the superior performance of deep neural models over traditional machine learning techniques, we develop advanced architectures to achieve state-of-the-art results in detecting COVID-19-related fake news. To situate our contributions within the evolving landscape of LLM-based misinformation detection, we conduct comprehensive evaluations comparing our approach with GPT-4o in zero-shot settings across English and Persian.

Our experiments reveal that while GPT-4o performs competitively in English, our model significantly outperforms it in [specific aspects, e.g., cross-lingual consistency, low-resource language adaptation (Persian), or domain-specific fine-tuning]. In later sections, we introduce our novel Hybrid CapsNet framework, which integrates RoBERTa for enriched text representation, and present empirical results demonstrating its advantages over both traditional methods and modern LLM-based approaches.

3. Hybrid CapsNet with Dynamic Transformer Model for COVID-19 Fake News Detection

This section presents our fake news detection model. The proposed model incorporates a dynamic pre-trained embedding model and two parallel classifiers. It leverages the benefits of both DCNN and CapsNet, two distinct neural network architectures previously employed as classifiers. We first review dynamic pre-trained language models, including BERT with dynamic embedding, which incrementally updates word embeddings during training. We then describe the classifiers used in the learning process.

Figure 1 illustrates the proposed model architecture. This architecture utilizes four parallel neural networks. Each network includes an n-gram convolutional layer and a CapsNet layer comprising a primary capsule layer, a convolutional capsule layer, and a feed-forward capsule layer, as previously described by Yang et al. [37]. Following global max-pooling and leaky ReLU activation, the outputs of the CNNs and CapsNet are concatenated. Subsequently, two dense layers process the concatenated output, producing the final prediction for the input news article's label. This architecture enables models to represent text meaningfully and extensively at various n-gram levels, depending on its length.

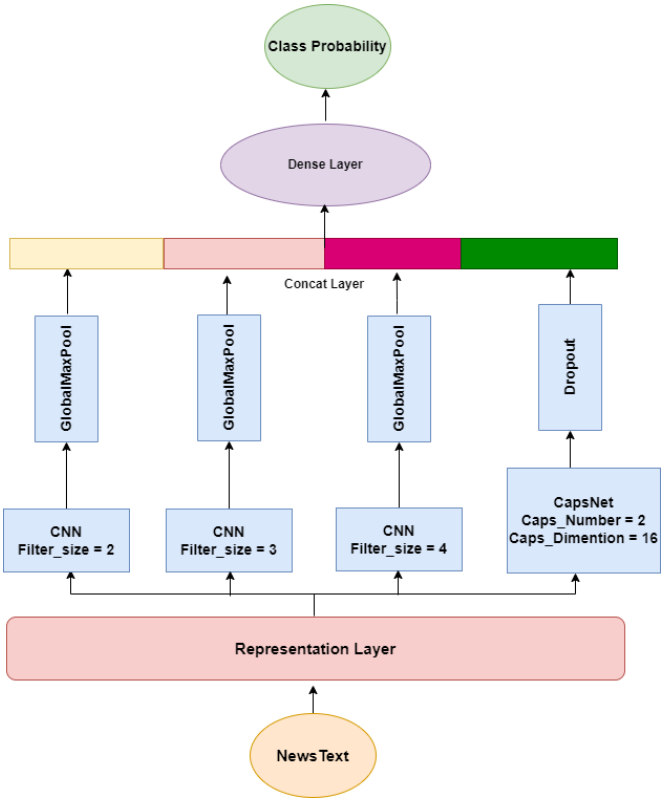

Figure 1. Proposed model for fake news detection.

3.1. Representation Layer

Kim [55] demonstrated the effectiveness of various training methods for word2vec in generating word vectors, showing superior performance compared to conventional pre-trained embeddings when integrated into a CNN model. Dynamic embedding initiates the learning process with pre-trained vectors, which are then fine-tuned during training for each specific task using task-specific data. Similarly, our proposed model utilizes dynamic Transformer Embedding, initializing transformer-based embeddings with pre-trained weights (e.g., BERT or similar models) and subsequently fine-tuning them during training for the fake news detection task. This allows the model to better adapt to the nuances of the task.

3.2. DCNN Layer

Figure 1 depicts the computational flow of the DCNN classifier. Zhong et al. [56] demonstrated that fake news detection can be approached using a standard text classification model consisting of an embedding layer, a one-dimensional convolutional layer, a max-pooling layer, and a prediction-based output layer [56]. Our proposed model draws inspiration from this idea, utilizing multiple parallel channels based on variable-size filters. Specifically, we incorporate three different filter sizes—two, three, and four—as n-gram convolutional layers for feature extraction.

3.3. CapsNet Layer

The proposed model employs CapsNet, consisting of two capsules of 16 dimensions followed by a leaky ReLU activation, as the parallelized neural network. Figure 2 illustrates the CapsNet layer used in the proposed model.

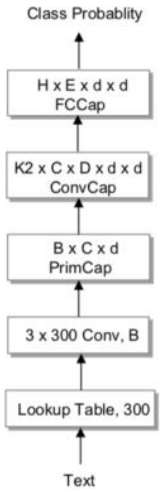

Figure 2. The architecture of capsule network proposed for text classification by Yang et al. [37].

