Deep transfer learning for COVID-19 fake news detection in Persian

Masood Ghayoomi1 | Maryam Mousavian2

1Faculty of Linguistics, Institute for Humanities and Cultural Studies, Tehran, Iran
2Computer Engineering Department, Amirkabir University of Technology, Tehran, Iran

Correspondence
Masood Ghayoomi, Faculty of Linguistics, Institute for Humanities and Cultural Studies, Tehran, Iran. Email: m.ghayoomi@ihcs.ac.ir

Abstract
The spread of fake news on social media has increased dramatically in recent years. Hence, fake news detection systems have received researchers' attention globally. During the COVID-19 outbreak in 2019 and the worldwide epidemic, the importance of this issue becomes more apparent. Due to the importance of the issue, a large number of researchers have begun to collect English datasets and to study COVID-19 fake news detection. However, there are a large number of low-resource languages, including Persian, that cannot develop accurate tools for automatic COVID-19 fake news detection due to the lack of annotated data for the task. In this article, we aim to develop a corpus for Persian in the domain of COVID-19 where the fake news is annotated and to provide a model for detecting Persian COVID-19 fake news. With the impressive advancement of multilingual pre-trained language models, the idea of cross-lingual transfer learning can be proposed to improve the generalization of models trained with low-resource language datasets. Accordingly, we use the state-of-the-art deep cross-lingual contextualized language model, XLM-RoBERTa, and the parallel convolutional neural networks to detect Persian COVID-19 fake news. Moreover, we use the idea of knowledge transferring across-domains to improve the results by using both the English COVID-19 dataset and the general domain Persian fake news dataset. The combination of both cross-lingual and cross-domain transfer learning has outperformed the models and it has beaten the baseline by 2.39% significantly.

KEYWORDS
contextualized text representation, COVID-19, deep neural network, fake news detection, transfer learning

1 | INTRODUCTION

The COVID-19 pandemic since 2019 has imposed many challenges to the communities around the world and their health systems. This event might be the first pandemic where technology and social media are used widely to keep people connected and informed about it on a massive scale. This might cause an infodemic, an overabundance of information. Infodemic causes to disseminate a large amount of correct or incorrect information that might undermine public health. One of the most important challenges in an infodemic situation is the widespread dissemination of fake news and rumours about medical and political information, such as prevention, treatment, causes, and consequences of the disease on humans' life (Al-Zaman, 2021; Barua et al., 2020; Swire-Thompson & Lazer, 2020). Dissemination of incorrect information among people is...
harmful to threaten people’s physical and mental health; as a result, it might lead to a higher prevalence of the disease. The general director of World Health Organization at the Munich Security Conference on 15 February 2020 proclaimed that ‘We’re not just fighting an epidemic; we’re fighting an infodemic’ (Zarocostas, 2020).

van der Linden et al. (2020) have specified some published fake news about the epidemic, including marketing for counterfeit treatments such as gargling with lemon or salt water, injecting bleach, or 5G cellular network that caused or exacerbated COVID-19 symptoms. The movie ‘Plandemic’ broadcasted online on 4 May 2020 attracted millions of views; and it quickly became one of the most widespread examples of misinformation about the COVID-19. This video offered dangerous health advice, for example, it falsely suggested that wearing a mask activates the COVID-19.

The mass media, health-care organization, and all community-based organizations have a duty to provide a platform for the transmission and dissemination of credible public health and political information and messages through their participation. But due to the infodemic and producing a large amount of information, the channel of transmitting correct information to people might not function properly. To this end, filtering the information is required, which cannot be done manually due to the large amount of data to be produced. Machinery methods that utilize natural language processing techniques can help to remove unreliable online content from social networks.

Due to the importance of the issue, researchers have provided various datasets based on different resources; including multilingual datasets (Lopez et al., 2020; Qazi et al., 2020), and mono-lingual datasets, including English, (Cui & Lee, 2020; Memon & Carley, 2020; Patwa et al., 2020).

In this article, we bring Persian, as a low-resource language, into consideration and improve the general ability of the model for Perdian COVID-19 fake news detection by leveraging a high-resource language dataset, that is, English, and a general domain Persian fake news dataset. The proposed model is based on language and domain transfer learning. We implement a parallel Convolutional Neural Networks (CNN) model and use the state-of-the-art deep cross-lingual contextualized representation model, XLM-RoBERTa (Conneau et al., 2020), for precisely detecting Persian COVID-19 fake news.