3.4. Fully Connected Layer

Generally, dense layers operate linearly, with all inputs connected to all outputs by a weight system. To ensure the proposed model's inherent density, we utilize two dense layers. The first dense layer receives the output from the concatenation layer, while the second dense layer generates the final model output predictions.

4. Evaluation

4.1. Dataset

During the early stages of the Coronavirus pandemic, social media users disseminated an increasing volume of unconfirmed information and fake news regarding the disease. Motivated by this, researchers sought to collect data from social media and propose machine learning models for evaluation. In recent publications, Patwa et al. [38] presented a dataset named COVID-19, comprising tweets that include both fake and real news. This dataset comprises over 10,000 posts about COVID-19 outbreaks shared on social media. A summary of the dataset is provided in Table1.

Table 1. COVID-19 dataset statistics by Patwa et al. [38].

Dataset	Label		Total
	Real	Fake	
Validation	1120	1020	2140
Training	3360	3360	4420
Test	1120	1120	2140

Furthermore, another COVID-19 dataset for the Persian language was presented by Ghayoomi and Mousavian [43]. Table 2 shows the dataset's statistics.

Table 2. Persian COVID-19 dataset statistics by Ghayoomi and Mousavian [43].

Dataset	Label		Total
	Real	Fake	
Persian COVID-19	265	265	530

4.2. Experimental Setup

In the experiments, we utilize dynamic transformer models and dynamic RoBERTa to obtain word embeddings. We initialize the embedding layer for the COVID-19 and Persian COVID-19 datasets using 100-dimensional word vectors. Additionally, we employ dynamic embedding as described in Section 3.1. For both datasets, we use the Adam optimizer [57] and the ReLU activation function.

The hyperparameter settings for the COVID-19 datasets are: batch size 50, lambda 0.25, epochs 5, embedding size 100, and learning rate 0.001. For the Persian COVID-19 dataset, we choose batch size 50, lambda 0.7, epochs 5, embedding size 100, and learning rate 0.001.

4.3. Results

This section evaluates the proposed model using the COVID-19 dataset for both English and Persian languages. The results are compared to those of other baseline methods. To demonstrate the proposed model's superiority, additional evaluations were performed on two parallel layers separately. Finally, a series of experiments on the dataset are discussed in the discussion subsection.

4.4. Classification Results on The COVID-19 Dataset

Following the presentation of the COVID-19 dataset by Patwa et al. [38], various machine learning and deep learning models were evaluated for fake news detection using this dataset. Patwa et al. [38] used conventional machine learning models, including LR, SVM, DT, and GDBT. By using additional data for training, Shifath et al. [39] proposed an MLP connected to RoBERTa's pooled output. The study by Wani et al. [32] evaluated several models, including a softmax layer connected to BERT for prediction. Samadi et al. [40] demonstrated the superiority of a CNN connected to RoBERTa's pooled output over prior models. To explore the zero-shot capabilities of modern large language models (LLMs), we also evaluated GPT-4o on both the English and Persian COVID-19 fake news datasets without any task-specific fine-tuning. To assess the robustness of the proposed model, we performed 5-fold cross-validation. The results are summarized in Table 3, which reports the mean performance metrics along with 95% confidence intervals (CI).

Table 3. 5-fold cross-validation results for the proposed model on the COVID-19 test set.

Metric	Mean	95%CI Lower	95%CI Upper
Accuracy	97.34	96.85	97.83
Precision	97.21	96.68	97.74
Recall	97.38	96.90	97.86
F1 Score	97.29	96.80	97.78

Table 4 compares the results of our proposed model with the state-of-the-art models, including GPT-4o, on the English COVID-19 test set. While GPT-4o achieved near-perfect precision (99.90%), its recall was significantly lower (69.50%), resulting in a moderate F1-score (81.92%). This suggests that LLMs like GPT-4o, despite their strong linguistic capabilities, may struggle with recall in zero-shot fake news detection scenarios. Based on the F1-score, our proposed Hybrid CapsNet outperforms all baseline models, including GPT-4o. Moreover, in fake news detection, recall is critical since missing fake news articles can have serious consequences. As shown in Table 4, our model achieves the highest recall (97.38%) among all competitors.

Table 4. Comparison of proposed model result with the result of other models on the COVID-19 test set.

Model	Acc	Prec	Rec	F1
DT [42]	85.37	85.47	85.37	85.39
LR [42]	91.96	92.01	91.96	91.96
SVM [42]	93.32	93.33	93.32	93.32
GDBT [42]	86.96	87.24	86.96	86.96
RoBERTa-MLP [48]	96.68	97.12	95.880	96.49
BERT-MLP [56]	95.79	98.94	92.15	95.43
RoBERTa-CNN [46]	**97.43**	98.30	96.27	97.27
GPT-4o (LLM)	84.75	**99.90**	69.50	81.92
Proposed Model	97.34	97.21	**97.38**	**97.29**

Table 5 compares the results of our proposed model with state-of-the-art models on the Persian COVID-19 test set. GPT-4o achieved the highest precision (79.14%) but exhibited lower recall (45.74%) and F1-score (57.97%) compared to our model. The proposed Hybrid CapsNet attains the best performance in accuracy (72.73%), recall (88.76%), and F1-score (73.83%), demonstrating its robustness across languages.

Table 5. Comparison of proposed model result with the result of other models on the COVID-19 Persian test set.

Model	Acc	Prec	Rec	F1
XLMRoBERTa [18]	-	75.98	64.97	69.74
GPT-4o (LLM)	66.84	**79.14**	45.74	57.97
Proposed Model	**72.73**	66.58	**88.76**	**73.83**

The results confirm that the Hybrid CapsNet model outperforms state-of-the-art baselines, including zero-shot LLMs, in terms of F1-score. Additionally, our model's superior recall highlights its effectiveness in minimizing missed detections

of fake news, a crucial factor in real-world applications.

4.5. Performance of Parallel Layers

Tables 6 and 7 show the proposed model's performance compared to the two parallel models. The results show that the baseline models perform better when they use different CNN feature extractors and CapsNet, which maintain detailed information about each object's location and pose throughout the network. The best result is achieved when both models are used together.

Table 6. Comparison of proposed model result with the result of parallel layers on the COVID-19 test set.

Model	Acc	Prec	Rec	F1
RoBERTa-CapsNet	93.22	93.92	91.91	96.31
RoBERTa-CNN	**97.43**	**98.30**	96.27	97.27
Proposed Model	97.34	97.21	**97.38**	**97.29**

Table 7. Comparison of proposed model result with the result of parallel layers on the Persian COVID-19 test set.

Model	Acc	Prec	Rec	F1
ParsBERT-CapsNet	65.24	59.05	**97.19**	71.39
ParsBERT-DCNN	69.52	63.83	87.31	71.32
Proposed Model	**72.73**	**66.58**	88.76	**73.83**

4.5.1. Ablation Study on The COVID-19 Dataset

The ablation studies presented in Tables 8 and 9 effectively demonstrate the individual contributions of the margin loss(ML) and CapsNet components to the overall performance of the DCNN-CapsNet COVID-19 fake news detection model. Let's discuss the results of each study separately: Table 8 compares a model with DCNN and CapsNet against the same model augmented with margin loss. The results clearly indicate a positive impact from incorporating the margin loss. Model B, which includes the margin loss, achieves significantly higher accuracy (97.34% vs 96.24%), precision (97.21% vs. 96.50%), recall (97.38% vs. 95.80%), and F1-score (97.29% vs 96.15%). This improvement suggests that the margin loss helps to better discriminate between true and false fake news, leading to a more robust and accurate classification. The increase across all metrics demonstrates a consistent improvement in performance, not just a marginal gain in one aspect. This improvement is likely due to the margin loss's ability to encourage larger margins between classes in the feature space, making the model less susceptible to noisy data or ambiguous examples.

Table 8. Ablation study: impact of margin-loss(ML) on the COVID-19 testset performance.

Model Configuration	Acc	Prec	Rec	F1
DCNN + CapsNet	96.24	96.50	95.80	96.15
DCNN + CapsNet + ML	**97.34**	**97.21**	**97.38**	**97.29**

Table 9 evaluates the effect of incorporating CapsNet. Model (C: DCNN + ML) utilizes DCNN and margin loss, whereas Model (D: DCNN + CapsNet + ML) incorporates CapsNet into this configuration. Again, the addition of CapsNet (Model D) leads to a substantial performance improvement compared to the model without it (Model C). The accuracy increases from 95.78% to 97.34%, precision from 95.90% to 97.21%, recall from 95.50% to 97.38%, and the F1-score from 95.70% to 97.29%. This significant boost in performance highlights the crucial role of CapsNet in enhancing the model's ability to capture the complex relationships and hierarchical structures within the fake news data. CapsNet's inherent ability to handle pose variations and part-whole relationships likely contributes to its effectiveness in this context, allowing the model to better understand the nuances of the text and identify subtle indicators of fake news.

Table 9. Ablation study: impact of CapsNet on the COVID-19 testset performance.

Model Configuration	Acc	Prec	Rec	F1
DCNN + ML	95.78	95.90	95.50	95.70
DCNN + CapsNet + ML	**97.34**	**97.21**	**97.38**	**97.29**

5. Discussion

5.1. English COVID-19 Dataset

This section further analyzes the training set of the COVID-19 dataset for real and fake news labels. Figure 3 and Figure 4 show word clouds for the real and fake news of the training set, respectively, after omitting stopwords. From the word clouds and the most frequent words, we observe an overlap of important words across fake and real news. Therefore, for further analysis, we list the ten most frequent words in real and fake news after removing stopwords in Table 10.

If we disregard the common words shared by the two groups, we find that among the fake news, words related to sensitive quotes and reports, such as those about vaccines, the coronavirus, and the names of politicians, are more frequent. Additionally, statistics about the number of infected cases and related terms in this domain, such as “number,” “death,” “day,” and the names of countries, are repeated more frequently in real news.