The rest of the article is organized as follows: Section 2 reviews related works about COVID-19 fake news detection. Section 3 presents the proposed model in this article. The datasets used in our research are described in Section 4. Section 5 reports the experimental results and analyzes errors; and finally, the article is summarized in Section 6.

2 RELATED WORKS

Recently neural network models have achieved researchers’ attention. Computational hardware capabilities and the availability of large amount of data have made it possible to propose advanced neural models in the text representation and text classification. The classification and vectorization based on this approach can be used in various text mining tasks. One of the recent hot topics that uses natural language processing methods is detecting misinformation, that is also called fake news or rumours. Liu, Wu, et al. (2019) proposed a two-stage model based on the Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) model to detect fake news. In the first stage, their proposed model classified news into two categories as the coarse-grained labels using statements and extra information. The predicted coarse-grained labels together with the statements and extra information were used for fine-grained label prediction. They evaluated the proposed model on the LIAR (Wang, 2017) dataset.

Huang and Chen (2020) proposed an ensemble model in fake news detection and investigated the cross-domain intractability. In this method, a combination of models called embedding Long Short-Term Memory (LSTM) network, depth LSTM, linguistic inquiry and word count with a CNN classifier as well as n-gram with a CNN classifier was implemented to find the appropriate features of the news. The self-adaptive harmony search algorithm was used to determine the optimized weights of the ensemble learning model.

Kaliyar et al. (2020) implemented the FNDNet model. This model was composed of three parallel convolutional networks and max-pooling layers followed by two dense layers and the output layer. The GloVe word representation proposed by Pennington et al. (2014) was used in the CNN model. Their experiments were performed on the Kaggle fake news dataset* which is related to the fake news published during the time of the 2016 U.S Presidential Election.

In the proposed model by Goldani, Momtazi, and Safabakhsh (2021), parallel capsule networks were presented to detect fake news. They used four parallel capsule networks using the non-static word embedding for medium or long news statements, and two parallel capsule networks using static word embedding for short news statements. In another research, they implemented a CNN network using static and non-static word embedding and margin loss (Goldani, Safabakhsh, & Momtazi, 2021). In both researches, they conducted their experiments on the two datasets, namely ISOT (Ahmed et al., 2017) and LIAR (Wang, 2017).

Kaliyar et al. (2021) proposed a BERT-based deep learning model, called FakeBERT, by combining three single-layer CNN with different kernel sizes and filters followed by two stacked single-layer CNNs, a flatten layer, and two dense layers to detect fake news. Their experiments were evaluated on the Kaggle fake news dataset.

Nasir et al. (2021) developed a combination of CNN and LSTM to extract local features and learn long-term dependencies to classify news as real or fake. They evaluated their experiments on two datasets, namely ISOT (Ahmed et al., 2017) and FA-KES (Salem et al., 2019).
2.1 | Related works on Persian fake news detection

In addition to the research studies described above on fake news detection of the English data, there is a limited number of researches on detecting Persian fake news. In most of these studies, conventional machine learning methods are used. Zamani et al. (2017) collected a dataset of 783 Persian rumour tweets from two Iranian websites, including gomaneh.com† and shayeaat.ir‡ and added an equal number of random non-rumour tweets to the dataset. They used structural and content-based features as the input of four different machine learning methods, including decision tree, naive Bayes, sequential minimal optimization, and instance-based learner.

Mahmoodabadi et al. (2018) collected about 3.5 million posts from almost 12,000 users’ profiles in Twitter. From the crawled data, 4345 tweets were labelled as rumours. The synthetic minority over-sampling technique (Chawla et al., 2002) was used to overcome the data imbalance problem. Classifiers were trained with three different types of features: (a) structural features, (b) content features, and (c) demographic features. These three types of features were defined as features related to the social media structure, the text of each tweet, and the user who sends the tweet, respectively. Different classifiers were selected for training with the crawled dataset, such as random forest, decision tree, sequential minimal optimization, and multilayer perceptron.

Zarharan et al. (2019) prepared a dataset that consists of 534 claims and 2124 associated news articles for stance classification. They developed logistic regression, Support Vector Machine (SVM), random forest, naive Bayes, and stacked LSTM models. Several features were used to train the classifiers like bag-of-words representation, term frequency-inverse document frequency, and cosine similarity between the claim and the article to find out the best feature set that obtains the highest results of a classifier.

Jahanbakhsh-Nagadeh et al. (2020) presented a model to calculate the Spread Power of Rumour (SPR). They computed SPR based on the two factors that Allport and Postman (1947) found as the reasons to spread rumours: (a) importance, and (b) ambiguity. They used 42 features in their study. They used 28 features to train a classifier to study the impact of the features on the importance factor. Moreover, another set of 14 features was used to study the ambiguity factor. Their experiments were conducted on the data developed by Zamani et al. (2017) as well as 882 posts from three Telegram channels of Iranian websites.

In another study, Jahanbakhsh-Nagadeh et al. (2021) presented a model for detecting Persian rumours based on valuable features from content information of social media texts. In this research, the data collected by Jahanbakhsh-Nagadeh et al. (2020) and Zamani et al. (2017) has been used. Additionally, the data developed by Seifkar et al. (2018) related to the Kermanshah earthquake was also used.

Samadi et al. (2021) proposed a processing model for detecting Persian fake news. In the model, the document representation and sequence representation of each word of a document have been used. The models took the advantage of contextualized representation (Devlin et al., 2019). Two classifiers were used in the model such that the single layer perceptron neural network was used in the first model, and the CNN was used in the second model. Furthermore, the research contributed to develop a fake news dataset for Persian in the news media domain. This dataset contained 1861 real news crawled from five reliable Persian online news agencies, namely IRNA, ISNA, Farsnews, Hamshahri and Mehrnews, and 1800 fake news collected from various news agencies.

2.2 | Related works on COVID-19 fake news detection

By the spread of COVID-19, some studies focused merely on misinformation about COVID-19. Al-Rakhami and Al-Amri (2020) collected data through the use of Twitter’s streaming application programming interface and annotated the data by human experts. They proposed a two-level stack-based ensemble learning model that depends on the quality and credibility of information about COVID-19. The proposed model was based on experiments using six different machine learning models, namely naive Bayes, Bayes net, k-nearest neighbour, decision tree, random forest, and SVM. They used SVM plus random forest as a first level model and the decision tree as a meta model. A set of 26 features were used in their proposed model, out of which 9 features were based on the tweets and 17 features were based on users’ information. They concluded that among the features ‘whether an account is verified or not’ is the most important user-level feature. Additionally, the features ‘the number of retweets’, ‘the number of hashtags’, ‘the number of mentions’ and ‘the following rate stay in the next positions’ were used.

Gundapu and Mamid (2021) compared the performance of various machine learning, deep learning, and transformer-based models using the English dataset developed by Patwa et al. (2020). In this study, they used some conventional classification algorithms like linear regression, SVM, passive aggressive classifier, and extreme gradient boosting. Moreover, they implemented different neural models, such as LSTM and Bidirectional LSTM (BiLSTM) with attention mechanism, CNN, and stacked CNN-BiLSTM. They developed individual and ensemble models of the three transformer models, namely BERT (Devlin et al., 2019), A Lite BERT (ALBERT) (Lan et al., 2019), and XLNet (Yang et al., 2019). An ensemble of the transformer-based model computed the average of softmax probabilities from these models, and it achieved the best F1 score by 98.55%.

Wani et al. (2021) developed models such as parallel CNN, single layer LSTM, hierarchical attention networks (Yang et al., 2016), and BiLSTM with attention mechanism. The models used GloVe (Pennington et al., 2014) and Fasttext (Joulin et al., 2017) embeddings. Furthermore they fine-tuned BERT (Devlin et al., 2019), a distilled version of BERT, called DistilBERT, (Sanh et al., 2019), and two publicly shared BERT-based pretrained models on the COVID corpus from the Hugging Face model hub, namely COVID-BERT-Base and COVID-Twitter-Base (Müller
et al., 2020), to adapt the model to the target classification task. They continued their experiments by training BERT and DistilBERT as a language model based on the corpus composed of COVID-19 tweets.** These pre-trained language models were used in a classification model to be adapted to the target task. They evaluated their models based on the COVID-19 fake news detection dataset (Patwa et al., 2020).

3 | PROPOSED MODEL

In this article, we propose a model to detect Persian COVID-19 fake news, but this language suffers from the lack of available data in this domain. The main idea of the proposed transformer model is based on transferring knowledge across languages and domains. Therefore, we perform cross-lingual and cross-domain transfer learning by using deep contextualized cross-lingual representation, namely Cross-lingual Language Model Robustly optimized BERT approach (XLM-RoBERTa) (Conneau et al., 2020), and parallel CNN model as the classifier.

Figure 1 shows the overall architecture of our proposed model. As can be seen, the model can be trained with different sets of data. In this research, we focus on two types of transfer learning models: (1) in-domain cross-lingual learning where we use an English COVID-19 dataset, and (2) in-language cross-domain learning where a general domain Persian dataset is used. We study the impact of using additional data belonged to another language and domain for detecting the Persian COVID-19 fake news to benefit from both types of language transfer and domain transfer models to achieve promising results in our target task.