Table 10. List of the ten most frequent words in English real and fake news after removing stopwords.

Fake news	Real News
COVID	case
coronavirus	COVID
people	new
claim	state
trump	test
virus	number
say	death
vaccine	India
new	total
case	day

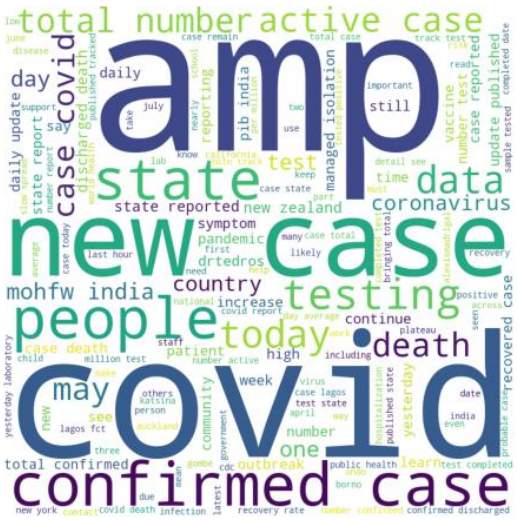

Figure 3. Word cloud of real news.

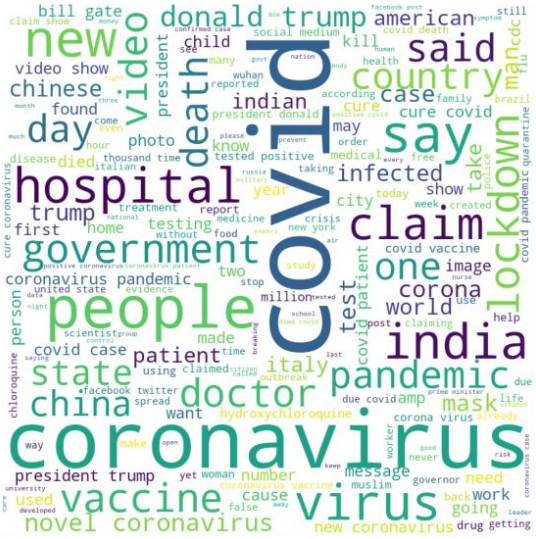

Figure 4. Word cloud of fake news.

If we disregard the common words shared by the two groups, we find that among the fake news, words related to sensitive quotes and reports, such as those about vaccines, the coronavirus, and the names of politicians, are more frequent. Additionally, statistics about the number of infected cases and related terms in this domain, such as "number," "death," "day," and the names of countries, are repeated more frequently in real news.

5.2. Persian COVID-19 Dataset

For the Persian COVID-19 dataset, we repeated the same analysis. Figure 5 and Figure 6, respectively, show word clouds for real news and fake news of the training set after omitting stopwords. From the word clouds and the most frequent words, we observe an overlap of important words across fake and real news. Therefore, for further analysis, we list the translations of the ten most frequent words in the Persian real and fake news after removing stopwords in Table 11.

Figure 5. Word cloud of real Persian news.

Figure 6. Word cloud of fake Persian news.

Table 11. List of the ten most frequent words in Persian real and fake news after removing stopwords (the words are translated to English for better understanding).

Fake news	Real News
corona	corona
country	virus
virus	Iran
disease	year
case	people
COVID	disease
people	country
thousand	news
vaccine	announcement
healthy	day

If we disregard the common words between the two groups, we find that among the fake news, words related to vaccines and cases are more frequent. Additionally, statistics about the number of infected cases and related terms in this domain, such as years and country names, are repeated more frequently in real news.

5.3. Prevalence and Patterns of Fake News Sharing

The spread of COVID-19 fake news in English has been extensively documented, with studies highlighting the role of social media platforms in amplifying false claims. For instance, research on English-language fake news reveals that fake news often leverages emotional and sensationalist language to increase engagement, with topics such as virus origins, treatments, and conspiracy theories dominating the discourse [58]. In contrast, Persian-language fake news, while less studied, exhibits unique cultural and linguistic nuances. For example, Persian-speaking communities have been observed to share fake news related to traditional remedies and religious interpretations of the pandemic, reflecting localized beliefs and practices [59].

5.4. Thematic and Emotional Characteristics

A cross-lingual analysis of COVID -19 fake news reveals distinct thematic and emotional characteristics. In English, fake news frequently revolves around political narratives, such as the alleged role of governments or corporations in the pandemic, often accompanied by negative emotional tones like fear and anger [60]. Conversely, Persian-language fake news tends to emphasize communal and religious themes, with emotional undertones of hope and faith, particularly in discussions about divine intervention or traditional healing methods [59]. These differences highlight the importance of culturally tailored approaches to combat fake news.