In the followings, we first review XLM-RoBERTa as a pre-trained cross-lingual language model; then, we describe the architecture of the parallel CNN classifier.

3.1 | XLM-RoBERTa

Devlin et al. (2019) introduced BERT that uses a two-step framework, namely pre-training and fine-tuning, as shown in Figure 2. BERT is a multi-layer bidirectional transformer introduced by Vaswani et al. (2017). The BERT developers presented two main models, namely BERT-base and BERT-large, which differ in terms of parameters, the number of layers, the hidden size, and the number of self-attention heads. They pre-trained BERT with two unsupervised tasks, namely Masked Language Model (MLM) and Next Sentence Prediction (NSP). The BERT input for each token
is made by summing the token embedding (the vocabulary ID for each token), segment embedding (a numeric to distinguish between different sentences), and position embedding (position of each token in the sequence). The BERT model provides two types of representations to be used in various natural language processing tasks: (a) the representation of the whole sentence used in classification tasks; and (b) a matrix containing representations of each token in a sentence.

The RoBERTa model (Liu, Ott, et al., 2019) was proposed to improve the BERT model. They trained a model with larger data and batch size, and longer sequences. They changed the pattern of masking; that is, they used dynamic masking versus static masking in the original model. In addition, they removed the NSP loss function during the pre-training process.

Conneau et al. (2020) applied transformer-based MLM to 100 languages using more than two terabytes of filtered CommonCrawl data. They took the proposed cross-lingual MLM (XLM) (Conneau & Lample, 2019) and RoBERTa (Liu, Ott, et al., 2019) models into consideration. The transformer model, trained with the multilingual MLM objective, uses monolingual data. They sampled streams of text from each language and trained the model to predict the masked tokens in the input. Using cross-lingual representation provided the possibility of using data from other languages to enrich the trained model.

3.2 Convolutional neural network

CNN is the most widely used deep learning architecture in computer vision. After its success in computer vision tasks, CNN was considered by researchers to be used in natural language processing to extract n-grams from texts. In our proposed model, we use the outputs of XLM-RoBERTa model as the input of three parallel CNN classifiers. In the next layers, for each parallel network, there are common layers of CNN-based models, including convolutional layer, and max-pooling layer to learn substantial patterns. At the end, concatenated output of parallel CNNs is fed into a fully-connected layer to make a prediction for the labels. The convolution layers of three parallel CNN use 30 filters with kernel sizes 3, 4, and 5. The model contains a single dense layer which contains 32 units. The architecture of the proposed deep neural model in our framework is represented in Figure 3.

4 DATASETS

In this article, we use three different datasets to apply the idea of cross-lingual and cross-domain transfer learning for fake news detection in Persian. The datasets are described in the followings.

a. An English COVID-19 fake news dataset;
b. A general domain Persian fake news dataset;
c. A Persian COVID-19 fake news dataset.
Two of the datasets that we use, namely the English fake news dataset on COVID19 (Patwa et al., 2020) and the Persian fake news belonged to the general domain (Samadi et al., 2021), have already been normalized and tokenized. The only common normalization that we implemented was separating punctuations from the previous word. For our developed Persian COVID-19 dataset, we used the preprocessing algorithm introduced by Ghayoomi (2019) which does both normalization and tokenization with the accuracy of 97.80%.

Patwa et al. (2020) introduced the English COVID-19 dataset. They manually annotated a set of 10,700 posts and articles of real and fake news about the pandemic from social media. The fake and real news were collected from several fact-checking websites and Twitter. Table 1 provides the statistics about the English COVID-19 dataset.

Persian, in general, is one of the low-resource languages. Although attempts have been made to develop language resources for this language, the developed data does not contain variability in the new research domains, such as fake news, and does not update regularly. The fake news detection data developed by Zamani et al. (2017) and Jahanbakhsh-Nagadeh et al. (2020) are mostly from social media. These two datasets, however, are not publicly available. Therefore, we used a new dataset recently developed by Samadi et al. (2021) from online news agencies. The statistics of this general domain Persian fake news dataset is provided in Table 2.

As mentioned, the three Persian fake news datasets provided by Zamani et al. (2017), Jahanbakhsh-Nagadeh et al. (2020) and Samadi et al. (2021) are in general domain, and none of them is in the domain of COVID-19. To fill the gap, we contributed to develop a new
dataset in this domain. This newly developed dataset is collected from social media, including Twitter and Instagram, as well as the websites of various Persian news agencies. Our Persian COVID-19 fake news dataset includes 265 real news and 265 fake news, reported in Table 3. The labels ‘fake’ and ‘real’ are utilized to annotate this dataset. Considering the small size of the dataset, in this research, we use this dataset as the test data. Only in the baseline experiment (the first row in Table 6) this dataset is used for training in a 5-fold cross-validation mode.

5 | EVALUATION

5.1 | Experimental setup

In our research, we used the XLM-RoBERTa model for text representation. We utilized English COVID-19 dataset and the Persian general domain dataset to build the model. The parameters of the proposed model were set during the model training process. The values of the parameters are summarized in Table 4.

5.2 | Results

We evaluated our model in three different ways such that each of them focuses on one of the datasets described in Section 4. Considering that three different datasets are available for our research, before focusing on Persian COVID-19 fake news detection, we first performed experiments on the English COVID-19 and Persian general domain datasets that contain a large number of data.

In the first experiment, we used the English COVID-19 data described in Table 1 for both training and testing. The performance of the CNN classifier for the English dataset as a rich language was 96.61% based on F-measure. This performance can be considered as the upper-bound of the performance for other languages to obtain a similar result.
In the second experiment, we focused on the general domain Persian fake news detection task, where both training and test data belong to the dataset were described in Table 2. According to the experimental results, our model achieved the performance of 87.09% based-on F-measure.

The results of these two experiments are reported in Table 5. The results show the status of fake news detection on the domain of COVID-19 in English as well as the general domain data of a low-resource language.

In the third experiments, we focused on detecting Persian COVID-19 fake news and proposed four learning scenarios for evaluating the model:

a. The baseline model: Training the model with the Persian COVID-19 data;
b. The Cross-Lingual Learning (CLL) model: Training the model with the English COVID-19 data;
c. The Cross-Domain Learning (CDL): Training the model with the Persian general domain data;
d. The Cross-Lingual and Cross-Domain Learning (CLCDL): Training the model with both English COVID-19 dataset and Persian general domain dataset.

The obtained results of the scenarios are reported in Table 6. In the first learning scenario, we only used the Persian COVID-19 data without any help of other available datasets. We consider this learning scenario as the baseline. In this experiment, both training and test data belong to the COVID-19 domain of Persian fake news as described in Table 3. We used 5-fold cross-validation to evaluate this model. According to the experimental results, when the training and test data belong to the same domain and language, the model achieves 69.74% based on F-measure. The result does not seem to be high due to the small amount of the available data for a specific domain in a low-resource language. To overcome the data sparsity problem, we proposed learning scenarios which benefit from transfer learning.

In the second learning scenario, the cross-lingual transfer learning method is used such that the model is trained with the English data and tested with the Persian data. The domain of both training and test data is the same and belongs to the COVID-19 domain. The F-measure of this model was 67.05% which was almost 2.7% lower than the baseline. This result showed that the cross-lingual transfer learning method could not improve the performance, and in addition to overcome the data sparsity problem another issue should be taken into consideration.

In the third learning scenario, the cross-domain transfer learning method was used such that the model was trained with the Persian data in the general domain of fake news and tested with the data belonged to the COVID-19 domain. The F-measure of this model was 70.55% which was slightly better than the baseline. This result showed that the cross-domain transfer learning method had a positive impact in our model to improve the performance but not significantly. The result determines that the equality of the domain between training and test data is important.

Finally, in the fourth learning scenario, a model constructed from the combination of cross-lingual and cross-domain learning models was proposed such that the English COVID-19 data along with the Persian general domain dataset on fake news were used for training the model, and the model was tested with the Persian COVID-19 fake news data. This learning scenario outperformed the other proposed learning scenarios. This model obtained 2.39% improvement compared to the baseline which is statistically significant (p < 5) according to the two-tailed t-test. In this learning scenario, the advantage of the cross-domain transfer learning method is enhanced with the cross-lingual transfer learning method to resolve the data sparsity problem for low-resource languages.