5.5. Impact of Social Media Platforms

Social media platforms play a pivotal role in the dissemination of fake news in both languages [61]. English-language fake news is predominantly shared on platforms such as Twitter and Facebook, where algorithmic amplification and viral content contribute to its rapid dissemination [58]. In Persian-speaking communities, platforms such as Telegram and Instagram are more influential, with encrypted messaging apps facilitating the unchecked circulation of false information [59]. This divergence highlights the need for platform-specific interventions to address fake news in different linguistic contexts.

5.6. Statistical Insights and Comparative Metrics

Quantitative studies provide valuable insights into the scale of fake news sharing. For example, a comparative analysis of English and Chinese fake news found that English-language fake news was shared 1.5 times more frequently than its Chinese counterpart, attributed to the higher prevalence of English on global social media platforms [58]. While similar statistics for Persian-language fake news are limited, anecdotal evidence suggests that the sharing rate is influenced by the level of digital literacy and access to verified information sources [59].

5.7. Strategies for Combating Fake News

Addressing fake news in both Persian and English requires a multifaceted approach. For English-speaking audiences, leveraging fact-checking organizations and promoting media literacy have proven effective [60]. In Persian-speaking communities, collaborations with religious leaders and community influencers can enhance the credibility of accurate information [59]. Additionally, cross-lingual AI models, such as the CrossFake framework, offer promising solutions for detecting and mitigating fake news in low-resource languages like Persian by leveraging high-resource language datasets [62].

5.8. Different Methods of Fake News Sharing with COVID-19 Subject

Apuke and Omar [11] proposed different methods of fake news sharing with COVID-19 subject on social media. This study's model was developed based on the U\&G theory and previous studies and includes six sharing methods: entertainment,

socialization, pastime, altruism, information seeking, and information sharing.

For further experiments on the subjectivity of fake news, fake news with subjectivity was extracted and classified. In Figure 7, we can see that most fake news with high subjectivity is shared for informational purposes on social media. Additionally, socialization is another common method used in the dissemination of fake news. As a result, one method of spreading fake news is to share interesting information and then use members to re-share that news on social networks. Fake news instances were categorized based on their level of subjectivity (high or low) and their alignment with the aforementioned sharing motivations. The results indicate that fake news with high subjectivity is most frequently shared for informational purposes, as users often perceive such content as relevant or important to their social networks. Additionally, socialization emerged as a significant factor, where fake news is shared to foster interactions and engage with others, especially in emotionally charged or controversial contexts.

To analyze the model's predictions, we evaluated its ability to correctly classify fake news across these categories. The model demonstrated strong performance in identifying fake news shared for information seeking and information sharing, as it effectively leveraged linguistic features such as subjective language, emotional tone, and context-specific keywords. The model also exhibited a high degree of accuracy in detecting fake news shared for socialization, which often involves emotionally resonant or sensational content.

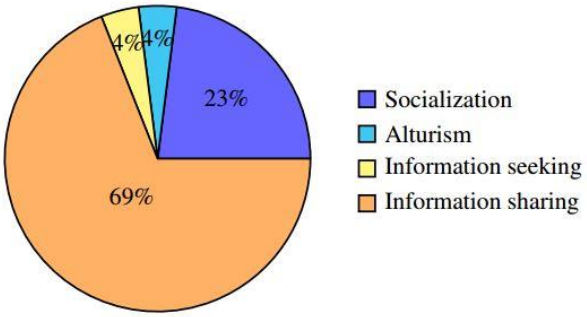

Figure 3. Distribution of different sharing methods of fake news on COVID-19.

6. Conclusion and Future Work

This paper proposes a novel architecture, Hybrid CapsNet with dynamic transformer and margin loss, for detecting COVID -19 fake news in English and Persian languages. The proposed architecture utilizes four parallel neural networks, comprising three different n-gram convolutional layers for feature extraction and a CapsNet layer that includes a primary capsule layer, a convolutional capsule layer, and a feed-forward capsule layer. By using a dynamic transformer that incrementally updates and refines the word embeddings during the training phase, and combining the outputs of CNNs and CapsNet, our model achieves high accuracy in detecting fake news. Our proposed architecture enables models to learn more meaningful and extensive text representations at different n-gram levels, depending on the text length. We evaluated our architecture using two recently introduced COVID -19 datasets in the field. Our experimental results, in terms of F1-score, demonstrate that the Hybrid CapsNet model outperforms state-of-the-art baselines in detecting fake news. In conclusion, our proposed model can effectively detect COVID -19 fake news in both English and Persian languages and has the potential to contribute to the fight against fake news during the pandemic. Future work could explore the application of our proposed architecture to other languages and domains, as well as the incorporation of additional features or data sources to further enhance the model's performance.

References

[1] X. Zhang and A. A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion," *Inf. Process. Manag.*, vol. 57, p. 102025, 2019.

[2] K. Ma, C. Tang, W. Zhang, B. Cui, K. Ji, Z. Chen, and A. Abraham, "DC-CNN: Dual-channel convolutional neural networks with attention-pooling for fake news detection," *Appl. Intell.*, pp. 1–16, 2022.

[3] A. Bondielli and F. Marcelloni, "A survey on fake news and rumour detection techniques," *Inf. Sci.*, vol. 497, pp. 38–55, 2019.

[4] X. Zhou and R. Zafarani, "Fake news: A survey of research, detection methods, and opportunities," *arXiv:1812.00315*, 2018.

[5] A. Bovet and H. A. Makse, "Influence of fake news in Twitter during the 2016 US presidential election," *Nat. Commun.*, vol. 10, no. 1, pp. 1–14, 2019.

[6] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer, "Fake news on Twitter during the 2016 US presidential election," *Science*, vol. 363, no. 6425, pp. 374–378, 2019.

[7] J. Kim, J. Aum, S. Lee, Y. Jang, E. Park, and D. Choi, "FibVid: Comprehensive fake news diffusion dataset during the COVID-19 period," *Telemat. Inform.*, vol. 64, p. 101688, 2021.

[8] G. Shrivastava, P. Kumar, R. P. Ojha, P. K. Srivastava, S. Mohan, and G. Srivastava, "Defensive modeling of fake news through online social

networks," *IEEE Trans. Comput. Soc. Syst.*, vol. 7, no. 5, pp. 1159–1167, 2020.

[9] A. Depoux, S. Martin, E. Karafillakis, R. Preet, A. Wilder-Smith, and H. Larson, "The pandemic of social media panic travels faster than the COVID-19 outbreak," *J. Travel Med.*, vol. 27, no. 3, 2020.

[10] L. Li, Q. Zhang, X. Wang, J. Zhang, T. Wang, T.-L. Gao, et al., "Characterizing the propagation of situational information in social media during COVID-19 epidemic: A case study on Weibo," *IEEE Trans. Comput. Soc. Syst.*, vol. 7, no. 2, pp. 556–562, 2020.

[11] O. D. Apuke and B. Omar, "Fake news and COVID-19: Modelling the predictors of fake news sharing among social media users," *Telemat. Inform.*, vol. 56, p. 101475, 2021.

[12] G. Pennycook, J. McPhetres, Y. Zhang, J. G. Lu, and D. G. Rand, "Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention," *Psychol. Sci.*, vol. 31, no. 7, pp. 770–780, 2020.

[13] CNN News, "Nigeria records chloroquine poisoning after Trump endorses it for coronavirus treatment," 2020. [Online]. Available: http://ediioncnn.com2020/03/23/africa/chloroquine-tum-nigeria-inl/ndex.html. [Accessed: May 18, 2023].

[14] N. Karimi and J. Gambrell, "Hundreds die of poisoning in Iran as fake news suggests methanol cure for virus," *Times of Israel*, 2020.

[15] R. Bal, S. Sinha, S. Dutta, R. Joshi, S. Ghosh, and R. Dutt, "Analysing the extent of misinformation in cancer related tweets," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 14, pp. 924–928, 2020.

[16] S. A. Memon and K. M. Carley, "Characterizing COVID-19 misinformation communities using a novel Twitter dataset," *CEUR Workshop Proceedings*, vol. 2699, 2020.

[17] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, et al., "RoBERTa: A robustly optimized BERT pretraining approach," *arXiv:1907.11692*, 2019.

[18] U. Naseem, I. Razzak, M. Khushi, P. W. Eklund, and J. Kim, "COVIDSenti: A large-scale benchmark Twitter data set for COVID-19 sentiment analysis," *IEEE Trans. Comput. Soc. Syst.*, vol. 8, no. 4, pp. 1003–1015, 2021.

[19] P. Gupta, S. Kumar, R. R. Suman, and V. Kumar, "Sentiment analysis of lockdown in India during COVID-19: A case study on Twitter," *IEEE Trans. Comput. Soc. Syst.*, vol. 8, no. 4, pp. 992–1002, 2020.

[20] Y. Li, Q. Pan, T. Yang, S. Wang, J. Tang, and E. Cambria, "Learning word representations for sentiment analysis," *Cogn. Comput.*, vol. 9, no. 6, pp. 843–851, 2017.

[21] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv:1301.3781*, 2013.

[22] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Trans. Assoc. Comput. Linguist.*, vol. 5, pp. 135–146, 2017.

[23] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543, 2014.

[24] A. Khikmatullaev, J. Lehmann, and K. Singh, "Capsule neural networks for text classification," Ph.D. dissertation, Apr. 2019.

[25] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 4171–4186, 2019.

[26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., "Attention is all you need," in *Advances in Neural Information Processing Systems*, pp. 5998–6008, 2017.

[27] M. Farahani, M. Gharachorloo, M. Farahani, and M. Manthouri, "ParsBERT: Transformer-based model for Persian language understanding," *Neural Process. Lett.*, vol. 53, pp. 3831–3847, 2021.

[28] J. A. Nasir, O. S. Khan, and I. Varlamis, "Fake news detection: A hybrid CNN-RNN based deep learning approach," *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 1, p. 100007, 2021.

[29] M. H. Goldani, R. Safabakhsh, and S. Momtazi, "Convolutional neural network with margin loss for fake news detection," *Inf. Process. Manag.*, vol. 58, no. 1, p. 102418, 2021.