### Table 5 Performance of the proposed model on English COVID data and Persian general domain data

| Dataset          | Results (%)                  |
|------------------|------------------------------|
|                  | Precision | Recall | F-measure |
| English COVID-19 | 94.22     | 99.12  | 96.61     |
| Persian general domain | 94.46 | 80.78  | 87.09     |

### Table 6 Performance of the proposed model using various training and test sets

| Mode   | Language | Domain | Results (%) |
|--------|----------|--------|-------------|
|        | Train    | Test   | Train       | Test       | Precision | Recall | F-measure |
| Baseline | Persian | Persian | COVID-19    | COVID-19   | 75.98     | 64.97  | 69.74     |
| CLL     | English  | Persian | COVID-19    | COVID-19   | 53.76     | 89.06  | 67.05     |
| CDL     | Persian  | Persian | General     | COVID-19   | 56.62     | 93.58  | 70.55     |
| CLCDL   | English + Persian | Persian | COVID-19 + General | COVID-19 | 62.96     | 85.27  | 72.13     |

Abbreviations: CDL, Cross-Domain Learning; CLCDL, Cross-Lingual and Cross-Domain Learning; CLL, Cross-Lingual Learning.
According to the experimental results, it can be concluded that the cross-lingual transfer learning method in the second learning scenario can extract information about COVID-19 from the English dataset, but this dataset cannot acquire appropriate information about the Persian language. On the other hand, in the third learning scenario, the cross-domain transfer learning method can learn about the properties of Persian fake news, but it does not learn much about the context of COVID-19 news. The results of these two models indicate that the cross-domain transfer model obtains the features better than cross-language transfer model to detect fake news; therefore, linguistic properties of a language has a positive impact on the performance of a model. Taking the advantages of both transfer learning approaches in the fourth learning scenario, the best result has been obtained which indicates that having data from related domains of the test data in the training data is important even if the information source is another language. In addition, having data from the same language as the test data is important even if the information source is in another domain. As a result, both domain and language specific properties are important features to achieve the best performance of the model.

5.3 | Error analysis

In the followings, four examples from the Persian COVID-19 dataset are discussed. The first two examples are true; but they are algorithmically recognized as fake news. The next two examples are fake; but they are algorithmically detected as real news.

The sample news in Figure 4, labelled ‘real’ in the gold data, is about a celebrity who has been hospitalized due to COVID-19 in a luxurious hospital in Tehran such that a large amount of Remdesivir for treatment of COVID-19 has been provided and food, such as chicken stroganoff, lobster, and Pizza, has been served. Since most of the news with such content about celebrities are fake, the content of the news seems to be fake and the algorithm assigned a wrong label to this news. None of our proposed learning scenarios assigned a correct label to this news.

The sample news in Figure 5, labelled ‘real’ in the gold data, is about taking advantages of using technologies such as cyberspace, social media, and the Internet for education of students either primary and secondary school and higher education during the pandemic to prevent distribution of the disease. Since most of the news about cyberspace and social media are fake, the algorithms assigned a wrong label to this news and detected it as fake news.

Figure 4 First example of Persian real news detected as fake

Figure 5 Second example of Persian real news detected as fake
The sample news in Figure 6, labelled ‘fake’, is about sending 40,000 COVID-19 test kits from Iran to Germany, as quoted by Bloomberg. It is also stated that other countries, like Spain, Brazil, Turkey, and Ecuador asked for the test kit. The content of the news can be generally true, because during the pandemic sending test kits to other countries as a charity was normal and such information appeared in many real news. But for this specific news, it is fake.

The content of the sample news in Figure 7 is very similar to the previous example. The news is lengthy and it contains direct speech of Iran’s vice president of science and technology. Similar to the previous example, the news is fake but it is algorithmically recognized as real news. The reason is that most of the news that contains direct speech is true. Therefore, it is not easy for the models to make a correct decision.

6 | CONCLUSION AND FUTURE WORKS

In this article, we implemented knowledge transferring across language and across domain models using a parallel CNN classifier and the XLM-RoBERTa language model to detect Persian COVID-19 fake news. We compared the results obtained from the proposed model in four different scenarios. Three different datasets were used to evaluate the proposed model: an English COVID-19 dataset (Patwa et al., 2020), a general domain Persian fake news dataset (Samadi et al., 2021), and a Persian dataset for fake news detection on COVID-19 which contributed to be developed in the current research. The combination of these two knowledge transferring models resulted in a 2.39% improvement in F-measure. Using the English COVID-19 fake news dataset solely in the cross-lingual transfer learning model did not improve the results, because the model did not learn anything about the linguistic properties of Persian. If the model is trained with a dataset in a language that is similar to the test data, say Persian in our experiments, it has a slightly positive impact on the performance of the model; because the model has learnt about the linguistic properties of the target language, but it might not perform properly due to lack of knowledge in the target domain.