[30] K. L. Tan, C. P. Lee, and K. M. Lim, "FN-Net: A deep convolutional neural network for fake news detection," in *2021 9th International Conference on Information and Communication Technology (ICOICT)*, pp. 331–336, 2021.

[31] R. K. Kaliyar, A. Goswami, P. Narang, and S. Sinha, "FNDNet–A deep convolutional neural network for fake news detection," *Cogn. Syst. Res.*, vol. 61, pp. 32–44, 2020.

[32] A. Wani, I. Joshi, S. Khandve, V. Wagh, and R. Joshi, "Evaluating deep learning approaches for COVID19 fake news detection," in *International Workshop on Combating Online Hostile Posts in Regional Languages during Emerging Situations*, pp. 153–163, 2021.

[33] Q. Jiang, L. Chen, R. Xu, X. Ao, and M. Yang, "A challenge dataset and effective models for aspect-based sentiment analysis," in *Proceedings of the 2019

Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 6280–6285, 2019.

[34] M. H. Goldani, S. Momtazi, and R. Safabakhsh, "Detecting fake news with capsule neural networks," *Appl. Soft Comput.*, vol. 101, p. 106991, 2021.

[35] B. Palani, S. Elango, and V. Viswanathan K, "CB-Fake: A multimodal deep learning framework for automatic fake news detection using capsule neural network and BERT," *Multimed. Tools Appl.*, pp. 1–34, 2021.

[36] S. Sridhar and S. Sanagavarapu, "Fake news detection and analysis using multitask learning with BiLSTM CapsNet model," in *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pp. 905–911, 2021.

[37] M. Yang, W. Zhao, J. Ye, Z. Lei, Z. Zhao, and S. Zhang, "Investigating capsule networks with dynamic routing for text classification," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 3110–3119, 2018.

[38] P. Patwa, S. Sharma, S. Pykl, V. Guptha, G. Kumari, M. S. Akhtar, et al., "Fighting an infodemic: COVID-19 fake news dataset," in *International Workshop on Combating Online Hostile Posts in Regional Languages during Emerging Situations*, pp. 21–29, 2021.

[39] S. Shifath, M. F. Khan, M. Islam, et al., "A transformer based approach for fighting COVID-19 fake news," *arXiv:2101.12027*, 2021.

[40] M. Samadi, M. Mousavian, and S. Momtazi, "Deep contextualized text representation and learning for fake news detection," *Inf. Process. Manag.*, vol. 58, no. 6, p. 102723, 2021.

[41] R. Vijjali, P. Potluri, S. Kumar, and S. Teki, "Two stage transformer model for COVID-19 fake news detection and fact checking," in *Proceedings of the 3rd NLP4IF Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda*, pp. 1–10, 2020.

[42] B. Koloski, T. Stepišnik-Perdih, S. Pollak, and B. Škrlj, "Identification of COVID-19 related fake news via neural stacking," in *International Workshop on Combating Online Hostile Posts in Regional Languages during Emerging Situations*, pp. 177–188, 2021.

[43] M. Ghayoomi and M. Mousavian, "Deep transfer learning for COVID-19 fake news detection in Persian," *Expert Syst.*, vol. 39, no. 8, p. e13008, 2022.

[44] D. K. Vishwakarma, P. Meel, A. Yadav, and K. Singh, "A framework of fake news detection on web platform using ConvNet," *Soc. Netw. Anal. Min.*, vol. 13, no. 1, p. 24, 2023.

[45] D. Mehta, A. Dwivedi, A. Patra, and M. Anand Kumar, "A transformer-based architecture for fake news classification," *Soc. Netw. Anal. Min.*, vol. 11, pp. 1–12, 2021.

[46] A. Choudhary and A. Arora, "Linguistic feature based learning model for fake news detection and classification," *Expert Syst. Appl.*, vol. 169, p. 114171, 2021.

[47] C. Blackledge and A. Atapour-Abarghouei, "Transforming fake news: Robust generalisable news classification using transformers," *arXiv:2109.09796*, 2021.

[48] K. Ma, C. Tang, W. Zhang, B. Cui, K. Ji, Z. Chen, and A. Abraham, "DC-CNN: Dual-channel convolutional neural networks with attention-pooling for fake news detection," *Appl. Intell.*, vol. 53, no. 7, pp. 8354–8369, 2023.

[49] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, et al., "Language models are few-shot learners," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 1877–1901, 2020.

[50] C. Chen and K. Shu, "Can LLM-generated misinformation be detected?" in *NeurIPS 2023 Workshop on Regulatable ML*, 2023. [Online]. Available: http://openreview.net/foru?id=Mzu0NIMvvh. [Accessed: May 18, 2023].

[51] C. Chen and K. Shu, "Combating misinformation in the age of LLMs: Opportunities and challenges," *AI Mag.*, vol. 45, no. 3, pp. 354–368, 2024.

[52] H. Wan, S. Feng, Z. Tan, H. Wang, Y. Tsvetkov, and M. Luo, "DELL: Generating reactions and explanations for LLM-based misinformation detection," in *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 2637–2667, 2024.