To continue this research in future, we can extract user’s information from news articles in social media as features and use it as meta-data to train our proposed neural network models and compare the results.
ENDNOTES

*https://www.openicpsr.org/openicpsr/project/113992/version/V1/view
http://gomaneh.com/
http://shayeaat.ir/
https://irma.ir/
https://isra.ir/
https://www.farsnews.ir/
https://www.hamshahrionline.ir/
https://mehrenews.com/
https://huggingface.co/models
https://huggingface.co/deepset/covid_bert_base
***https://www.kaggle.com/gpreda/covid19-tweets

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Massood Ghayoomi https://orcid.org/0000-0001-6685-1332

REFERENCES

Ahmed, H., Traore, I., & Saad, S. (2017). Detection of online fake news using n-gram analysis and machine learning techniques. In International conference on intelligent, secure, and dependable systems in distributed and cloud environments (pp. 127–138).
Allport, G. W., & Postman, L. (1947). The psychology of rumor. Henry Holt.
Al-Rakhami, M. S., & Al-Amri, A. M. (2020). Lies kill, facts save: Detecting COVID-19 misinformation in twitter. IEEE Access, 8, 155961–155970.
Al-Zaman, M. S. (2021). COVID-19 related social media fake news in India. Journalism and Media, 2(1), 100–114.
Barua, Z., Barua, S., Aktar, S., Kabir, N., & Li, M. (2020). Effects of misinformation on covid-19 individual responses and recommendations for resilience of disastrous consequences of misinformation. Progress in Disaster Science, 8, 100119.
Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique.
Conneau, A., & Lample, G. (2019). Cross-lingual language model pretraining. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, & R. Garnett (Eds.), Proceedings of the 58th annual meeting of the association for computational linguistics (pp. 8440–8451). Association for Computational Linguistics.
Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., & Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale. In Proceedings of the 58th annual meeting of the association for computational linguistics (pp. 8440–8451). Association for Computational Linguistics.
Goldani, M. H., Momtazi, S., & Safabakhsh, R. (2021). Detecting fake news with capsule neural networks. Applied Soft Computing, 101, 106991.
Goldani, M. H., Momtazi, R., & Momtazi, S. (2021). Convolutional neural network with margin loss for fake news detection. Information Processing & Management, 58(1), 102418.
Goulangos, S., & Mamid, R. (2021). Transformer based automatic COVID-19 fake news detection system. CoRR. https://arxiv.org/abs/2101.00180
Huang, Y.-F., & Chen, P.-H. (2020). Fake news detection using an ensemble learning model based on self-adaptive harmony search algorithms. Expert Systems with Applications, 159, 113584.
Jahanbakhsh-Nagadeh, Z., Feizi-Derakhshi, M.-R., Ramezani, M., Rahkar-Farshi, T., Asgari-Chenaghlu, M., Nikzad-Khamsaki, N., & Balafar, M.-A. (2020). A model to measure the spread power of rumors. CoRR. https://arxiv.org/pdf/2002.07563.pdf
Jahanbakhsh-Nagadeh, Z., Feizi-Derakhshi, M.-R., & Sharifi, A. (2021). A model for detecting of Persian rumors based on the analysis of contextual features in the content of social networks. Signal and Data Processing, 18(1), 29–50.
Joulin, A., Grave, E., Bojanowksi, P., & Mikolov, T. (2017). Bag of tricks for efficient text classification. In Proceedings of the 15th conference of the European chapter of the association for computational linguistics (pp. 427–431). Valencia, Spain: Association for Computational Linguistics.
Kaliyar, R. K., Goswami, A., & Narang, P. (2021). FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. Multimedia Tools and Applications, 80, 1–24.
Kaliyar, R. K., Goswami, A., Narang, P., & Sinha, S. (2020). Fndnet – A deep convolutional neural network for fake news detection. Cognitive Systems Research, 61, 32–44.
Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). ALBERT: A Lite BERT for self-supervised learning of language representations. CoRR. http://arxiv.org/abs/1909.11942
Liu, C., Wu, X., Yu, M., Li, G., Jiang, J., Huang, W., & Lu, X. (2019). A two-stage model based on BERT for short fake news detection. In International conference on knowledge science, engineering and management (pp. 172–183).
Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. https://arxiv.org/abs/1907.11692

Lopez, C. E., Vasu, M., & Gallemore, C. (2020). Understanding the perception of COVID-19 policies by mining a multilanguage twitter dataset. CoRR. https://arxiv.org/abs/2003.10359

Mahmooodabad, S. D., Farzi, S., & Bakhtiarvand, D. B. (2018). Persian rumor detection on twitter. In 2018 9th international symposium on telecommunications (IST) (pp. 597–602).