[53] M. Garry, W. M. Chan, J. Foster, and L. A. Henkel, "Large language models (LLMs) and the institutionalization of misinformation," *Trends Cogn. Sci.*, 2024.

[54] A. Liu, Q. Sheng, and X. Hu, "Preventing and detecting misinformation generated by large language models," in *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 3001–3004, 2024.

[55] Y. Kim, "Convolutional neural networks for sentence classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1746–1751, 2014.

[56] B. Zhong, X. Xing, P. Love, X. Wang, and H. Luo, "Convolutional neural network: Deep learning-based classification of building quality problems," *Adv. Eng. Inform.*, vol. 40, pp. 46–57, 2019.

[57] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations*, 2015.

[58] G. Pennycook, J. McPhetres, Y. Zhang, and D. G. Rand, "The COVID-19 misinformation pandemic: A call for action," *Nat. Hum. Behav.*, vol. 4, no. 5, pp. 460–461, 2020.

[59] S. Shahsavari, P. Holur, T. Wang, and T. R. Tangherlini, "Misinformation and disinformation in Persian-language social media: A case study of COVID-19," *J. Med. Internet Res.*, vol. 22, no. 12, p. e22501, 2020.

[60] J. S. Brennen, F. M. Simon, P. N. Howard, and R. K. Nielsen, "Types, sources, and claims of COVID-19 misinformation," *Reuters Institute for the Study of Journalism*, 2020.

[61] F. Hafezi and M. Khodabakhsh, "Coronavirus incidence rate estimation from social media data in Iran," *J. Artif. Intell. Data Min.*, vol. 11, no. 2, pp. 315–329, 2023.

[62] X. Zhang and A. A. Ghorbani, "CrossFake: A cross-lingual framework for detecting fake news in low-resource languages," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1234–1245, 2021.

تشخیص چندزبانه اخبار جعلی مرتبط با کووید-۱۹ با استفاده از شبکه‌های کپسولی ترکیبی

محمدهادی گلدانی، سعیده ممتازی* و رضا صفابخش

دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر (پلیتکنیک تهران)، تهران، ایران.

ارسال ۲۰۲۵/۰۱/۲۱؛ بازنگری ۲۰۲۵/۰۵/۲۶؛ پذیرش ۲۰۲۵/۰۷/۱۴

چکیده:

استفاده گسترده از پایگاه‌های مبتنی بر وب و شبکه‌های اجتماعی منجر به افزایش مصرف اخبار شده است. برای کاهش تأثیر اخبار جعلی بر تصمیمات مرتبط با سلامت کاربران، توسعه مدل‌های یادگیری ماشین که قادر به تشخیص و مقابله خودکار با اخبار جعلی باشند، امری حیاتی است. در این مقاله، ما یک مدل چندزبانه نوآورانه به نام Hybrid CapsNet را پیشنهاد می‌کنیم که از یک مدل مبدل پویا برای تشخیص اخبار جعلی مرتبط با کووید-۱۹ در هر دو زبان انگلیسی و فارسی استفاده می‌کند. مدل ما شامل دو مدل نمایش از پیش آموزش‌دیده پویا است که به‌صورت افزایشی، امبدینگ کلمات را در طول فاز آموزش به‌روزرسانی می‌کنند: روبرتا پویا برای انگلیسی و پارس‌برت پویا برای فارسی. علاوه بر این، مدل ما از دو دسته‌بندی موازی با یک تابع هزینه جدید به نام Margin Loss بهره می‌برد. با استفاده از مبدل پویا و ترکیب شبکه عصبی پیچشی عمیق (DCNN) **و** شبکه عصبی کپسولی (CapsNet)، عملکرد بهتری نسبت به مدل‌های پایه پیشرفته فعلی به دست می‌آوریم. برای ارزیابی مدل پیشنهادی، از دو مجموعه‌داده اخیر کووید-۱۹ به زبان‌های انگلیسی و فارسی استفاده شده است. نتایج ما بر اساس معیارF1-Score ، اثربخشی مدل Hybrid CapsNet را نشان می‌دهد. مدل ما از مدل‌های پایه موجود پیشی می‌گیرد، که نشان‌دهنده قابلیت آن به‌عنوان ابزاری مؤثر برای تشخیص و مقابله با اخبار جعلی مرتبط با کووید-۱۹ در چندین زبان است. به‌طور کلی، مطالعه ما بر اهمیت توسعه مدل‌های مؤثر یادگیری ماشین برای مقابله با اخبار جعلی در رویدادهای بحرانی مانند همه‌گیری کووید-۱۹ تأکید می‌کند. مدل پیشنهادی پتانسیل کاربرد در زبان‌ها و حوزه‌های دیگر را دارد و می‌تواند به‌عنوان ابزاری ارزشمند برای محافظت از سلامت و ایمنی عمومی مورد استفاده قرار گیرد.

کلمات کلیدی: تشخیص اخبار جعلی، مبدل پویا، شبکه‌های عصبی پیچشی عمیق، شبکه‌های عصبی کپسولی.