Memon, S. A., & Carley, K. M. (2020). Characterizing COVID-19 misinformation communities using a novel twitter dataset. CoRR. https://arxiv.org/abs/2008.00791

Müller, M., Salathé, M., & Kummervold, P. E. (2020). Covid-twitter-BERT: A natural language processing model to analyse COVID-19 content on twitter. CoRR. https://arxiv.org/abs/2005.07503

Nasir, J. A., Khan, O. S., & Varlamis, I. (2021). Fake news detection: A hybrid CNN-RNN based deep learning approach. International Journal of Information Management Data Insights, 1(1), 100007.

Patwa, P., Sharma, S., Pyki, S., Guptha, V., Kumari, G., Akhtar, M. S., & Chakraborty, T. (2020). Fighting an infodemic: COVID-19 fake news dataset. CoRR. https://arxiv.org/abs/2011.03327

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532–1543).

Qazi, U., Imran, M., & Ofli, F. (2020). Geocov19: A dataset of hundreds of millions of multilingual COVID-19 tweets with location information. SIGSPATIAL Special, 12(1), 6–15.

Saleem, F. K. A., Al Fee, R., Elbasuoni, S., Jaber, M., & Farah, M. (2019). Fa-kes: A fake news dataset around the Syrian war. In Proceedings of the international AAAI conference on web and social media (Vol. 13, pp. 573–582).

Samadi, M., Mousavian, M., & Montazi, S. (2021). Persian fake news detection: A deep neural representation and deep learning approach. ACM Transactions on Asian and Low-Resource Language Information Processing, 21(1), 10.

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. CoRR. https://arxiv.org/abs/1910.01108

Seifikar, M., Farzi, S., & DehghaniMahmooodabad, S. (2018). Kermanshah earthquake event tracking through Persian tweets. In Proceedings of the 9th international symposium on telecommunications (p. 424–428).

Swire-Thompson, B., & Lazer, D. (2020). Public health and online misinformation: Challenges and recommendations. Annual Review of Public Health, 41(1), 433–451.

van der Linden, S., Roozenbeek, J., & Compton, J. (2020). Inoculating against fake news about COVID-19. Frontiers in Psychology, 11, 2928.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). Attention is all you need. In I. Guyon, et al. (Eds.), Advances in neural information processing systems (Vol. 30). Curran Associates, Inc.

Wang, W. Y. (2017). "Liar, Liar pants on fire": A new benchmark dataset for fake news detection. In Proceedings of the 55th annual meeting of the association for computational linguistics (Vol. 2: Short papers, pp. 422–426). Vancouver, Canada: Association for Computational Linguistics.

Wani, A., Joshi, I., Khandve, S., Wagh, V., & Joshi, R. (2021). Evaluating deep learning approaches for Covid19 fake news detection. CoRR. https://arxiv.org/abs/2101.04012

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). XLNet: Generalized autoregressive pretraining for language understanding. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché Buc, E. Fox, & R. Garnett (Eds.), Advances in neural information processing systems (Vol. 32). Curran Associates, Inc. (p. 32).

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016). Hierarchical attention networks for document classification. In Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Human language technologies (pp. 1480–1489).

Zamani, S., Asadpour, M., & Moazzami, D. (2017). Rumor detection for Persian tweets. In 2017 Iranian conference on electrical engineering (ICEE) (pp. 1532–1536).

Zarharian, M., Ahangar, S., Rezvaninejad, F., Bidhendi, M., Pilevar, M. T., Mineai, B., & Etemad, S. (2019). Persian stance classification data set. In Test. Zarocostas, J. (2020). How to fight an infodemic. The Lancet, 395(10225), 676.

AUTHOR BIOGRAPHIES

Masood Ghayoomi is an assistant professor at the Institute for Humanities and Cultural Studies, Tehran, Iran. He received his PhD degree in 2014 from the Department of Mathematics and Computer Science at Freie Universität Berlin, Berlin, Germany. His main research interest focuses on computational linguistics, machine learning, syntax and lexical semantics. During his PhD, he was a visiting researcher at Stuttgart University, Stuttgart, Germany. After his PhD, he worked as a research assistant at Pädagogische Hochschule Karlsruhe and Duale Hochschule Baden-Württemberg, Karlsruhe, Germany.

Maryam Mousavian is currently an M.Sc. holder of Artificial Intelligence from the Amirkabir University of Technology, Tehran, Iran. She received her B.Sc. degree in Software Engineering in 2017 from the Isfahan University, Isfahan, Iran. Natural language processing is her main research interest.